



Ontology: A Case for Disease and Drug Knowledge Discovery

Onuri Ernest E.
Department of
Health Informatics,
School of Health Professions,
Rutgers Biomedical and
Health Sciences
(RBHS) of New Jersey,
USA

Oyindolapo Komolafe
Department of
Computer Science
Babcock University
Ilishan-Remo
Ogun State, Nigeria.

Shade O. Kuyoro
Department of
Computer Science
Babcock University
Ilishan-Remo
Ogun State, Nigeria.

Awodele Oludele
Department of
Computer Science
Babcock University
Ilishan-Remo
Ogun State, Nigeria.

ABSTRACT

A number of medical conditions are still incurable today, some are even auto-immune. When people come down with disease conditions like lupus, Alzheimer, multiple sclerosis, schizophrenia, diabetes, cancer, asthma, Creutzfeldt-Jakob, AIDS and a host of others, they become incapacitated, lose major functions and most just wait until they die. These conditions are usually very cruel to those who suffer them. The lack of cure for these conditions is partly due to the fact that their causative agents are not clearly known or understood. One might be tempted to presume that with the completion of the Human Genome Project, solutions would have been derived for such disease conditions. It is only wise to think that such conditions have so far remained major challenges to medical researchers because they have multiple causes. Ontology, a widely accepted Knowledge Representation (KR) paradigm is therefore proposed as a KR technique to holistically attempt to address the gaps by first identifying all the causative elements, and then being able to proffer viable solutions to such conditions.

General Terms

Medical Informatics, Artificial Intelligence, Knowledge Representation, Intelligent Systems, Genomics

Keywords

Ontology, Expert System, Drug Discovery, Disease, Prediction

1. INTRODUCTION

Engaging in wholesome interactions and understanding language in any form requires understanding connections that exist between words, phrases, concepts and thoughts in a meaningful way. Many of the problems faced today in Artificial Intelligence (AI) depend in some way in understanding this network of relationships which represent the facts known and perceived about the world. AI is the mechanization of tasks requiring intelligence. Basically, AI is

a multi-disciplinary science that combines the fields of computer science, mathematics, philosophy, psychology and linguistics in an effort to make machines think and behave like humans. AI, a field predominantly domiciled in computer science and engineering is concerned with the computational understanding of intelligent behavior, as well as the creation of artifacts that display such behavior [1]. However, one of the essential challenges of AI is the computerized or systematic acquisition of knowledge from data along with the representation of this knowledge to support reasoning and understanding [2]. Taking a quick look at the backbone paradigms of AI such as expert systems, data mining, neural networks, machine learning, natural language processing, genetic algorithm, computer vision and emerging paradigms like semantic web, bioinformatics and recognition systems, one thing is for sure; in the midst of the accumulation of large amounts of data, lies the imperative need for knowledge extraction in a meaningful way such that similar concepts can be mapped together. The approach of choice in carrying KR goes a long way in determining the efficacy of the AI system in view. Suffice it to say that the goal of this research paper is to show the viability of using ontology to drive knowledge discovery for diseases that have no known cure today especially because they have multiple causative agents.

2. KNOWLEDGE-BASED SYSTEMS

A knowledge-based System (KBS) is a problem solving and decision making system, based on knowledge of its task and logical rules as well as procedures for using knowledge [3]. Knowledge engineering which is the process of designing an expert system as depicted in figure 1, is predicated on the domain knowledge of a human expert. It has three stages [4]:

- i. Knowledge Acquisition (KA): this is the method of obtaining the knowledge from human expert(s) through interviews, observation and/or structured research so as to enable feature extraction. KA requires the ability to focus attention on the relevant features of the problem. In addition, any KA

approach entails the ability to learn rules relating the selected features to the categories of interest, and ability to acquire knowledge incrementally and detect when knowledge is changing [2]. The first step of KA is the collection of data and selection of the features to represent the data. The use of collaborative filtering and multi-agent systems can be used as a method of acquiring new knowledge.

- ii. Knowledge Representation (KR): this is a subarea of AI implied with understanding, designing and implementing ways of representing information in computers so that programs (agents) can use the information to derive information that is implied by

it, to converse with people in natural languages, to decide what to do next, to plan future activities and to solve problems in areas that normally require human expertise, as well as solve problems that humans find complicated and unable to solve. It involves selecting the most appropriate structures to represent the knowledge (lists, sets, scripts, decision trees, object-attribute value triplets etc.). There are two basic components of KR; reasoning and inference.

- iii. Knowledge Validation (KV): this involves testing that the knowledge of the expert system is correct and complete.

