MEDICAL EXPERT SYSTEMS

The field of medical informatics (also termed health informatics) concerns application of information science and information technology to health care, clinical care, education and biomedical research. Most countries have national societies in this area, and some 40 of them are organized in the International Medical Informatics Association (*IMIA*). Artificial intelligence (*AI*) methods specifically refer to the application of computer-based programs simulating human experts. Recent developments in medical informatics benefit from the availability of powerful personal computers (workstations), advanced information processing techniques such as the artificial neural network (*ANN*), and increased acceptance by the clinical community. The latter seemingly trivial factor should not be underestimated by engineers. In fact, wider acceptance is only partly due to improved user interfaces, but largely by the gradual recognition that computers form a useful tool in the doctor's office. In 1995, still fewer than 1% of the family practitioners in the United States use a computerized patient record, but enthusiasm is increasing. It has been shown that patient satisfaction does not decrease when the computer is employed in the physician's examination room (1). Information, in general, requires that locally available knowledge can be communicated. Indeed, facts are only meaningful if they can be uniquely described and successfully transmitted from one location or person to another. Trivial examples from everyday life concern the combination of coding of messages by writing news reports and the distribution of newspapers, and the formulation of integrated weather reports and subsequent radio broadcasting. Similar lines of communication apply to medical informatics, although the implementation of advanced techniques started not earlier than around 1975. To understand this delay that surprisingly impeded an important issue such as medical care, and also to appreciate the potential progress that can be realized, it is essential to indicate the circumstances that make health care differ from other areas in the natural sciences. First, the primary information stems from humans (or animals for the sake of veterinary informatics) inflicted with shortcomings regarding their functioning. Second, in medicine it is difficult to define what is "normal." Normality does not refer to a single numerical value, but rather to a certain range defined by reference values. Therefore it seems almost impossible to define a deviating process to begin with, then to assess the severity of any abnormality, to judge whether the defect is dangerous for your health, next to evaluate the impact of therapeutic intervention, and finally to determine the prognosis for each individual. Clearly, a vast number of communication steps are to be taken, thus limiting the efficiency of the process. Moreover, existing knowledge on a particular disease may not be immediately available to any physician, because it is an impossible task to scrutinize weekly or monthly all medical journals published anywhere in the world. Further limitations regarding communication of medical knowledge refer to clinical terminology and classification of health data. With the knowledge that medical informatics deals with enormous amounts of data, often located at widely distributed locations, it is not surprising that computer support in this area will be of great impact on efficiency, accuracy, and advancement of health care. Many projects, often concerted international efforts, address the issue of how to handle an ever-increasing amount of medical information. Universal classifications have been designed and regularly refined, while other approaches aim not only to collect, but also to structure and disclose this exponentially growing body of medical information. The following sections will be devoted to a more general description of various topics of relevance to the field of medical informatics and may be of interest for the average reader.

PATIENT DESCRIPTION AND THE ELECTRONIC PATIENT FILE

The basic goals of the use of computers in medicine concern communication and clinically relevant combination of data. This electronic medium is expected to enhance and facilitate such interaction and data interpretation. Ideally, every citizen should carry a patient data card, which in an emergency case presents valuable information to the physician. The Medical Records Institute is an instrumental force in the movement toward such an electronic patient record. Locally, most hospitals have developed an information system [hospital information system (*HIS*)]. A patient card may include information on medical history, familial traits, use of prescription drugs, allergies, lifestyle (including sports activities and use of alcohol and/or tobacco), availability of x-ray pictures, electrocardiogram recording, and blood chemistry (2). Obviously, these initiatives involve delicate ethical issues, as well.

MEDICAL TERMINOLOGY AND EPONYMS

Knowledge obviously can be represented by symbols, words, definitions, and their interrelations. Knowledge may be expressed by spoken or written words, flow charts, (mathematical) equations, tables, or figures. Aspects of language and text interpretation are central issues in AI. A powerful abstraction of language also provides a powerful representation of knowledge. Various strategies have been explored: semantic networks offer a versatile tool for representing knowledge of virtually any type that can be captured in words by employing nodes (representing things) and links (referring to meaningful relationships), thus expressing causal, temporal, taxonomic, and associational connections. Other approaches (such as frame systems and production rule systems) have also been investigated. *Conceptual graphs* (3) are an emerging standard for knowledge representation, and the method is particularly suited to the representation of natural language semantics. Free-text data have limitations due to spelling errors, ambiguity, and incompleteness. However, formalisms that collect data in a structured and coded format are more likely to increase the usefulness regarding biomedical research, decision support, quality assessment, and clinical care (4). However, the lack of standardized medical language limits the optimal use of computers in medicine. Incorporation of knowledge bases containing equivalent expressions may be required for the practical use of medical information systems (5). The Generalized Architecture for Language Encyclopaedias and Nomenclature in medicine (*GALEN*) project, funded by the European Union, develops the architecture and prototypes for a terminology server (6). Indeed, medical language forms one of the greatest obstacles for the practical use of AI in the field of medicine (5, 7). Natural language often has remote roots, for example, adrenaline and epinephrine are the same chemical substances. Similarly, (pontine) angle tumor, acoustic (nerve) neurinoma, and acoustic neurilemmoma all have the same meaning. The Latin word *os* means both mouth and bone. The terms heterogenous, heterogeneous, and heterogenic look similar but all have a different meaning. Also, several English words have a dual meaning, for example, apprehension, aromatic, attitude, auricle, bladder, capsule, cast, cervical, cream, and cystectomy. Eponyms (8) further complicate descriptions. These examples illustrate the problem of translating medical phrases into concise "computer-storable" language. In addition to problems inherent to the understanding of natural language, additional difficulties pertaining to medical terminology can be indicated, as follows.

