Human Plausible Reasoning for Question Answering Systems

E. DARRUDI, M. RAHGOZAR, F.OROUMCHIAN

Abstract--**Question Answering (QA) systems aim to communicate with users directly through natural language to better meet the needs of information seekers. In this paper we describe our knowledge-based QA system that employs the theory of Human Plausible Reasoning (HPR) in its core reasoning engine. The experiments conducted in with CIA World Factbook (CWFB) show that system's answers are analogous to human beings' replies when the required relations by HPR are added to the underlying factual data provided in CWFB.**

Index Terms--**Question Answering, Human Plausible Reasoning, Chaining Reasoning, CIA World Factbook**

I. Introduction

In recent years, Question Answering (QA) systems have evolved out of the field of Information Retrieval to meet better the needs of information seekers. Unlike the simple keyword-based information retrieval systems (e.g. Web search engines), they aim to communicate directly with users through a natural language. They accept natural language questions and return exact answers, eliminating the burden of query formulation and reading lots of irrelevant documents to reach the answer.

An ideal QA system would converse with users to fully understand their information needs. This way, it bears much resemblance to the machine that would ultimately pass the symbolic Turing Test [1].

QA systems have become increasingly complex over the few years such that there is now little in common across all systems. This complexity and diversity is apparent from open-domain QA systems (systems which deal with unrestricted questions upon large-scale text corpora) take part annually in the Text Retrieval Conference (TREC) QA track [2]. It has been the same for domain-specific (closed domain or restricted domain) QA systems that seek to answer questions in much smaller controlled domains via well-structured knowledge sources. As questions asked by users are getting harder, QA systems will have to understand the meanings exploiting more sophisticated AI techniques compared to merely statistical and shallow ones used by most QA systems at present. Indeed, a few prototypes of such intelligent QA systems exist now. Among those participated in the QA track of TREC 2002 [2], PowerAnswer [3], a logic-based QA system, ranked first. It utilized a logic prover and WordNet [4] to pick out the correct answer from retrieved candidate answers.

In This paper, we describe our knowledge-based QA system that exploits the theory of Human Plausible Reasoning (HPR) [5] as the core of its reasoning engine. The theory was proposed by Collins & Michaski in 1989. It was developed based on experimental observations that simulate question-answering situations where information is incomplete or uncertain and dynamically changing. A human-like reasoning can extract implicit knowledge and boost the QA's performance. Our proposed approach replies with *Exact Answers* whenever possible, just like other QA systems. Moreover, it tries to reason and return a *Plausible Answer* when an exact answer is not available.

The remaining of this paper is structured as follows. In section II, we scan the literature for related work on HPR. In section III, we take a closer look at HPR and its cornerstones. The architecture of our QA system is discussed in section IV. Section V describes the knowledge representation in our QA system and Section VI designates the set of basic human plausible inferences implemented in the system. Our core reasoning algorithm using these basic inferences is explained in section VII and the experimental results using this algorithm are discussed in section VIII. Finally, section IX concludes the paper and provides some guidelines for future research.

II. Related Work

HPR has been employed successfully in several information-based applications so far. In [6], the author describes a pilot version of the theory on a problem in the domain of the chemical periodic table. Kelly [7] developed an expert system for grass identification based on HPR. Oroumchian applied HPR for an experimental information retrieval system called PLIR [8]. In [9], [10] and [11] authors suggest some applications of the theory for adaptive filtering, intelligent tutoring and document clustering, respectively. The theory has also been beneficial in designing intelligent graphical user interfaces offer pieces of advice to their users

Manuscript received November 2, 2004.

E. Darrdui is with Iran Telecommunication Research Center, P.O.BOX 14155-3961, Tehran, Iran (telephone: +98-21-8497483, e-mail: darrudi@ itrc.ac.ir).

M. Rahgozar is with the Electrical & Computer Engineering Department, University of Tehran and a member of Control and Intelligent Processing Center of Excellence, Faculty of Eng. in University of Tehran, P.O.BOX 14395-515, Tehran, Iran (Telephone: +98-21-8020403-3504, email: rahgozar@ut.ac.ir).

F. Oroumchian was with Electrical & Computer Engineering Department, Faculty of Eng. University of Tehran. He is now with University of Wollongong in Dubai, P.O.BOX 20183, Dubai, UAE (telephone: +971-503585420, email: foroumchian@acm.org).

[12][13]. All these implementations confirm the usefulness and flexibility of HPR for applications that need to reason about and interact with human beings.