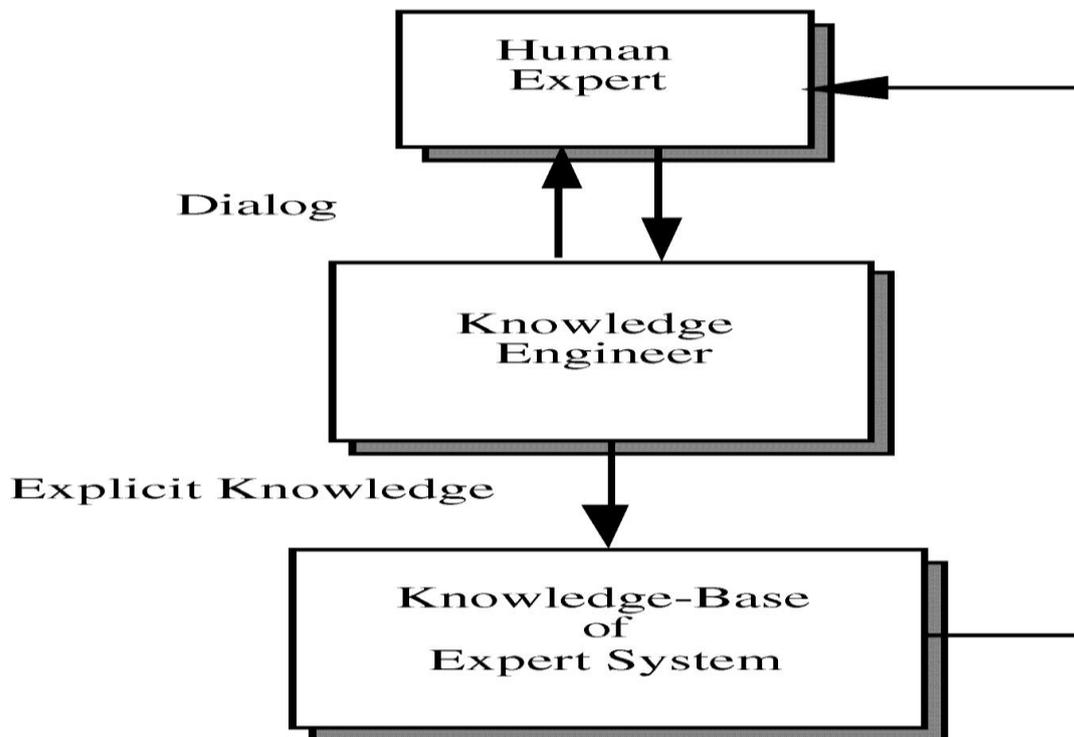


Fig 1: Development of and Expert System

There are various types of knowledge:

- i. Declarative: it defines what is known about a problem. This includes simple statements which are asserted to be either true or false.
- ii. Procedural: this defines how a problem is solved. It contains rules, strategies, agenda and procedures.
- iii. Heuristic: it describes a rule-of-thumb that helps to guide the reasoning process.
- iv. Meta knowledge: describes knowledge about knowledge to enhance the efficiency of problem solving.
- v. Structural Knowledge: it offers descriptive insight into knowledge structures. In addition, it contains rule sets, concept relationships and concept to object relationships [5].
- vi. Factual Knowledge: this form of knowledge is explicit and verifiable through experiments and formal methods. It is usually well documented in textbooks and manuals.
- vii. Tacit Knowledge: it is implicit, unconscious knowledge that can be difficult to express in words or other forms of representations.
- viii. Priori/Prior knowledge: it is independent of the experience or empirical evidence e.g. “everybody born before 1999 is older than 15 years”
- ix. Posteriori/Posterior knowledge: dependent of experience or empirical evidence, such as “X was born in 1990” [3].

2.1.Fundamental Roles of KR

KR basically underlies a set of ontological commitments, that is, an answer to the question; in what terms should I think

about the world? It is a fragmentary theory of intelligent reasoning, expressed in terms of three components:

- i. the representation's fundamental conception of intelligent reasoning;
- ii. the set of inferences the representation sanctions; and
- iii. The set of inferences it recommends.

KR is a medium for pragmatically efficient computation, i.e., the computational environment in which thinking is accomplished. One contribution to this pragmatic efficiency is supplied by the guidance a representation provides for organizing information so as to facilitate making the recommended inferences. Furthermore, it is a medium of human expression, a language in which one says things about the world [6]. KR faces issues which include resolution details, scope, modularity, understandability, explicit versus implicit knowledge as well as procedural versus declarative knowledge.

2.2. Evolution of Knowledge Representation

KR has a long history in AI research, philosophy and declarative thought. AI turned to KR after the failure of early attempts at machine translations, which were based to a large extent on simple dictionary look-up to teach a computer to understand natural language [7]. As depicted in figure 2, KR has its roots in the design of formal representation languages of the 1980s. In the early days, researchers initiated methods like heuristic questions & answers, neural networks and theorem proving [8]. Translation is not purely a map from lexical item to lexical item, and ambiguity is inherent in a large number of utterances, hence a means is required to encode what the actual meaning of the sentence is. Therefore, there must be a representation of meaning independent of the words used for example in medical diagnosis, the systems MYCIN and DENDRAL written in LISP at Stanford University, could identify bacteria and propose a therapy. During this period, a number of KR-friendly languages like PROLOG and LISP were designed to express the content of natural language statements [9] and commonsense reasoning [10] as directly as possible.