American Versus British Spelling. Two standard differences are evident, namely the use of the digraph in British spelling (e.g., anaemia versus anemia) and preference for using c (e.g., in leucocyte) rather than k (as in the American word leukocyte). Interestingly, the British equivalent of the American spelling of the word leukemia is spelled as leukaemia.

Preferred Terminology. In radiology "air" means gas within the body, regardless of its composition or site, but the term should be reserved for inspired atmospheric gas. Otherwise, the preferred term is "gas". Sometimes the preferred terminology refers to simplicity; the expression "lower extremity" must be replaced by "leg", for example. On other occasions the preferred terminology pertains to technical vocabulary that permits high precision if the available information is exact; the word "clumsiness" describes defective coordination of movement in general, whereas "dysdiadokokinesis" refers to a defect in the ability to perform rapid movements of both hands in unison (9).

Meaning Within a Certain Context. The quality "blue" primarily refers to a particular color. The actual meaning in medical language may, however, relate to a specific noun, for example, blue asphyxia, blue baby, blue bloater, blue diaper syndrome, blue dome, blue line, blue nevus, blue pus, blue sclera, blue stone, and blue toe syndrome (7).

Implicit Information. A particular statement may imply many relevant components, for example, if urinalysis is normal, then this result implies the absence of proteinuria, hematuria, glucosuria, and casts. Also antonyms may apply: leukopenia in particular implies "no leukocytosis". This mutual exclusion principle applies to all terms beginning with hypo- or hyper-.

Imprecise Terminology. (10) Some terms may carry a vague meaning, for example, tumor, swelling, mass, and lump. To a large extent, however, the use of such terms reflects the uncertainty related to an observation. In that respect it is justifiable: indeed it would be incorrect to specify an observation in greater detail than the facts permit. This notion has consequences for the selection of equivalent expressions.

Certainty Versus Uncertainty. Decision analysis itself does not reduce our uncertainty about the true state of nature, but as long as we must make some choice it does enable us to make rational decisions in the light of our uncertainty (11). Another aspect concerns subjective interpretation of percentage figures about prognosis (12). Outcomes perceived with certainty are overweighted relative to uncertain outcomes. Thus, the formulation of information affects its interpretation by humans.

Limited Scope of a Thesaurus. Thus far, no agreement exists regarding directives for coding diseases. Major sources are organized in different ways, for example, in the International Classification of Diseases[1] (*ICD*) one finds "Bladder, see condition (e.g., Leukoplakia)," whereas the book *Current Medical Information and Technology* (*CMIT*) (13) reads "leukoplakia (of bladder), see bladder." Notably, "leukoplakia of the bladder" as such is not listed in the book *Medical Subject Headings* (*MeSH*) (14).

Knowledge Engineering. This type of engineering implies various levels of translation. Thoughts by the human expert are formulated as precisely as possible, the engineer provides feedback using his or her own phrases to ensure an exact match between both minds, and subsequently the resulting expression is translated into a format usable for the computer program. These steps involve transformations of language while yet assuming that the ultimate user of the program fully appreciates the scope of the original thoughts of the expert.

Information Source Versus Actual Patient. (15) Current medical information sources tend to adhere to preference terminology to promote the use of uniform medical language. However, such standard vocabulary is not used by the average patient to describe individual health problems (16). Then it is left to the clinician to transpose, for example, "puffy face" and "moon face" if appropriate. Indeed, better health care can be realized by educating the patient about the value of structured communication with the physician (17).

Synonyms. For example, "icterus" is identical to "jaundice." Thrombocytosis and thrombocythemia are two words to indicate that the number of platelets in the peripheral circulation is in excess of 350,000 per microliter.

Subspecialty Interpretation. When naming a "hollow space" you may choose anything out of the following set: cavity, crypt, pouch, gap, indentation, dell, burrow, crater, concavity, excavation, gorge, pocket, cave, cavern, cistern, or lacuna. However, every expression may exhibit a nuance within a certain context. Then there is jargon: the terms "show," "engagement," and "station," for example, have a particular meaning within the field of obstetrics (7). The term "streaking" has a different meaning for the microbiologist and the radiologist.

Eponyms. Many disease names refer to the first author (e.g., Boeck's disease for sarcoidosis) who described that particular disorder, to the first patient analyzed in detail (e.g., Mortimer's disease, again for sarcoidosis), or to the geographical area (e.g., Lyme disease) where the illness was first detected. But variations may occur: The Plummer–Vinson syndrome (sideropenic dysphagia), as it is known in the United States and Australia, is termed Paterson–Kelly syndrome in the United Kingdom, but Waldenström–Kjellberg syndrome in Scandinavia (8).