III. The Theory of Human Plausible Reasoning

For 15 years, Collins and his colleagues have been collecting and organizing a wide variety of human plausible inferences people do in everyday's life [14]. His collaboration with Michalski led to development of a formal system based on Michalski's variable valued logic [15] that characterized different patterns of plausible inferences people use in reasoning about the world. [5] [16]. They attempted to formalize plausible inferences that frequently occur in people's responses to questions for which they don't have ready answers. In the formal notation of the theory, a *Statement* like "we are almost sure that GDP per capita in Luxembourg is $55,100" is written as:

GDP per capita (Luxembourg) = {$55,100}, $\gamma = 0.93$

where "GDP per capita" is a *Descriptor*, "Luxembourg" is an *Argument*, "$55,100" is a *Referent*, and γ is the certainty in the statement (in this case it shows high confidence). There are many other certainty parameters that give a descriptive power to plausible expressions. Some parameters which have been used in our implementation include frequency (φ), dominance (δ), similarity (σ) and conditional likelihood (α). The φ parameter quantifies the frequency of the referent in domain of the descriptor. For example the following statement conveys that only 10% of Albanians are Roman Catholic.

Religion (Albania) = {Roman Catholic}, $\varphi = 0.1$

The δ parameter is used to show the dominance of a subset in a set in ISA and PARTOF hierarchies. For example, we might express that Australia constitutes a large percentage of Oceania as stated below.

PARTOF (Australia) = {Oceania}, δ = very high

The σ parameter demonstrates the degree of similarity between two concepts in a particular context. The following statement expresses that Somalia and Yemen are very similar in respect to their population growth rate.

Somalia SIM Yemen, σ = high, CX: Population Growth Rate

Besides simple statements, *Dependencies* and *Implications* can also form logical expressions. A marked dependency shows that the increase in one side of the relation makes the other side concept decrease (negative dependency) or increase (positive dependency). For example, the positive dependency "Money Supply (Country) +→ Inflation Rate (Country)" cites that by increasing the money supply in a country we can cause the inflation rate to rise in that country (the Monetarist view [17]). Unmarked dependencies just show that one concept has an effect on another, e.g., terrain (area) → agricultural products (area).

Implications are bounded dependencies which can represent simple rules in the KB. The below expression mentions that: "if a country has a tropical climate, then we can expect mangos as an agricultural product for that country with 70% probability."

Climate (Country) = {Tropical} => Agricutural Product (Country) = {Mangos}, $\alpha = 0.7$

The α parameter in the above statement shows the conditional likelihood that the right-hand side of the implication (or dependency) has a particular value (referent), given the left-hand side value. In this case it shows only a probability of 70%.

The theory has a rich set of transforms and inferences which operate on the above expressions to infer new statements which might answer the question in hand. Section VI elaborates on the set of inferences implemented in our system.

IV. The Question Answering Makeup

Fig. 1 depicts the architecture of our knowledge-based QA system. The system is composed of four interconnected components: Question Analyzer, Reasoning Engine, Natural Language Generator and a knowledge base (KB) which supplies the axioms and rules about the desired domain. Virtually, any KB which possesses essential HPR's relations can be plugged into the system to change its domain of utilization. Currently, we carry out experiments with CIA World Factbook [18] KB. The process of making an appropriate KB from the original Factbook is described is the section VIII.

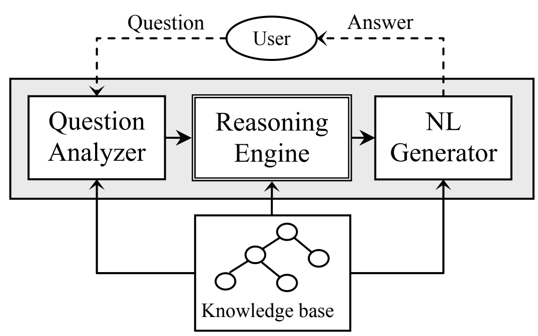

Fig. 1. The architecture of our QA system. Question Analyzer parses the question in English. Reasoning Engine finds the answer and Natural Language Generator present the answer to the user in English again. All the components need a Knowledge base to carry out their intended tasks.

The user enters his question in natural language. In the first step, Question Analyzer parses the inquiry and converts it to a partial statement called *Plausible Question*. The component tries to map question keywords to concepts which exist in the KB. In terms of HPR, it should determine a descriptor and an argument to form a plausible question. Table I, shows some sample questions and their plausible representations produced by Question Analyzer.

Given a plausible question, the reasoning engine tries to infer the missed referent and convert the plausible question to a statement. This process is described in section VII. After finding the answer, a natural language generator presents it to the user with natural language expressions. The system also possesses the capability to justify its answers. As all inferences are carried out in the semantic layer, the inferences and intermediate results are understandable to end users. The justification is provided if the user asks for one.