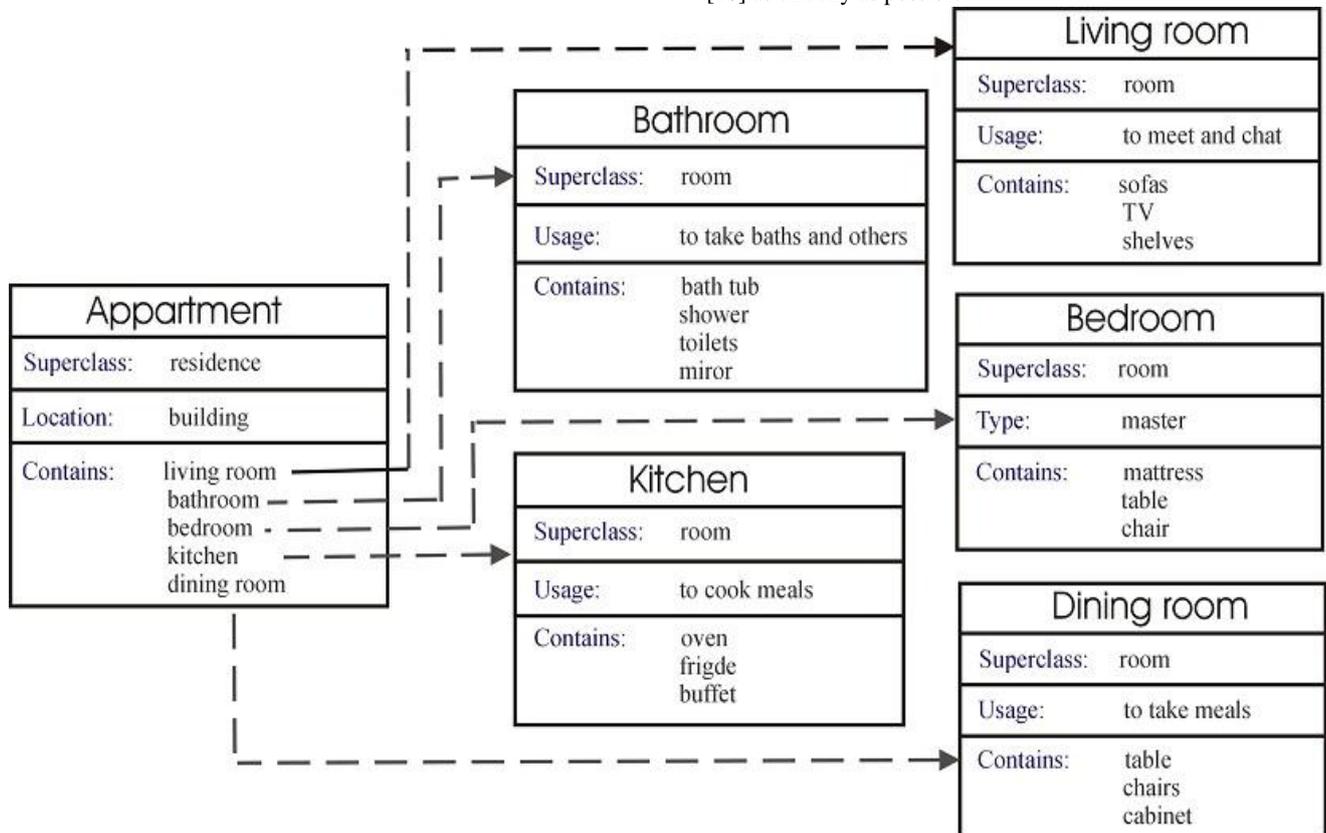


Fig 2: Representation of data, information, knowledge and wisdom [8]

2.3. Knowledge Representation Techniques

AI needs to represent things like objects, properties, categories and relationships between objects, situations, events, states and time, causes and effects, KR provides the way to represent them [11]. KR techniques are divided into two major categories, which are declarative representation and procedural representation. The declarative representation techniques are used to represent objects, facts, relations while the procedural representation are used to represent the action performed by the objects. Some of the techniques for KR are semantic nets, frames, ontology and hybrid representation. However, discussion will be limited to just ontology.

2.3.1. Ontology

An ontology is a formal and explicit specification of a shared conceptualization of a domain of interest. Ontology is a branch of metaphysics and denotes the philosophical investigation of existence. It is concerned with the fundamental questions of “what is being?” and “what kinds of things are there?” [12]. The definition of ontology captures several characteristics of the concept as a specification of domain knowledge, namely the aspects of formality, explicitness, being shared, conceptuality and domain-specificity.

- i. **Formality:** ontology is expressed in a KR language that provides formal semantics. This ensures that the specification of domain knowledge in an ontology can be processed by machines and can be interpreted in a well-defined way.
- ii. **Explicitness:** ontology states knowledge in a clear and definite way to make understanding easy for machines. Notions that are not explicitly included in the ontology are not part of the machine-

interpretable conceptualization it captures, although humans might take them for granted by common sense.