Multilingual Approaches. The relation between a concept and the various corresponding terms in different languages is in general not unique. This implies that a multitude of different words from different syntactical categories may represent a single concept. Particularly the European countries are confronted with additional natural language problems. The Commission of the European Communities supports research activities in this area through the Advanced Informatics in Medicine (*AIM*) project, such as *EPILEX* (a multilingual lexicon of epidemiological terms in Catalan, Dutch, English, etc.) (18), and the development of a multilingual natural language system (19).

Frequency of Occurrence. The meaning of semiquantitative indicators such as "always" and "often" is not transparent when screening a medical text. The intuitive interpretation of some quasinumerical determinants is summarized elsewhere (20).

Noise Terms in Patient Description. When analyzing 104 patient cases, we found (21) that the input consisted on average of 75 terms; the required number of terms for establishing the primary diagnosis was only 15. This implies that 80% of the input data consisted of "noise terms" that may blur the process of hypothesis formation for humans (22).

Illogical Terminology. Certain terms contain paradoxical details, for example, hayfever is usually not accompanied by fever, while acute rheumatic fever typically has a chronic course.

CLASSIFICATION AND CODING SYSTEMS

With the exception of one British project, all classification or coding systems have been developed in the United States. The following survey lists all projects along with some of their characteristics.

- The ICD system just entered its tenth version, although the ninth edition is still used. It is applied worldwide for classifying diagnoses and also permits diagnosis-related group (*DRG*) assignment employed for billing and reimbursement purposes.
- Another widely accepted system is called Systematized Nomenclature of Medicine (*SNOMED*) (23), which offers a structured nomenclature and classification for use in human as well as in veterinary medicine. It covers about 132,600 records, with a printed and a CD-ROM version.
- The Current Procedural Terminology volume (*CPT*) (24) provides a uniform language for diagnostic as well as surgical and other interventional services. The system is distributed by the American Medical Association (*AMA*) and has been incorporated in the Medicare program.
- MeSH (14) is a systematic terminology hierarchy that is used to index the MEDLINE medical publications system, with annual updates.
- The National Library of Medicine[2] (*NLM*) in 1986 started a project called the Unified Medical Language System[3] (*UMLS*) (25), aiming to address the fundamental information access problem caused by the variety of independently constructed vocabularies and classifications used in different sources of machine-readable biomedical information.
- Gabrieli (26) constructed a computer-oriented medical nomenclature based on taxonomic principles. His system covers 150,000 preferred terms and a similar number of synonyms. The partitioning method employed for medical classification readily permits replacement of English names with terms of any other language, thus creating the perspective of a worldwide standard.
- Read from the United Kingdom designed a classification for various computer applications (27). Its design adheres to the following criteria: comprehensive, hierarchical, coded, computerized, cross-referenced, and dynamic. The system is closely connected to the British National Health Service (*NHS*). Version 2 includes 100,000 preferred terms, 250,000 codes, and 150,000 synonyms.

MEDICAL KNOWLEDGE BASES

Ideally a medical knowledge base (*KB*) integrates text, graphics, video, and sound. Furthermore, it should be accurate, verifiable, and easily accessible where doctors see patients, and the system should be adaptable to doctors' own preferred terms or abbreviations (28). Future developments will certainly include the use of ANNs (29), and examples realized thus far include myocardial infarction, diabetes mellitus, epilepsy, bone fracture healing, appendicitis, dermatology diagnosis, and electroencephalogram (*EEG*) topography recognition. An overview will be given of current KB systems. With the exception of the Oxford System of Medicine (*OSM*) and Medwise, all projects originate in the United States. One approach (CONSULTANT) addresses the field of veterinary medicine. A survey referring to the year 1987 has been published before (15).

Obviously, the Internet, a rapidly expanding network for computer-to-computer communication, nowadays offers a convenient window to medical resources. A useful guide, called *Medical Matrix* is the result of a project devoted to posting, annotating, and continuously updating full content and unrestricted access to this medium. The system can be reached at www.medmatrix.org and features a ranking system based on the utility for point-of-care clinical application.