TABLE I
SOME SAMPLE QUESTIONS AND THEIR PLAUSIBLE REPRESENTATIONS

Natural Question	Plausible Question
What languages are spoken in Norway?	Language (Norway) = {?}
Where is Norfolk Island?	Location (Norfolk Island) = {?}
What is the gross domestic production in China?	GDP (China) = {?}
How is the population growth rate in Middle East in general?	Population Growth Rate (Middle East) = {?}

V. KNOWLEDGE REPRESENTATION

To be able to reason efficiently, one should have mechanisms to represent effectively the knowledge about the desirable domain. In our system the domain knowledge is stored in a specific implementation of semantic networks. In semantic networks concepts are represented with nodes in the knowledge graph. Relations are usually denoted by links between these concepts. We extend this idea in which we treat both concepts and relations as simple nodes. This way, each node in the semantic network may represent an argument, referent or descriptor. This scheme, gives extra power to the knowledge representation, lets relations own properties and relations to other concepts. Indeed, it is a relaxed implementation of conceptual graphs [19] which allows also representation of more complex information such as rules and contextual data, in spite of its simplicity. Fig. 2 depicts some statements about France in this representation.

Some relations play an important role in HPR. ISA and PARTOF relations make the hierarchies needed to draw basic inferences such as generalization and specification. Likewise, SIM relations are needed for analogy. It is important to note that each SIM relation is valid in a particular context, e.g. France is similar to Spain in respect to their people's life expectancy, as depicted in Fig. 2.

In addition to simple axioms, we can also store simple rules using the proposed approach. Fig. 3, illustrates a typical rule in this representation. It states that if climate of a place is tropical then it implies, with a conditional likelihood of 70%, the products of that place are mangoes; and we are certain about the source of this piece of knowledge.

VI. BASIC PLAUSIBLE INFERENCES

Our selected set of plausible inferences from HPR consists of three statement transforms and four dependency and implication-based inferences. They are explained here in an abstract manner. See the experiments' section for practical examples of their usage.

The simplest inferences in the HPR theory are statement transforms. If a statement is believed to be true with some confidence, then there are three inferences that allow plausible conclusions to be drawn. They are *Generalization Transform*, *Specification Transform,* and *Similarity Transform*.

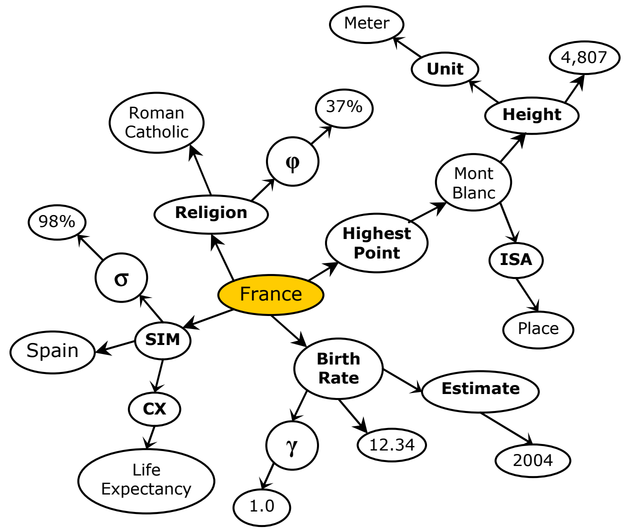

Fig. 2. Some statements about France in the proposed knowledge representation. Descriptors have been shown in bold face.

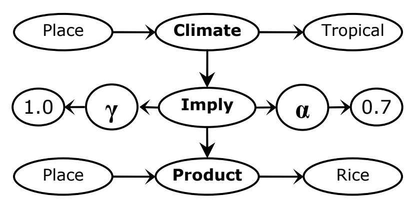

Fig. 3. Representation of an implication in the system. It states that if the climate of a place is tropical then, with a conditional likelihood of 70%, mangos are a product of that place; and we are definitely certain about this piece of knowledge.

Fig. 4 depicts the Generalization Transform[1]. The final certainty in the conclusion is a function of the degree of belief in the original statement, certainty in the ISA or PARTOF relation and the respective frequency and dominance parameters. Currently, this function is the normal product of all parameters. Unknown parameters are excluded when computing the certainty.

1. Arg_1 **ISA or PARTOF** Arg_2 : γ_{isa}, δ
2. Des (Arg_1) = {Ref} : γ_{stat}, φ

Des (Arg_2) = {Ref} : $\gamma_{con} = F(\gamma_{stat}, \gamma_{isa}, \varphi, \delta)$

Fig. 4. Generalization Transform. It accepts an expression and a hierarchical relation (ISA or PART OF) and produces another expression with a generalized argument.

