From Search Engines to Question-Answering Systems—The Need For New Tools

Lotfi A. Zadeh*

Extended Abstract

Search engines, with Google at the top, have many remarkable capabilities. But what is not among them is the deduction capability—the capability to synthesize an answer to a query by drawing on bodies of information which are resident in various parts of the knowledge base. It is this capability that differentiates a question-answering system, Q/A system for short, from a search engine.

Construction of Q/A systems has a long history in AI. Interest in Q/A systems peaked in the seventies and eighties, and began to decline when it became obvious that the available tools were not adequate for construction of systems having significant question-answering capabilities. However, Q/A systems in the form of domain-restricted expert systems have proved to be of value, and are growing in versatility, visibility and importance.

Search engines as we know them today owe their existence and capabilities to the advent of the Web. A typical search engine is not designed to come up with answers to queries exemplified by "How many Ph.D. degrees in computer science were granted by Princeton University in 1996?" or "What is the name and affiliation of the leading eye surgeon in Boston?" or "What is the age of the oldest son of the President of Finland?" or "What is the fastest way of getting from Paris to London?"

Upgrading a search engine to a Q/A system is a complex, effort-intensive, open-ended problem. Semantic Web and related systems for upgrading quality of search may be viewed as steps in this direction. But what may be argued, as is done in the following, is that existing tools, based as they are on bivalent logic and probability theory, have intrinsic limitations. The principal obstacle is the nature of world knowledge.

The centrality of world knowledge in human cognition, and especially in reasoning and decision-making, has long been recognized in AI. The Cyc system of Douglas Lenat is a repository of world knowledge. The problem is that much of world knowledge consists of perceptions. Reflecting the bounded ability of sensory organs, and ultimately the brain, to resolve detail and store information, perceptions are intrinsically imprecise. More specifically, perceptions are f-granular in the sense that (a) the boundaries of perceived classes are fuzzy; and (b) the perceived values of attributes are

* Professor in the Graduate School and Director, Berkeley initiative in Soft Computing (BISC), Computer Science Division and the Electronics Research Laboratory, Department of EECS, University of California, Berkeley, CA 94720-1776; Telephone: 510-642-4959; Fax: 510-642-1712;
E-Mail: zadeh@cs.berkeley.edu . Research supported in part by ONR N00014-00-1-0621, ONR Contract N00014-99-C-0298, NASA Contract NCC2-1006, NASA Grant NAC2-117, ONR Grant N00014-96-1-0556, ONR Grant FDN0014991035, ARO Grant DAAH 04-961-0341 and the BISC Program of UC Berkeley.

granular, with a granule being a clump of values drawn together by indistinguishability, similarity, proximity or functionality. What is not widely recognized is that f-granularity of perceptions put them well beyond the reach of computational bivalent-logic-based theories. For example, the meaning of a simple perception described as “Most Swedes are tall,” does not admit representation in predicate logic and/or probability theory.

Dealing with world knowledge needs new tools. A new tool which is suggested for this purpose is the fuzzy-logic-based method of computing with words and perceptions (CWP), with the understanding that perceptions are described in a natural language. A concept which plays a key role in CWP is that of Precisiated Natural Language (PNL). It is this language that is the centerpiece of our approach to reasoning and decision-making with world knowledge.

A concept which plays an essential role in PNL is that of precisiability. More specifically, a proposition, p, in a natural language, NL, is PL precisiable, or simply precisiable, if it is translatable into a mathematically well-defined language termed precisiation language, PL. Examples of precisiation languages are: the languages of propositional logic; predicate logic; modal logic; etc.; and Prolog; LISP; SQL; etc. These languages are based on bivalent logic. In the case of PNL, the precisiation language is a fuzzy-logic-based language referred to as the Generalized Constraint Language (GCL). By construction, GCL is maximally expressive.