- iii. **Being shared:** an ontology reflects an agreement on a domain conceptualization among people in a community. The larger the community, the more difficult it is to come to an agreement on sharing the same conceptualization. Thus, an ontology is always limited to a particular group of people in a community, and its construction is associated with a social process of reaching consensus.
- iv. **Conceptuality:** an ontology specifies knowledge in a conceptual way in terms of symbols that represent percepts and their relations. The concepts and relations in an ontology can be intuitively grasped by humans, as they correspond to the elements in our mental model. Moreover, an ontology describes a conceptualization in general terms and does not only capture a particular state of affairs. Instead of making statements about a specific situation involving particular individuals, an ontology tries to cover as many situations as possible that can potentially occur [13].
- v. **Domain specificity:** specifications in an ontology are limited to knowledge about a particular domain of interest. The narrower the scope of the domain for the ontology, the more an ontology engineer can focus on axiomatising the details in this domain, rather than covering a broad range of related topics. In this way, the explicit specification of domain knowledge can be modularized and expressed using several different ontologies with separate domains of interest [14].

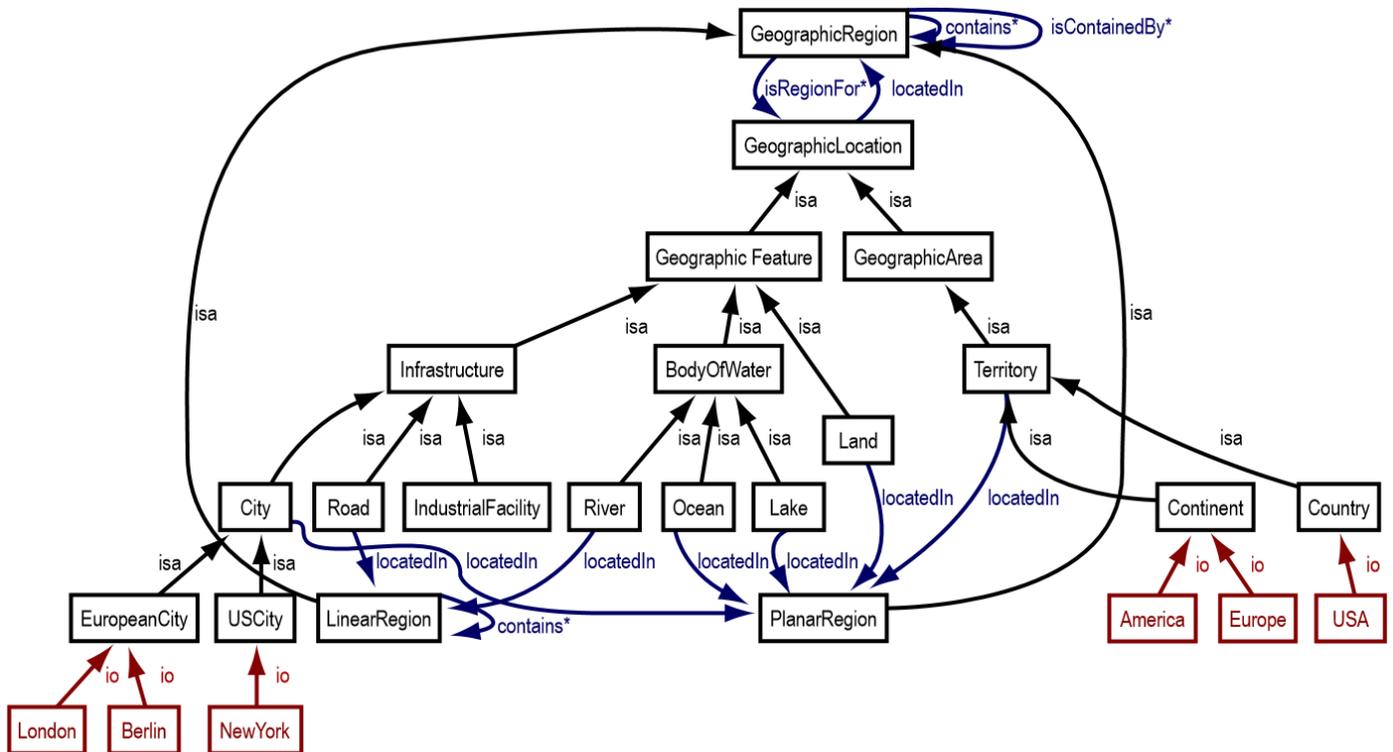


Fig 3: Example of a geographic ontology [14]

Not all the information in an ontology can easily be visualized in a graph. As shown in figure 3, ontology engineering environments provide extra means for displaying and editing complex axiomatic information, using a special-purpose ontology language or logical formal notation. The computational domain model of an ontology can be used for various purposes, depending on the application scenario. Furthermore, figure 4 is an ontological depiction of the main

types of diabetes highlighting types of causative chains and goes to show that a relationship exists between the various sub-types of diabetes. In addition, figure 5 is an example of a Resource Description Framework (RDF) representation of diseases which also elucidates inherent component relationships whereas figure 6 specifies the relationships between the components of a medical history questionnaire.