[1] In the figures Des, Arg, Ref, Stat, Dep, Imp and Con stand for Descriptor, Argument, Referent, Statement, Dependency, Implication and Conclusion, respectively.

Specification Transform acts in the same way but traverses ISA hierarchies downward. The Similarity transform employs SIM relations to perform a sideway traversal to alter the statement in hand to one that may answer the question. Indeed, it implements the classical analogy. Fig. 5 shows this inference. Although there may exist several similarities between two concepts from different viewpoints (contexts), however, those with relevant contexts to the question are appropriate. For example, to answer the question "can a goose quake?" one may consider that geese are similar to ducks in the context of their beaks and the shape of the beak has an effect on the bird's sound therefore geese can quack as well as ducks do.

1. Arg_1 **SIM** Arg_2, CX: Des_1 : γ_{sim}, σ
2. $Des_1 (Arg) \rightarrow Des_2 (Arg)$: γ_{dep}, α
3. Arg_1 **ISA** Arg : γ_{isa1}, δ_1
4. Arg_2 **ISA** Arg : γ_{isa2}, δ_2
5. $Des_1 (Arg_1) = \{Ref\}$: γ_{stat}, φ

$Des_2 (Arg_2) = \{Ref\}$: $\gamma_{con} = F (\gamma_{stat}, \gamma_{sim}, \gamma_{dep}, \gamma_{isa1\text{-}2}, \varphi, \sigma, \alpha, \delta_{1\text{-}2})$

Fig. 5. Similarity Transform. Each SIM relation is expressed in a context (represented with a descriptor). The transform is valid when the question's descriptor is satisfied by the context (the dependency relation).

Dependency and implication-based inferences include *Derivation from Dependency*, *Transitivity Inference, Derivation from Implication* and *Dependency-based Analogy*. Derivation from Dependency is applicable to marked dependencies. Fig. 6 depicts a derivation from positive dependency. For Derivation from negative dependency the increase and decrease happen in opposite directions.

1. $Des_1 (Arg) +\rightarrow Des_2 (Arg)$: γ_{dep}, α
2. Arg_1 **ISA** Arg : γ_{isa}, δ
3. $Des_1 (Arg_1) = \{High \mid Medium \mid Low\}$: γ_{stat}, φ

$Des_2 (Arg_1) = \{High \mid Medium \mid Low\}$: $\gamma_{con} = F (\gamma_{stat}, \gamma_{dep}, \gamma_{isa}, \varphi, \alpha, \delta)$

Fig. 6. Derivation from a positive dependency. When the left hand side of the dependency is low, medium or high, then the right hand side is low, medium or high, respectively.

Derivation from Implication inference checks the preconditions of an implication and infers the consequence if all preconditions are met. Fig. 7 illustrates this inference.

Transitivity Inference combines two dependencies or two implications when there is no direct dependency or implication between desired concepts. For example, if the reasoning engine holds that Net National Product (NNP) of a country affects the Gross Domestic Product (GDP) of that country and GDP in turn influences the unemployment rate, then it concludes NNP affects the unemployment rate directly, though with less certainty and conditional likelihood.

1. $Des_1 (Arg) = \{Ref_1\} \Rightarrow Des_2 (Arg) = \{Ref_2\}$: γ_{imp}, α
2. Arg_1 **ISA** Arg : γ_{isa}, δ
3. $Des_1 (Arg_1) = \{Ref_1\}$: γ_{stat}, φ

$Des_2 (Arg_1) = \{Ref_2\}$: $\gamma_{con} = F (\gamma_{stat}, \gamma_{dep}, \gamma_{isa}, \varphi, \alpha, \delta)$

Fig. 7. Derivation from Implication. The consequence of the implication is deduced if the antecedent exists in the KB.

The weakest type of inference is Dependency-based Analogy. Indeed, it is a reformulation of Similarity Transform discussed above nevertheless it is more flexible in which it can discover similarities online. This inference is not in the original proposed set of HPR and we adopted it form [7] and extended it to take account of inexact similarities. Fig. 8 demonstrates this inference.

1. $Des_1 (Arg) \rightarrow Des_2 (Arg)$: γ_{dep}, α
2. Arg_1 **ISA** Arg : γ_{isa1}, δ_1
3. Arg_2 **ISA** Arg : γ_{isa2}, δ_2
4. $Des_1 (Arg_2) = Ref_1$: $\gamma_{stat1}, \varphi_1$
5. $Des_1 (Arg_1) = Ref_3$: $\gamma_{stat2}, \varphi_2$
6. Ref_1 **SIM** Ref_3 , CX: Any : γ_{sim}, σ
7. $Des_2 (Arg_1) = Ref_2$: $\gamma_{stat3}, \varphi_3$

$Des_2 (Arg_2) = \{Ref_2\}$: $\gamma_{con} = F (\gamma_{stat1\text{-}3}, \gamma_{dep}, \gamma_{sim}, \gamma_{isa1\text{-}2}, \varphi_{1\text{-}3}, \alpha, \delta_{1\text{-}2})$

Fig. 8. Dependency-based Analogy. This is the weakest type of basic inferences because of the lengthy chain of preconditions and vast number of influencing parameters.