A basic assumption underlying GCL is that, in general, the meaning of a proposition, p, in NL may be represented as a generalized constraint of the form X isr R, where X is the constrained variable; R is the constraining relation, and r is a discrete-valued variable, termed modal variable, whose values define the modality of the constraint, that is, the way in which R constrains X. The principal modalities are; possibilistic (r=blank); probabilistic (r=p); veristic (r=v); usuality (r=u); fuzzy random set (r=rs); fuzzy graph (r=fg); and Pawlak set (r=ps). In general, X, R and r are implicit in p. Thus, precisiation of p, that is, translation of p into GCL, involves explicitation of X, R and r. GCL is generated by (a) combining generalized constraints; and (b) generalized constraint propagation, which is governed by the rules of inference in fuzzy logic. The translation of p expressed as a generalized constraint is referred to as the GC-form of p, GC(p). GC(p) may be viewed as a generalization of the concept of logical form. An abstraction of the GC-form is referred to as a protoform (prototypical form) of p, and is denoted as PF(p). For example, the protoform of p: “Most Swedes are tall” is Q A’s are B’s, where A and B are labels of fuzzy sets, and Q is a fuzzy quantifier. Two propositions p and q are said to be PF-equivalent if they have identical protoforms. For example, “Most Swedes are tall,” and “Not many professors are rich,” are PF-equivalent. In effect, a protoform of p is its deep semantic structure. The protoform language, PFL, consists of protoforms of elements of GCL.

With the concepts of GC-form and protoform in place, PNL may be defined as a subset of NL which is equipped with two dictionaries: (a) from NL to GCL; and (b) from GCL to PFL. In addition, PNL is equipped with a multiagent modular deduction database, DDB, which contains rules of deduction in PFL. A simple example of a rule of deduction in PFL which is identical to the compositional rule of inference in fuzzy logic, is: if X is A and (X, Y) is B then Y is A∘B, where A∘B is the composition of A and B, defined by $\mu_B(v) = \sup_u(\mu_A(u) \wedge \mu_B(u,v))$, where μ_A and μ_B are the membership functions of A and B, respectively, and $\wedge$ is min or, more generally, a T-norm. The rules of deduction in DDB are organized into modules and submodules, with each module and submodule associated with an agent who controls execution of rules of deduction and passing results of execution.

In our approach, PNL is employed in the main to represent information in the world knowledge database (WKD). For example, the items:

- If X/Person works in Y/City then it is likely that X lives in or near Y
- If X/Person lives in Y/City then it is likely that X works in or near Y

are translated into GCL as:

Distance (Location (Residence (X/Person), Location (Work (X/Person) isu near,
where isu, read as ezoo, is the usuality constraint. The corresponding protoform is:

F (A(B(X/C), A(E(X/C)) isu G.

A concept which plays a key role in organization of world knowledge is that of an epistemic (knowledge-directed) lexicon (EL). Basically, an epistemic lexicon is a network of nodes and weighted links, with node i representing an object in the world knowledge database, and a weighted link from node i to node j representing the strength of association between i and j. The name of an object is a word or a composite word, e.g., car, passenger car or Ph.D. degree. An object is described by a relation or relations whose fields are attributes of the object. The values of an attribute may be granulated and associated with granulated probability and possibility distributions. For example, the values of a granular attribute may be labeled small, medium and large, and their probabilities may be described as low, high and low, respectively. Relations which are associated with an object serve as PNL-based descriptions of the world knowledge about the object. For example, a relation associated with an object labeled Ph.D. degree may contain attributes labeled Eligibility, Length.of.study, Granting.institution, etc. The knowledge associated with an object may be context-dependent. What should be stressed is that the concept of an epistemic lexicon is intended to be employed in representation of world knowledge—which is largely perception-based—rather than Web knowledge, which is not.

As a very simple illustration of the use of an epistemic lexicon, consider the query "How many horses received the Ph.D. degree from Princeton University in 1996." No existing search engine would come up with the correct answer, "Zero, since a horse cannot be a recipient of a Ph.D. degree." To generate the correct answer, the attribute Eligibility in the Ph.D. entry in EL should contain the condition "Human, usually over twenty years of age."

In conclusion, the main thrust of the fuzzy-logic-based approach to question-answering which is outlined in this abstract, is that to achieve significant question-answering capability it is necessary to develop methods of dealing with the reality that much of world knowledge—and especially knowledge about underlying probabilities is perception-based. Dealing with perception-based information is more complex and more effort-intensive than dealing with measurement-based information. In this instance, as in many others, complexity is the price that has to be paid to achieve superior performance.