- CMIT developed by the AMA (13) forms a reference for the selection of preferred medical terms including certain synonyms and generic terms with builtin arrangements to provide maximum convenience in usage, currency, and timely publication.
- Blois was the first to apply CMIT as a diagnostic tool in his RECONSIDER project (30). The application was released in 1981 and covered 3,262 disease entities, while 21,415 search terms were listed in a directory along with their frequency of occurrence. The program is extensively described in his book (30).
- DXplain (31) is also based on CMIT (13). The project had close connections with the AMA, and information is distributed using the World Wide Web. The KB contains information on 2,000 diseases and understands over 4,700 terms, with 65,000 disease-term relationships.
- QMR patient diagnostic software (32) covers some 600 disease profiles, and is the personal computer version of the INTERNIST-I prototype. Unfortunately the size of its disease KB remained remarkably constant over the last few years.
- MEDITEL (33) addresses the issue of diagnosis in adults. Over the last few years not much news was reported in the literature, apart from a comparative study (34).
- ILIAD (35) is a software package designed to aid students and residents in their clinical decision logic. The project stems from the health evaluation through logical processes (*HELP*) system developed at a major hospital in Salt Lake City. Its KB covers 1,300 diseases and 5,600 manifestations, mainly subspecialties of internal medicine.
- The Oxford System of Medicine (OSM) project, initiated by the Imperial Cancer Research Fund for use in primary care, helps general practitioners during routine work to support decision-making tasks such as diagnosing, planning investigations and patient treatment schedules, prescribing drugs, screening for disease, assessing the risk of a particular disease, and determining referral to a specialist (36).
- Medwise was founded in 1983 and now covers some 3,900 disease entities, with 29,000 different keywords (21). It includes a separate KB with almost 500 equivalent terms that each refer, on average, to three related terms. Equivalent terms are automatically generated to assist the user during the process of data entry. The matrix structure of the Medwise KB permits semantic differentiation, with corresponding weight factors for disease profile matching.
- The Framemed system (37) divides medical information into 26 domains and arranges the items in a hierarchical sequence, thus yielding a logical framework for a standardized terminology. The objective is to achieve a standard coded terminology (including synonyms) to which all existing systems can relate, with obvious use as an electronic encyclopedia and for differential diagnosis.
- STAT!-Ref (38) offers the contents of a first-choice medical library (including several standard textbooks, e.g., on primary care) as well as Medline on CD-ROM.
- MD-Challenger (39) offers a clinical reference and educational software for acute care and emergency medicine (everything from abdominal pain to zygoapophyseal joint arthritis), with nearly 4,000 annotated questions and literature references. MD-Challenger also includes continuing medical education (CME) credits.
- Labsearch/286 is a differential diagnosis program allowing input of up to two abnormal laboratory findings plus information on symptoms and signs. Laboratory data concentrate on body fluids (blood, urine, cerebrospinal, ascitic, synovial, and pleural fluid) entered as high or low (40). The system includes 6,500 diseases and 9,800 different findings.
- CONSULTANT (41), a KB for veterinary medicine, was developed at Cornell University. This database for computer-assisted diagnosis and information management is available on a fee-for-service basis in North America.
- Griffith and Dambro (42) compiled an annually updated book. The information is compiled by a group of contributing authors whose names are listed in conjunction with each disease profile. The first edition appeared in 1993 and contains chartlike presented information on 1,000 topics, along with their ICD code. The printed version shows similarity with CMIT [13]. Publication using electronic media recently became available.
- The *Birth Defects Encyclopedia* (43) is a comprehensive, systematic, illustrative reference source for the diagnosis, delineation, etiology, biodynamics, occurrence, prevention, and treatment of human anomalies of clinical relevance. A unique feature of the printed edition is the Fax service for requesting a current, daily updated version of any article in the KB. The related birth defects information system (*BDIS*) is a sophisticated computer-based profile-matching system that helps research and diagnostic tasks associated with complex syndromes.
- DiagnosisPro by MedTech, U.S. (44), is a differential diagnosis system including 8,500 diseases, designed by the internist C. Meader and the clinical pathologist H. C. Pribor. It covers all major specialties and is considered a useful tool for primary care professionals.

EXPERT SYSTEMS AND COMPUTER ASSISTED DIAGNOSIS

The term *expert system* or *knowledge-based system* (KBS) refers to a computer program that simulates the professional capabilities of a human expert (45–49). In the field of medicine, the expression *computer-assisted decision system* is also used for an expert system. KB systems are used for interpretation of actual data about a specific problem, considering the knowledge represented in the domain of the KB, to develop a problem-specific model and then to construct plans for problem solution. KBs usually include facts about the problem domain, and procedural knowledge to manipulate facts. In production systems this procedural knowledge adopts the form of IF-THEN or IF-THEN-ELSE rules, where the IF part is the antecedent, the THEN part is the conclusion, and the ELSE part, if exists, is the alternative conclusion. Candidate hypotheses are derived through some pattern-matching system. A reasoning "engine" (also termed *inference machine*) carries out the manipulation specified to obtain an answer. An inference engine is no more than a program, the function of which is to decide what to do at any given moment, that is, it recognizes and activates the appropriate rules. Generally, an inference engine should include an interpreter, which activates the relevant rules at any given moment, taking into account the current state of the active memory; a search strategy, which includes exploration heuristics; a self-knowledge mechanism, which permits the identification of the structures being utilized, the state of the problem, and changes in the active memory; and a termination mechanism for the inferential processes.

The overall functioning of the inference engine occurs in cycles called *basic production system cycles*. The nature of these basic cycles is very different depending on whether the search process is directed by the data or by the objectives. Given that the production systems are essentially based on rules, it will be necessary to define how the propagation of knowledge within the system can be affected. Let us look at two basic propagation methods.