VII. The Reasoning Algorithm

Generally, invoking a basic inference may not answer immediately the question in hand. The HPR theory does not specify how these basic inferences may be combined together in a controlled manner to find the answer. We propose an algorithm to do this by taking into account some criteria affect the actual reasoning process which takes place in the human brain. In this algorithm, basic plausible inferences can join together to form a chain of reasoning as the preconditions of each basic inference can be the queries for another inferences. The reasoning stops when either an exact answer is found in the KB or some confining conditions are met (it is a kind of backward chaining reasoning).

Reasoning Engine accepts a plausible question converted from a natural language question by Question Analyzer. At first, it searches for the answer in the KB. If the answer could not be found explicitly, the reasoning engine tries to infer it using all possible basic inferences mentioned in section VI. Each inference creates a new plausible question to work on (e.g., the Generalization transform make a question with the generalized argument), then it checks whether the system knows the answer for the new question. It may launch, in turn, other inferences if the answer for the new plausible question isn't again in the KB. The process continues recursively. In each step of the reasoning, after performing all possible inferences, the reasoning engine combines all evidences to produce one answer. At present, the combination is done by adding all supportive evidences for repetitive answers and sorting them according to their belief values (γ). The most certain answer, if any, and its normalized certainty will be returned to the inference initially called the current one.

There are several criteria which control the chain of reasoning in humans. For example, when there are several possible solutions (reasoning paths) to a problem, people tend to try first some inference patterns which are more likely to reach answer quickly. In our system, inferences are prioritized based on their inference types. This heuristic attempts to steer the reasoning in directions that have the greatest chance of finding evidences. Thus, in each step, the reasoning engine tries to carry out inferences in a predetermined order. Generalization and specification seem to be the most used inferences by people, therefore the reasoning engine administers them first. After that, similarity transform and derivation from dependency are examined in order. Transitivity inference is called whenever needed.

Human beings usually don't think concurrently about diverse topics. Additionally, they don't carry deep chaining inferences because the certainty of the expected answer quickly falls below some threshold of plausibility. Our system first performs a limited depth-first search. Therefore, the reasoning engine puts a constraint on the length of a solution path (currently 3). The path length is used to terminate search down a path after a globally defined length exceeded.

Another constraint that may prevent a search path to expand is the certainty in the prospective answer. In any depth of search, the certainty accumulated from utilized inferences and relations to that point determines an upper bound for the answers that may be found thereafter in that path. If the certainty is smaller than a predetermined threshold value, the inference will not proceed further.

VIII. Experiments

There exist well-defined criteria to evaluate open-domain QA systems in related conferences such as TREC. The comparison makes sense because all the systems use the same text corpora as the main source of knowledge for QA (though there is an increasing interest in using parallel sources such as Web). On the other hand, domain-specific QA systems usually benefit from other knowledge sources which are decidedly diverse and highly specific for the intended domain that make the comparison impractical. A question set that seems standard for one or two systems bear no meaning at all for others.

A. *CIA World Factbook as Our Testbed*

To evaluate our reasoning algorithm we chose the CIA World Factbook (CWFB) [18] as the domain KB for QA. CWFB is an annual publication by the Central Intelligence Agency of the United States with basic almanac-style information about the various countries of the world. The factbook gives a two-to-three-page summary of the demographics, location, telecommunications capacity, government, industry, military capability, etc., of almost all countries in existence. It is available free of charge to use and reproduce. There is also an online version of the book which offers the information in semi-structured HTML pages [18]. The wealth of information in this knowledge source persuaded us to use it for our experiments. Although there were ready digital versions of CWFB available [20][21]; however, we decided to build the KB from scratch using the original HTML pages. It was mainly due to their outdated information or little coverage of original data in CWFB.

We developed programs to parse automatically HTML pages from the web site and converted the contained knowledge to our proposed semantic network discussed in section V. Several implicit relations also extracted throughout the process. The ultimate KB contains more than 80,000 relations in about 200 relation categories. The KB is publicly available from [22].