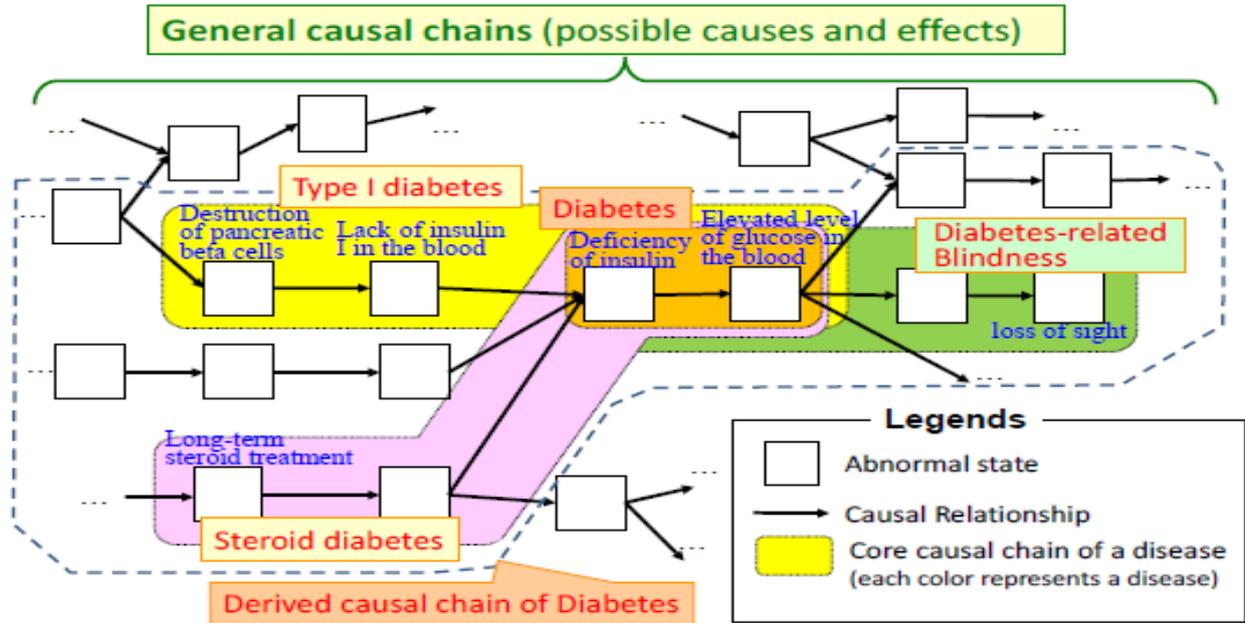


Fig 4: types of diabetes constituted of causal chains [15]

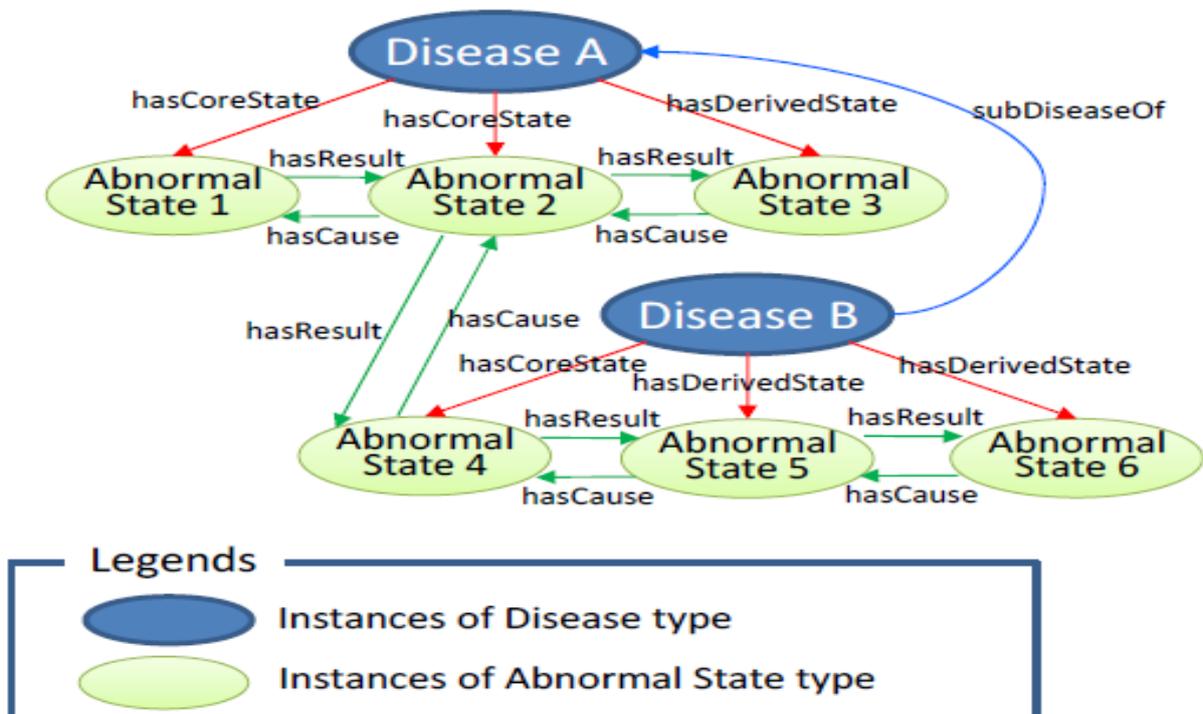


Fig 5: An RDF representation of diseases [15]