1. Forward chaining of rules, which is implemented as a search process directed by the data, and where the search initiates with the antecedents and leads to the conclusions of the rule
2. Backward chaining, which is implemented as a search process directed by the objectives, and where the search, by means of an evocative process, initiates with the conclusions of the rules, established as hypotheses, and extends to the antecedents

Verification and validation are the two most important stages in the evaluation of an expert system's behavior and functioning (48–53). Verification endeavors to ensure that the system has been constructed correctly. This means that it ensures that the software contains no errors and that the final product satisfies the initial design specifications and requirements. Validation, on the other hand, refers more precisely to a detailed analysis of the quality of the expert system in the context of its work environment, which permits determining whether or not the developed product adequately meets the expectations deposited therein. Verification of an expert system necessarily involves a detailed analysis of the knowledge contained within the system. Particularly if we refer to production systems, the rules may be the origin of many errors, among which the following can be identified:

Conflictive knowledge

$$p(x) \to q(x)$$
$$p(x) \to \text{not } q(x)$$

Circular knowledge

$$p(x) \to q(x)$$
$$q(x) \to r(x)$$
$$r(x) \to p(x)$$

Redundant knowledge

$$p(x) \quad \text{and} \quad q(x) \to r(x)$$
$$q(x) \quad \text{and} \quad p(x) \to r(x)$$

Unnecessary knowledge

$$p(x) \quad \text{and} \quad r(x) \quad \to q(x)$$
$$p(x) \quad \text{and} \quad \text{not } r(x) \to q(x)$$

Rules included in or contained within others

$$p(x) \quad \text{and} \quad r(x) \to q(x)$$
$$p(x) \to q(x)$$

Rules never executed

$$p(x) \quad \text{and} \quad r(x) \to q(x)$$
$$r(x) \quad \text{cannot be obtained}$$

The preceding erroneous situations must be detected and resolved, and in order to do this, the verification of the expert system may be approached from two different perspectives: first, verification dependent on the domain and second, verification independent of the domain. In the preceding section we have described several systems with KBs that are used for establishing a differential diagnosis. Recently, the diagnostic performance of four such commercially available programs (Dxplain, Iliad, Meditel, and QMR) was evaluated (34). The fraction of correct diagnoses by the computer ranged from 0.52 to 0.72, while half of the candidate diagnoses proposed by the human experts were not generated by the computer. However, on average each program suggested two additional diagnoses per case that the human expert did find relevant but which the researchers failed to include in their original differential diagnosis list. An obscure limitation of the study design is that the researchers themselves created a set of cases to be analyzed by the computer programs. Also, they employed the vocabulary provided by the program's developer. Our written requests to receive a copy of the patient cases used in this study for further testing of other programs were not honored thus far. In his editorial, Kassirer (45) wrote that the results of the study indicate substantial progress, but he found them disappointing from a physician-skeptic point of view. He concluded that the structure of the KBs, the computational *inference engine* that integrates clinical data into a diagnosis, the methods of capturing clinical data from patient's records, and the human-computer interface are still in their infancy. However, several field prototype expert systems have been successfully validated (55–59), and some methodologies for expert systems validation have been proposed (60–62).

ARTIFICIAL NEURAL NETWORKS (ANN)

In 1956, a group of researchers met in Dartmouth college in order to discuss the possibility of constructing genuinely intelligent, technologically advanced machines. This meeting laid the foundations for the science of AI. Principal among the participating researchers of note were Samuel, McCarthy, Minsky, Newell, Shaw, and Simon. Following this meeting two major breakaway groups were formed, both of which continued working more or less independently of each other. Thus, Newell and Simon formed a team with the idea of developing human behavior models, whereas McCarthy and Minsky formed another team dedicated to the development of intelligent machines, not being particularly concerned with human behavior as such. The first approach entailed an emulation of cerebral activity and, wherever possible, the copying of its structure. The second approach implied the construction of systems in which the problem-solving procedures applied are such that, were human beings to apply them, they would be considered intelligent.

However, in practice a combination of both approaches is necessary in order to obtain results that may be considered useful. Both approaches comply with the fundamental objectives of modern AI, namely, the understanding of human intelligence and the use of intelligent machines to acquire knowledge and to resolve complicated problems satisfactorily. Both approaches lead to AI programs, KB systems, expert systems and, finally, to ANNs.

AI programs can be said to exhibit a certain intelligence as a result of the skillful application or use of heuristics in the broadest sense. Heuristic knowledge is considered the fruit of experience, which is difficult to formulate, and which is established implicitly in order to find answers to a specific problem, answers that may be more or less accurate but that are nevertheless always valid. Although the use of heuristic knowledge does not guarantee the finding of optimal solutions, it does offer acceptable solutions, if they exist, by means of so-called inferential processes.

The next epistemological level is that of knowledge-based and expert systems, for which knowledge of the specific domain and knowledge of the control structures used to manipulate said knowledge are physically separate. Therefore, it is necessary to define and implement architectures different from the ones we are accustomed to, in which knowledge and control structures can be developed independently of one another in such a way that one specific control structure can be applied to knowledge from different domains. Expert systems can be considered as specialized knowledge-based systems, in that they resolve real-life problems, which, although limited in terms of size, are complex. The construction of an expert system requires the employment of techniques developed to construct AI programs, in addition to architectures defined for the development of knowledge-based systems. However, it is absolutely essential to place more emphasis on differential aspects such as the acquisition of knowledge and learning.

ANNs can be defined as massively parallel distributed processors with a natural capacity not only for storing experience-based, that is, heuristic, knowledge, but also as a facility for making such knowledge available for use. ANNs allow limitation of the brain in two ways. First, knowledge is acquired by means of a learning process, and second, the synaptic weights are used for storing the knowledge.