Some text snippets in the original HTML pages specify a time in which the information was gathered or estimated. This information serves as the certainty parameter (γ) for extracted relations applying a probability density function of the uniform cumulative distribution on the interval [1960, Current Time]. For example, a sentence states "the number of heliports estimated for Italy in 2003 is 4" is assigned a certainty of 0.93. All data gathered in the current year (2004) get a certainty of 1. The certainty parameters are recalculated whenever the system reloads the KB.

The dominance parameters (δ) are computed for PARTOF relations. For a country, δ is the proportional size of the country to the total size of the area. For example, we have PARTOF (Australia) = {Oceania}, $\delta = 0.94$. Although δ can be defined ISA relations too; however, there is not enough information in CWFB to compute them correctly.

The frequency parameters (φ) are worked out form quantified data present in CWFB. For example, it is stated that "the people of Lesotho practice Christian 80% and indigenous beliefs 20%." It is transformed to two statements: Religion (Lesotho) = {Christian}, $\varphi = 0.80$ and Religion (Lesotho) = {indigenous beliefs}, $\varphi = 0.20$.

The CWFB provides the underlying factual data needed for reasoning; however to exploit full potential of plausible reasoning it is also required to have some knowledge about the hierarchy of concepts (ISA and SIM relations), dependencies between them (Dependencies) and consequences of statements (Implications). This kind of knowledge is a mixture of domain-specific, lexical and commonsense facts, which has not been encoded in CWFB or in any other KB at present. We added this knowledge manually and asked for help from experts whenever needed.

B. Experiments

Based on CWFB some questions about the world's countries were selected and fed to the system. We tried to choose questions that a typical adult already know or can infer their answers given the required information to compare the results of the systems with normal humans. Answers and justifications of the system are studied below.

Question 1: What is the former name of Iran?
Plausible Question: Former Country Name (Iran) = {?}
Answer: Persia
Certainty: 100%
Justification:
Exact answer was found in the knowledge base.

For the above question system found the answer exactly in the CWFB KB. Thus the certainty in the answer is at the highest possible.

Question 2: What are the main export commodities of Oceania?
Plausible Question: Export Commodity (Oceania) = {?}
Answer: Wheat, Wool
Certainty: 94%
System's Justification:

1. Australia **PARTOF** Oceania, $\gamma = 1$, $\delta = 0.94$
2. Export Commodity (Australia) = {Wheat, Wool}, $\gamma = 1$
3. **Generalization Transform:** Export Commodity (Oceania) = {Whaet, Wool}, $\gamma = 0.94$

The system can't find an answer explicitly in the KB so launches the reasoning process. The reasoning engine is sure that Australia exports wheat and wool and because Australia constitutes a large percent of Oceania (high dominance), it concludes with high confidence that Oceania exports wheat and wool in general. The δ parameter prevents the reasoning algorithm from drawing implausible conclusions. For example, the system also knows that Vanuatu is part of Oceania and it produces cocoa but the system doesn't generalize the cocoa as a product of Oceania because of the low dominance of this country in Oceania. Other plausible parameters have the same responsibility to help the reasoning process avoid implausible results as we are looking into in next examples.

***Question 3**:* How is the fertility rate in Finland?
Plausible Question: Fertility Rate (Finland) = {?}
Answer: Low
Certainty: High
System's Justification:

1. Population Literacy (Country) → Fertility Rate (Country)
2. Norway **ISA** Country, $\gamma = 100\%$
3. Finland **ISA** Country, $\gamma = 100\%$
4. Norway **SIM** Finland, CX: Population Literacy, σ = high
5. Fertility Rate (Norway) = {1.78 children born/woman}, $\gamma = 0.95$
6. Fertility Rate (Norway) = {Low}, $\gamma = 0.95$
7. **Similarity Transform:** Fertility Rate (Finland) = {Low}, γ = high

The systems don't know anything about the fertility rate in Finland however it knows that the population growth rate affects the fertility rate and that Norway and Finland are very similar in accordance to this criterion. The fertility rate in Norway is 1.78 children born/woman with 95% confidence (it has been just an estimation in 2004 hence the certainty in the statement is less than 1). The reasoning engine then compares this quantity with other countries' fertility rate to get a fuzzy value because the question implies the user wants such an answer. Finally, the Similarity Transform concludes that Finland has a low fertility rate just like Norway.