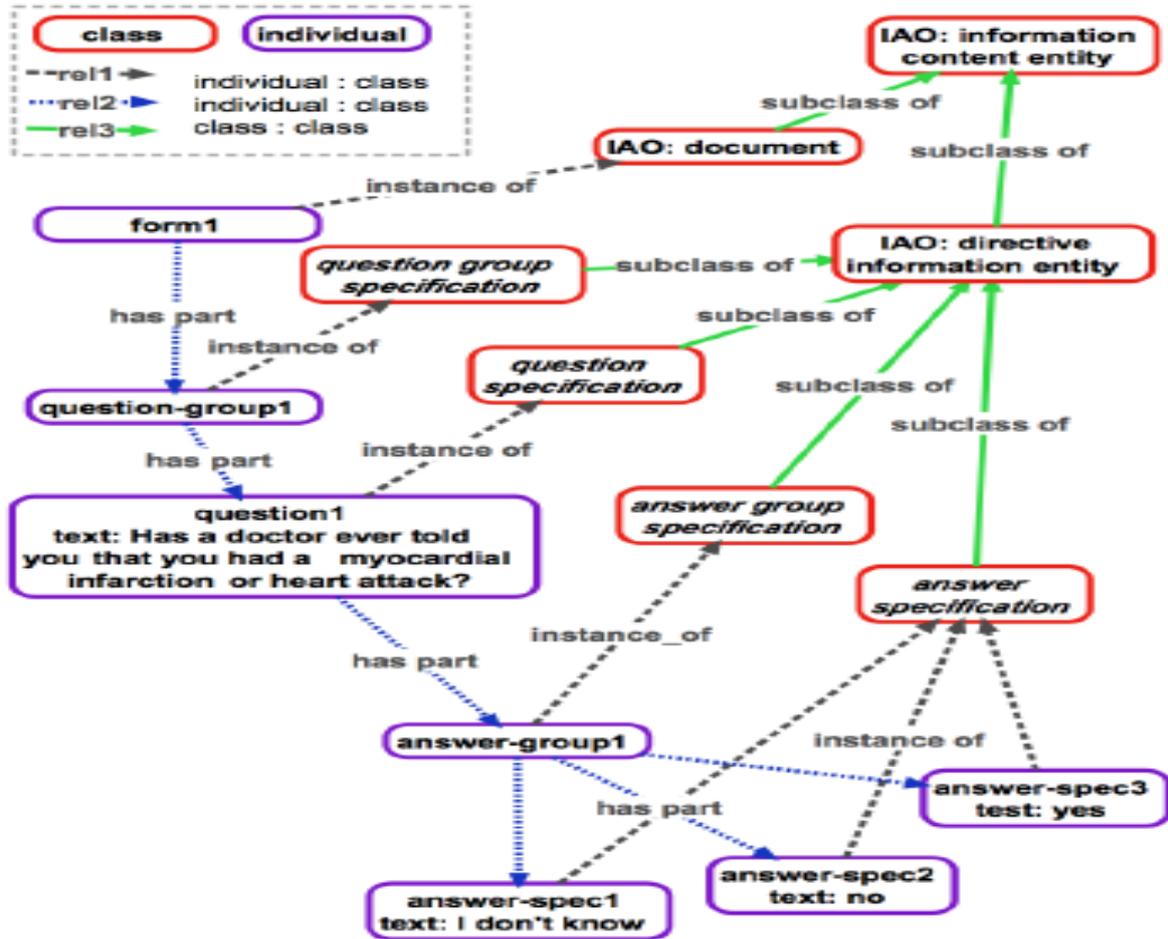


Fig 6: Representation scheme specifying questionnaire component [16]

3. APPLICATIONS OF KR

KR techniques were applied mainly to extend database properties and capabilities in the beginning. Then, as research matured, several sophisticated applications such as KBSs, Expert Systems, and Intelligent Decision Support Systems emerged. The propaganda surrounding KR applications was so high on the international scene in the early 1980s that Edward Feigenbaum made in his book on the Japanese Fifth Generation Computer project, a remarkable prediction about the power, the knowledge and capability to serve humans in medical diagnosis, product design and education.

Among the first really successful applications of KR are Expert Systems which are software designed to emulate the decision making, the thinking process of a human expert. An Expert System relies on symbolic languages like LISP and PROLOG, uses an inference engine (set of rules) to reason about a rich knowledge base, draws conclusions and then makes recommendations to humans. Several Fortune 500 companies used ES in the 1980s as practical tools to solve concrete problems. Medical diagnosis (e.g., Mycin by Shortliffe and associates) was a major application area where the first Medical Expert Systems appeared. In addition, KR researchers and engineers contributed to the creation of a plethora of software based on KR technology in education systems (intelligent tutoring), quality assurance, medical imaging and drug administration.

Other AI paradigms that rely heavily on KR include Artificial Neural Networks, data mining, bioinformatics, natural language processing, genetic algorithm, semantic web, machine learning. As a matter of fact, the efficacy and efficiency of these range of predictive systems predictive systems are hugely dependent on how well the knowledge has been represented whether by ontology or other KR techniques [17].