It is obvious that, in order to obtain acceptable results at any of the levels of AI described above, we need to draw from other fields such as mathematics (its language and procedures), medicine (especially the neurophysiological models), computer science (particularly software engineering and systems architecture), linguistics (especially syntax and semantics), psychology (which allows us to analyze intelligent behavior models), and finally, even philosophy.

Advanced Aspects of AI

Dealing with Uncertainty. AI is not only concerned with general mechanisms related to the search for solutions within a given space or with how to represent and utilize the knowledge of a specific discourse domain. Another aspect, up to now just mentioned in passing, is that concerning inferential mechanisms and/or processes, which are considered as the starting point for the so-called *reasoning models*.

In any domain, the propagation of knowledge by means of AI programs is always carried out by following a well-defined reasoning model. These reasoning models contribute in a decisive way to the correct organization of the search for solutions. Normally, the domain characteristics and the characteristics of the problems to be solved determine the type of reasoning model to be employed. Thus, there are domains of a markedly *symbolic* nature, in which solutions can be established with absolute confidence. In these cases the use of categorical reasoning models is indicated (63). There are, on the other hand, domains that are of a *statistical* nature, where unique solutions cannot be obtained and where in addition, a decision must be made as to which of the possible solutions arrived at is the most probable. In these cases it is preferable to reason with statistical models of which, given the peculiarities of the inferential processes that AI deals with, the *Bayesian scheme* is the most widely used (64, 65).

There are other domains in which the concept of uncertainty appears and which may be inherent to the data of the problem and the facts of the domain, or to the inferential mechanisms themselves. In such cases reasoning models are chosen that are capable of correctly manipulating such uncertainty (66–68).

Finally, there are domains in which the inferential elements include nuances of a linguistic nature where hierarchies and classifications can be established. Indicated in these cases are reasoning models based on *fuzzy sets* (69, 70).

Obviously there are domains that manifest more than one of the characteristics just mentioned, in which case the reasoning model most appropriate to the characteristics of the domain or a combination of different models can be used.

The Dempster-Shafer Theory of Evidence

This reasoning scheme has a solid foundation in theory to the extent, in fact, that the original reasoning model proposed by Dempster was subsequently formalized and converted into a genuine theory by Shafer (71). This scheme is attractive, principally for the following reasons: (1) it permits the modeling of uncertainty associated with pieces of evidence and hypotheses in a simplistic manner; (2) it permits the consideration of sets of hypotheses without the confidence in each set having to be distributed in any way between each of the individual hypotheses of the set; (3) it permits an elegant but precise representation of the lack of knowledge so frequently associated with reasoning processes; (4) it deals with the probability theory as a special case; (5) it contains some of the combinatory functions of the certainty factors model.

But how is it possible to deal with both the inexact knowledge and the lack of knowledge in the Dempster-Shafer model? In the first place, and given any discourse universe whatsoever, Dempster and Shafer introduce the concept of a *discernment frame* that can be defined as the finite set of all hypotheses that can be established in the problem domain. The discernment frame should form a complete and thus exhaustive set of hypotheses that are mutually exclusive. On the other hand, the effect of a specific piece of evidence on the overall set of hypotheses is not determined by the contribution of the confidence deposited in the individual hypotheses. On the contrary, each piece of evidence affects, generally speaking, a subset of hypotheses within the discernment frame. This approach is consistent with the reality of almost all real-life routine problems. In real-life problems, the reality is that the evidence e permits discrimination between groups or sets of alternative hypotheses. However, at the same time, within a set, uncertainty with respect to the alternative hypotheses is maintained. According to this argument:

Z is the discernment frame
A is any subset of the frame
$h_1, \ldots, h_n$ are the hypotheses of the discernment frame

In this context, the appearance of specific evidence e will favor a determined subset A within Z, in such a way that the degree to which A is favored by e is represented by $m(A)$, where m is indicative of the confidence that the evidence e permits in A. The values of m are represented by the closed interval [0, 1]. We will use the following notation:

$$e: \quad A = [h_a, h_b, h_c] \to m(A) = x, \; x \in [0, 1]$$

The fact that the evidence e supports the subset A does not imply, as already pointed out, that the individual hypotheses divide, explicitly, the confidence deposited in A itself. This fact diverges considerably from classical theories of probability. Each subset of the discernment frame, for which given evidence e it is established or verified that $m(A) \neq 0$, is called a *focal element*. Returning briefly to the basic probability theory, Dempster and Shafer define the following conditions for m:

$$\sum_{A \subset Z} m(A) = 1$$
$$m(\varnothing) = 0$$

Both conditions are a direct consequence of the restrictions imposed on the discernment frame. Evidence theory provides us with a neat way to deal with the lack of knowledge associated with reasoning processes. Let us take a discernment frame Z and evidence such that

$$e: \quad A \subset Z \to m(A) = s, \qquad 0 \le s \le 1$$

The first condition required for m establishes that $\Sigma A \subset Z$ $m(A) = 1$.