***Question 4**:* Is Luxembourg a wealthy nation?
Plausible Question: Is Wealthy (Luxembourg) = {?}
Answer: True
Certainty: 83%
System's Justification:

1. GDP per capita (Country) = {High} => Is Wealthy (Country) = {True}, $\gamma = 1.0$, $\alpha = 0.90$
2. Luxembourg **ISA** Country, $\gamma = 1.0$
3. GDP per capita (Luxembourg) = {$55,100}, $\gamma = 0.93$
4. GDP per capita (Luxembourg) = {High}, $\gamma = 0.93$
5. **Derivation from Implication:** Is Wealthy (Luxembourg) = {True}, $\gamma = 0.83$

A rule (implication) that connects wealth of a nation with its GDP per capita has been found. It is checked whether Luxembourg meets the precondition of the rule. A comparison with other nations' GDP per capita turns out the rule is applicable so the consequence is valid.

***Question 5**:* What is the climate like in the Netherlands?
Plausible Question: Climate (Netherlands) = {?}
Answer: Temperate
Certainty: 74%
System's Justification:

1. Latitude (Place) → Climate (Place), $\alpha = 0.8$
2. Country **ISA** Place, $\gamma = 1.0$
3. **Specification Transform:** Latitude (Country) → Climate (Country), $\alpha = 0.8$
4. Netherlands **ISA** Country, $\gamma = 100\%$
5. England **ISA** Country, $\gamma = 100\%$
6. Latitude (England) = {54 00 North}, $\gamma = 100\%$
7. Latitude (Netherlands) = {52 30 North}, $\gamma = 100\%$
8. "54 00 North" **SIM** "52 30 North", CX: Coordinates, $\sigma = 0.98$
9. Climate (England) = {Temperate}, $\gamma = 100\%$
10. **Dependency-based Analogy:** Climate (Netherlands) = {Temperate}, $\gamma = 74\%$

The system has to do a chaining reasoning for this question. It fails to retrieve the answer directly from the KB hence tries to find a concept that affect the climate. We have such a dependency in our KB which says the climate of an area is dependent to its latitude. The reasoning process first does a specification transform to make this dependency applicable for countries too. It discovers that Netherlands and England have very similar latitudes (the system knows how to compute similarities between quantities of some particular types such as coordinates and numbers. The computation takes place online). In the KB, the climate of England has been specified as temperate. A Dependency-based Analogy concludes that Netherlands has the temperate climate too. The certainty has dropped noticeably because of the prolix chained inferences.

Question 6*:* What are the prospective markets in Japan?
Plausible Question: Prospective Market (Japan) = {?}
Answer: Medical device for elderly people
Certainty: 68%
System's Justification:

1. 64 years old population and over (Country) = {High} => Faces (Country) = {Aging Problem}, $\alpha = 0.90$
2. Faces (Country) = {Aging Problem} => Prospective Market (Country) = {Medical device for elderly people}, $\alpha = 0.80$
3. **Transitivity Inference:** 64 years old population and over (Country) = {high} => Prospective Market (Country) = {Medical device for elderly people}, $\alpha = 0.72$
4. Japan **ISA** Country, $\gamma = 100\%$
5. 64 years old population and over (Japan) = {19%}, $\gamma = 95\%$
6. 64 years old population and over (Japan) = {High}, $\gamma = 95\%$
7. **Derivation from Implication:** Prospective Market (Japan) = {Medical device for elderly people}, $\gamma = 0.68$

Here, the system tries to do market prediction. It combines two implications using Transitivity Inference then applies derivation from Implication to figure out what might be the prospective market in Japan. The certainty is low because of the length of the chaining reasoning and the uncertain information about the Japan's population who are 64 years old an over (estimate in 2004).

IX. Conclusion & Future Work

In this paper, we presented a knowledge-based question answering system that replies questions with Plausible Answers. In the core of our reasoning engine, we have an implementation of the theory of Human Plausible Reasoning that tries to deduce reasonable answers when explicit information doesn't exist in the KB. The experiments show promising results after adding the required relations by HPR (dependencies, implications and hierarchical relations ISA and SIM) to CWFB.

Human beings learn pieces of knowledge from different sources and they can combine and use them simultaneously. When reasoning, one may utilize domain-specific knowledge, common sense, and lexical understandings. Adding such knowledge to QA systems would boost the performance greatly. However, there are not clear boundaries among these sources of knowledge. When "axiomatizing" domain-specific information, it is inevitable to put some commonsense knowledge to alleviate the brittleness of knowledge-based systems [23]. Paraphrasing used extensively by people is another problem for machines' comprehension about the same concepts expressed in different ways. Employing lexical knowledge may help surmount this barrier. Currently, there are good sources of commonsense and lexical knowledge available such as OpenCyc [24], ThoughtTreasure [25], ConceptNet [26] and WordNet [4]. For future, we aim to apply HPR to a very specific domain (more specific than CWFB) in conjunction with these resources to find out how HPR can imitate humans' question answering abilities in the presence of a rich KB.