Their applications are also widespread:

- i. Spam filtering
- ii. Precision medicine
- iii. Self-driven cars
- iv. Drones
- v. Weather forecasting
- vi. Disease diagnosis
- vii. Intelligent tutoring systems
- viii. Search engines
- ix. Semantic web
- x. Wolfram Alpha's computational engine



4. DISCUSSION

KR is the area of AI concerned with how knowledge can be represented symbolically and manipulated in an automated way by reasoning programs. It is at the very core of a radical idea about how to understand intelligence: instead of trying to understand or build brains from the bottom-up, attempt is made to try to understand or build intelligent behavior from the top-down. In particular, the question of what an agent would need to know in order to behave intelligently, and what computational mechanisms could allow this knowledge to be made available to the agent as required. KR entails the study of how what is known can at the same time be represented as comprehensibly as possible and reasoned with as effectively as possibly [18,19].

The proposition in this discourse is based on viable research works [15,16] that have adopted ontology as the KR technique required to design and implement complex biomedical systems. Drug discovery is no doubt a complex process which entails the following phases:

- i. Drug discovery also referred to as drug design which starts with target identification, also referred to as translational bioinformatics. A definite target which is also akin to the disease is identified and validated. Target validation implies ascertaining if a modification to the function of the target could have therapeutic effect.
- ii. Lead identification which is the next phase aims at compound identification which also doubles as the starting point for refinement. The approaches to the compound identification is more measured and scientific when compared to the traditional approach.
- iii. The next phase, lead optimization entails fine-tuning the lead compound to elicit greater drug activity as well as eliminate undesirable side-effects. This is carried out using either the Structure Based Drug Design (SBDD) or the Quantitative Structure Activity Relationship (QSAR).
- iv. The fundamental difference between the traditional method and the rational design is the concept of translational bioinformatics which is key to this new method of drug design. This makes the drug discovery process more intentional and tailored toward the production of effective therapeutic interventions. It also ensures reduction in the time frame devoted to the process.

Drug discovery has become a major component of bioinformatics, a field that ensures the interleaving of computing and biosciences. However, the predictive component of computing inherent in the computing/statistical tools, range from machine learning, to genetic algorithm and also include artificial neural networks and data mining. All these find their bearing majorly in AI and this in turn is heavily reliant on KR.

Bioinformatics entails the study and processing of biological data especially the information content of genomes. It majorly deals with aiding sequencing and annotations of genomes and their observed mutations. Some proponents argue that bioinformatics is the creation of tools (algorithms and

databases) that solve problem. The goal is perceived to be that of building useful tools that work on biological data (it is about the engineering). Bioinformatics tools became necessary especially in the last 10 to 15 years because;

- i. With the conclusion of the Human Genome Project in 2003,
- ii. Data grew bigger. The human genome for instance has about 3 billion genes and some other organisms like some big trees have larger genomes.
- iii. In light of the data explosion it became difficult to manually carry out analysis
- iv. Predictive analysis became necessary and what better way to do this efficiently than with tools made available by bioinformatics.

The roles of genomics, transcriptomics and proteomics in drug discovery is that they provide a sequential mapping of how the DNA is transcribed to messenger RNA which goes on to prompt the translation process required for the manufacture of the structural proteins.

In view of the foregoing, it becomes imperative that the knowledge of these concepts are necessary in drug discovery and development, due to the fact that changes might occur during either or both transcription or translation activities leading to disease. Hence, this knowledge becomes relevant in identifying and isolating the target responsible for the disease.

Furthermore, because the cells in the body continuously undergoes these processes due to the fact that proteins are essential in the structure and function of living cells, as well as in replacing worn out tissues, one wrong transcription/translation error and disease may ensue. If the correction mechanism of gene replication is incapable of fixing the error, then there is the chance for the error to perpetuate and continue to cause problems, unless a clinically proven intervention is introduced to cause a reversal of the disease condition.

An understanding of genomics, transcriptomics and proteomics is necessary to make this intervention vis-a-vis drug discovery development possible. In line with the problem statement which entails the painstaking efforts, yet a lack of results capable of nipping some terminal diseases in the bud through drug discovery, mediated by bioinformatics mechanisms. Hence, the proposition that ontology be adopted as a KR technique to model systems that embody components of these diseases in a view to systematically identifying their causative agents triggering such deadly conditions.

In addition, it is worth noting that the knowledge-base is responsible for storing factual and heuristic knowledge using one or more KR schemes to express knowledge about the application domain of interest. The inference engine is the mechanism that is used to manipulate symbolic information and knowledge so as to solve problems through reasoning. The inference engine determines which rule antecedent are satisfied by the fact. It consists of formulas to combine the rules with datasets in the knowledge-base. Inference engine can be forward chaining, backward chaining or a combination of both. The working storage houses the data that is specific to the problem being solved. The individuals who interacts with the system includes the domain expert (the human expert whose domain knowledge is being sought), knowledge



engineer (the one who encodes the expert's knowledge in a declarative form that can be used by the expert system), system engineer (the one who builds the user interface, designs the declarative format of the knowledge-base and implements the inference engine) and the user (the one who will consult the system to get advice as would have been provided by the expert) [19].