What happens to the rest of the confidence that has not been assigned to the focal element A? To answer to this question, Dempster and Shafer postulate that if

$$e: \quad A \subset Z \to m(A) = s, \qquad 0 \le s \le 1$$

then

$$m(Z) = 1 - m(A) = 1 - s$$

This formula should be read as "given that the evidence e supposes the assignation of a given confidence to a specific focal element A within the discernment frame, then the rest of the unassigned confidence represents a 'lack of knowledge' and therefore should be assigned to the discernment frame itself." This situation leads us to reflect as follows: Unassigned confidence is ignorance or lack of knowledge with respect to the importance of the evidence in relation to the focal element under consideration. In other words, it is known that the evidence supports the focal element to the extent of s. However, referring to the unassigned confidence $(1 - s)$, we do not know if it contributes or not to A (or to any other subset within the frame). The unassigned confidence $(1 - s)$ should be assigned to the frame since we constructed the frame; what we do know as a fact is that the solution is within the frame. The complete formulation for the approach is as follows:

The discernment frame $Z = [h_1, \ldots, h_n]$
The focal element $A \subset Z$
Evidence e referring to A
A measure of the "basic probability" of A given e: $m(A)$
$e: A \to m(A) = s$
$m(Z) = 1 - s$
$m(B) = 0 \; \forall \; B \subset Z, B \neq 0, B \neq A$

If the approach were probabilistic, the same evidence would support the focal element A as well as the complement of the focal element. Thus,

$$p(A) = s \to p(\text{not}\, A) = 1 - s$$

This was precisely one of the major drawbacks of the probabilistic models. With this new theory, if

$$Z = [h_1, h_2, h_3, h_4]$$

and

$$A = [h_1, h_2]$$

with

$$e: \quad A \rightarrow m([h_1, h_2]) = s$$

then

$$m([h_1, h_2, h_3, h_4]) = 1 - s$$

Generalizing broadly, it could be said that the way in which the lack of knowledge in the evidential theory is managed more than compensates for the defects in the probabilistic models.

Fuzzy Systems. Uncertainty does not only occur as a consequence of an absence of information or of other circumstances that may be, to a greater or lesser extent, formalized. On the contrary, uncertainty could well be associated with the very way in which humans express themselves. In practice, most human statements are ambiguous and this ambiguity is an essential characteristic not only of language but also of the processes of classification, of the establishment of taxonomies and hierarchies, and of the reasoning process in itself. Hence, if we define *living things* as organized molecular structures that are born, grow, reproduce, and die, it is clear that even the humble beetroot does precisely the same and therefore should also be considered a living thing. On the other hand, a sliver of a stone does not behave in the same way, and thus should not be considered a living thing. But what about a virus?

The difficulty in defining a virus so as to include it in the set of living things lies with the selfsame definition of the concept *living things*. In other cases, however, the difficulty arises due to questions of a subjective nature. For example, characterizing the set of *beautiful people* is not easy, since each person has a very personal idea of exactly what attributes an ideal should have in order to be beautiful. But it is particularly difficult to say if one specific object is beautiful or not. In this particular situation, subjective nuances appear that render impossible the very idea of classification.

Furthermore, it is not only problems of definition or of subjective nuances that may complicate categorical classification. In other cases, the context may modify the criteria. For example, the concept of *a tall man*—which in itself is intrinsically ambiguous—differs notably depending on whether one refers to a Scandinavian or to a Pygmy, and to make matters worse, probably both are right!

So it is possible to conclude rapidly that ordinary sets, in which an element of a determinate universe may or may not belong, are not sufficiently complete so as to represent the knowledge normally utilized within the command of a human being, not to speak of reasoning with that knowledge. The fields of mathematics and AI, at their outset concerned with interesting problems of a cognitive nature, were from an early stage intrigued by this concept. Finally, in 1965 Lofti Zadeh published in his famous article "Fuzzy Sets" the results of his investigations in this area (72).

Let us take any universe whatsoever, for example, the universe formed by the set N of natural numbers. Let us define a subset of N called A, characterized by the following description: A is the subset formed by natural even numbers of a value less than 10. Thus A, a subset of N, is perfectly defined as follows: $A = [2, 4, 6, 8]$, and obviously 2 belongs to A, but 3 does not belong to A, and 10 does not either. In this particular example, we have no difficulty in establishing the *degree of belonging* of an element of the discourse universe to a particular subset.

Now let us look at the universe C characterized by the following description: C is the set formed by all human beings; let B be a subset of C characterized by the description "B is a subset of C that includes tall, dark men." In this situation it is difficult to establish the degree of belonging of an element of the universe C to the subset B. Clearly, a common set may be defined as a collection of elements and if an element of the universe is represented in the collection, the element in question belongs to the said set. In these cases, it can be said that the degree of belonging of any particular element of the referential universe has a Boolean value. Thus if the element belongs to the set, the Boolean value is 1. If the element does not belong to the set, the Boolean value is 0. In this way we can construct a function f (which for common sets is a Boolean function) such that, given an element x from the universe U, and given (also) a subset of U, A, then

$$f_A(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if } x \notin A \end{cases}$$