Acknowledgment

This project is supported in part by Iran Telecom Research Center. We'd like to express our gratitude to Omid Asghari for his help in understanding and exploiting several economic concepts used in this paper.

References

[1] A.M. Turing, "Computing Machinery and Intelligence," *Mind* 49: 433-460, 1950.
[2] V.M. Vorhees, "Overview of the TREC 2002 Question Answering Track," *NIST Special Publication: SP 500-251*, *The Eleventh Text Retrieval Conference (TREC 2002).*
[3] D. Moldovan, S. Harabagio, R. Girju, P. Morarescu, F. Lacatsu, A. Novischi, et al. "LCC Tools for Question Answering," *NIST Special Publication: SP 500-251*, *The Eleventh Text Retrieval Conference (TREC 2002).*
[4] C.D. Fellbaum (Ed.), *WordNet: An Electronic Lexical Database*, The MIT Press, Cambridge, London, England.
[5] A. Collins, R. Michalski, "The logic of Plausible Reasoning A core theory*," Cognitive Science*, vol. 13, pp. 1-49, 1989.
[6] Kejitan Dontas, Maria Zemankova, "APPLAUSE: An implementation of the Collins-Michalski theory of plausible reasoning," *Information Science*, 52(2): 111-139, 1990.
[7] J.D. Kelly, *PRS: A System for Plausible Reasoning*, Master's Thesis, University of Illinois, Urbana-Champaign, 1989.
[8] F. Oroumchian, R.N. Oddy, "An Application of Plausible Reasoning to Information Retrieval," *SIGIR 1996*: 244-252.
[9] F. Oroumchian, B.N. Araabi, E. Ashoori, "An Application of Plausible Reasoning and Dempster-Shafer Theory of Evidence in Information Retrieval," *FSKD 2002*.
[10] F. Oroumchian, B. Khandzah, "Modeling an Intelligent Tutoring System by Plausible Inferences," *FSKD 2002*: 529-533.
[11] A. Jalali, F. Oroumchian, "An Evaluation of Document Clustering by means of Plausible Inferences," *International Journal of Computational Intelligence*, 2004.
[12] M. Virvou, K. Kabassi, "IFM: An Intelligent Graphical User Interface Offering Advice*," In proceeding of 2nd Hellenic Conf. of AI*, SETN 2002, Greece, Companion Volume, pp. 155-164.
[13] K. Kabassi, M. Virvou, "Combination of a Cognitive Theory with the Multi-Attribute Utility Theory," In V. Palade, R. J. Howlett, L. Jain (eds.): *Knowledge-Based Intelligent Information and Engineering Systems* – KES 2003, Lecture Notes in Artificial Intelligence, Vol. 2773, Springer, Berlin, Part I, pp. 944-950.

[14] A. Collins, “Fragments of a Theory of Human Plausible Reasoning,” *Theoretical Issues in Natural Language Processing II*, pp. 194-201, 1978.

[15] R.S. Michalski, “Variable-Valued Logic and Its Applications to Pattern Recognition and Machine Learning,” *Computer Science and Multiple Valued Logic Theory and Applications*, D.C. Rine (Ed.), Amsterdam: North-Holland, 506-534, 1975.

[16] R.S. Michalski, K. Dontas, D. Boehm-Davis, “Plausible Reasoning: An Outline of Theory and Experiments,” *Proceedings of the Fourth International Symposium on Methodologies for Intelligent Systems*, pp. 17-19, Charlotte, NC, October 1989.

[17] Milton Friedman, *Monetarist Economics*, Cambridge: Basil Blackwell, 1991.

[18] http://www.cia.gov/cia/publications/factbook/ visited on October 26, 2004.

[19] J.F. Sowa, "Conceptual graphs for a database interface," *IBM Journal of Research and Development*, vol. 20, no. 4, pp. 336-357, 1976.

[20] http://www.cs.umd.edu/projects/plus/Parka/parka-db.html visited on October 26, 2004.

[21] http://www.daml.org/2003/09/factbook/ visited on October 26, 2004.

[22] http://sourceforge.net/projects/cwfb-pl/ visited on October 26, 2004.

[23] Doug Lenat, Mayank Prakash, Mary Shepherd, “CYC: Using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks,” *AI Magazine*, Volume 6, Issue 4, Winter 1986, American Association for Artificial Intelligence, Menlo Park, CA, USA.

[24] http://www.opencyc.org/ visited on October 26, 2004.

[25] http://www.signiform.com/tt/htm/overview.htm visited on October 26, 2004.

[26] Push Singh, Barbara Barry, and Hugo Liu., “Teaching Machines about Everyday Life,” *BT Technology Journal* 22(4). pp. 227-240, 2004, Kluwer Academic Publishers.