In view of the foregoing, advances have been seen in the workings of high-level robotics with machine learning, genetic algorithm and neural network mechanisms embedded into them to bring about the finesse evident only in the man. There are now robots that come close in replicating the decision making capabilities of mankind. Intelligent tutoring systems now teach as well as generate examination questions with varying degrees of difficulty that is dependent on the ability of the student. It can be seen that the deployment of KR in intelligent intrusion detection and prevention systems that proactively prevent unauthorized breaching of systems. Furthermore, predictive systems implemented in complex disciplines like bioinformatics in gene therapy and sequence analysis now exist. Also, this is seen in systems implemented to determine the influence of family background factors on student academic performance. KR makes possible data mining and knowledge discovery which has influenced the emergence of computational engines such as wolfram alpha and revolutionized search. The list is indeed endless.

5. CONCLUSION

Suffice it to say that KR is at the heart of intelligent systems. KR as a major feature of AI systems is paramount in that it embodies the ability for intelligent systems to optimally carry out expected tasks which have ties to one form of AI mechanism or the other and ensure that decision reached by such systems tend towards what is obtainable in reality. It is therefore expected, that cures for some of the terminal diseases highlighted in this discourse will soon be found if ontological models evolve steadily to capture in a robust way the underlying components likely inherent in the multiple causative agents that trigger such conditions. It is only when such are known and clearly understood that viable drugs can be engineered using already discussed methods.

6. REFERENCES

- [1] Stuart C.S. (2010): Knowledge Representation and Reasoning Logics for Artificial Intelligence. Department of Computer Science and Engineering and center for Cognitive Science, University at Buffalo, The state university of New York.
- [2] Holder L.B, Markov Z. and Russell I (2006): advances in knowledge acquisition and representation. International journal on Artificial Intelligence Tools Vol 15, No 6 867-874
- [3] Rajeswari P.V.N. and Prasad T.V. (2012): Hybrid systems for knowledge representation in artificial intelligence. IJARAI international journal of advanced research in artificial intelligence Vol 1, No 8
- [4] Belavkin R.V (2002): knowledge Representation, a lecture note in Computer science BIS3226
- [5] Habil. S. (2008):, "Knowledge representation and networked schemes", a Lecture notes of Fundamentals of artificial intelligence
- [6] Okafor E.C and Osuagwu C.C (2007): Issues in Structuring the Knowledge-base of Expert Systems Electronic Journal of Knowledge Management, Vol. 5 Issue 3
- [7] Briggs R. (1985): Knowledge Representation in Sanskrit and Artificial Intelligence, RIACS, NASA Ames Research Center, Moffet Field, California. AI Magazine volume 6 number 1.
- [8] Kone M.T (2014): knowledge representation. The future of human Evolution retrieved from <http://futurehumanevolution.com/artificial-intelligence-future-human-evolution/knowledge-representation> on November 13, 2014.
- [9] Helbig H, (2006): *Knowledge Representation and the Semantics of Natural Language*, Springer, Berlin, Heidelberg, New York.
- [10] Davis, E. (1990). *Representations of Common-Sense Knowledge*. Morgan Kaufmann.
- [11] Tanwar P., Prasad T.V., Aswal M. S (2010). "Comparative Study of Three Declarative Knowledge Representation Techniques", Int. J. on Comp. Sc. and Engg, Vol. 02, No. 07
- [12] Craig E.(1998). Ontology. In E. Craig, editor, *Routledge Encyclopedia of Philosophy*, pages 117–118. Routledge, New York, 1998.
- [13] Guarino N (1998): . Formal Ontology and Information Systems, Preface. In N. Guarino, editor, *Proceedings of the 1st International Conference on Formal Ontologies in Information Systems, FOIS-98, Trento, Italy*, pages 3–15. IOS Press, 1998.
- [14] Grimm S, Hitzler P and Abecker A (2007): knowledge representation and ontologies; logic, ontologies and semantics web languages.
- [15] Kozaki K, Yamagata Y, Mizoguchi R, Imai T and Ohe K, (2015). Disease Compass --Navigation System for Disease Knowledge based on Ontology and Linked Data Techniques. ICBO2015, International Conference on Biomedical Ontology 2015, Proceedings of the Main Conference July, 27-30, Lisbon, Portugal.
- [16] Bona J, Kohn G and Ruttenberg A (2015). Ontology-driven Patient History Questionnaires. ICBO2015, International Conference on Biomedical Ontology 2015, Proceedings of the Main Conference July, 27-30, Lisbon, Portugal.
- [17] Onuiri E.E, Awodele O. and Idowu S.A (2014), "Framework for Knowledge-Based Intelligent Clinical Decisionsupport to Predict Comorbidity" International Journal of Advanced Research in Artificial Intelligence(IJARAI), 3(4), 2014. <http://dx.doi.org/10.14569/IJARAI.2014.030402>
- [18] Brachman, R.J., & Levesque, H.J., (2003). Knowledge Representation and Reasoning.
- [19] Chakraborty R.C. (2011). Expert Systems: AI Course Lecture 35 – 36, notes, slides. www.myreaders.info/html/artificial_intelligence.html. Retrieved November 12, 2014.