We will now extend the discussion to those special kinds of sets that we have called *fuzzy sets*. In reference to these, we said that linguistic, subjective, and other nuances impeded a precise establishment of the degree of belonging to the fuzzy set in question, of particular elements of the universe. Thus there will be elements from the universe that clearly belong to the set, others that clearly do not belong, and yet others that belong to a certain extent, but not totally. Following the approach previously described, the problem is easy to resolve if we consider the function f to have the following values, given an element x from the universe U, and (from U) a fuzzy subset A (from U):

$$f_A(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if } x \notin A \end{cases}$$
$$0 < f_A(x) < 1 \text{ if } x \text{ partially belongs to } A$$

The function f quantifies the degree of belonging of an element from the universe to the fuzzy set in question. Thus, a fuzzy set is one for which there does not exist a clear dividing line between belonging and not belonging of determinate elements from the universe. In order to establish the *fuzzy limits* of the corresponding set, we shall require a criterion, which naturally will be arbitrary. Let us examine the universe of *living persons* along with the fuzzy set A of U answering the description "A is the set of young living persons." a property we consider appropriate for the characterization of the fuzzy subset A is the "age" of the universe elements, but how should we define age? We are faced with the not insignificant problem of the definition of criteria for the "fuzzification" of sets. In our example, we will consider as "young" all those elements from the universe whose age permits them to legally obtain Youth

Travel Cards, Inter-Rail tickets, etc. (i.e., elements of age ≤25 years old), and "not young" all those elements that can legally obtain Pensioner Travel Cards (i.e., elements of age >65 years old). Thus,

$$f_A(x) = \begin{cases} 1 \ \forall\, x/\text{age}(x) \leq 25 \\ 0 \ \forall\, x/\text{age}(x) \geq 65 \end{cases}$$

But what happens to all those elements from the universe aged 26 to 64 years old? What exactly is their degree of youth with respect to the criterion of age? We are faced with yet another problem, which is the characterization of the diffuse area or zone. In order to get out of the conundrum, we will construct a linear function as follows:

$$\begin{aligned} f_A(x) &= [65 - \text{age}(x)]/(65 - 25) \\ &= [65 - \text{age}(x)]/40 \ \forall\, x/\text{age}(x) \in [25, 65] \end{aligned}$$

In this way we can segment our numeric space [0 −1] in three zones, two of which are not fuzzy and which refer to those elements of the universe that clearly belong or that clearly do not belong, to the fuzzy subset in question, and a third zone that is fuzzy and that corresponds to those elements of the universe that belong, to a certain extent, to the fuzzy subset in question.

We will now examine a situation in which Tom is 17 years old, Dick 31, and Harry 73. Conscious of the fact that each one is a "living person," what can we say about their "youth"? In accordance with the established criteria, and carrying out the appropriate substitutions, we can establish the values of their respective functions of belonging to the fuzzy subset of "young living persons" as follows:

$$\begin{aligned} f_{\text{young}}(\text{Tom}) &= 1.00 \\ f_{\text{young}}(\text{Dick}) &= 0.85 \\ f_{\text{young}}(\text{Harry}) &= 0.00 \end{aligned}$$

It is obvious that as the ages of our friends increase, their degree of belonging to the fuzzy subset decreases. The approach is coherent but is not very natural in linguistic terms. Just to illustrate our meaning, have a look at the following dialogue: "By the way, Sally, how old is Dick?" "I think he is 31." "Oh, so he is 0.85 young!" Absurd! Nobody talks in this way! We are faced again with a new problem, that of *linguistic classification* of fuzzy sets. The basic idea is that once we have managed to segment the numeric space (indicative of the degree of belonging of each element to the fuzzy subset in question), we need to segment the linguistic space by means of labels containing information of a semantic nature and then to match each linguistic label to a specific numeric interval, on the basis of a (minimally) reasonable criterion. Returning to our example, we define a linguistic scale to which we assign concrete values from our function degree of belonging to the fuzzy subset in question. Thus

$$\begin{aligned} f_A(x) &= 0.00 \rightarrow \text{not at all young} \\ 0.00 < f_A(x) &< 0.20 \rightarrow \text{very slightly young} \\ 0.20 \leq f_A(x) &\leq 0.40 \rightarrow \text{a little young} \\ 0.40 < f_A(x) &< 0.60 \rightarrow \text{to some extent young} \\ 0.60 \leq f_A(x) &\leq 0.80 \rightarrow \text{moderately young} \\ 0.80 < f_A(x) &< 1.00 \rightarrow \text{fairly young} \\ f_A(x) &= 1.00 \rightarrow \text{absolutely young} \end{aligned}$$

According to this scale, and using the facts from the example, we may now say that Tom is "totally young" (or simply "young"), Dick is "fairly young," and Harry is "not at all young" (or simply "not young"). These expressions represent a natural, human way of expressing judgments with respect to the ages of our friends.

Although an in-depth discussion of the problems deriving from knowledge representation and from fuzzy reasoning is way beyond the scope of this text, it is, however, necessary to include a reference to both. Remember that, from the perspective of AI, the fuzzy model permits us to represent and manipulate expressions appropriate to the language of human beings. In such expressions we come across fuzzy predicates, fuzzy quantifiers, and fuzzy probabilities. Other, more conventional approaches to the representation of knowledge lack the means for efficiently representing the meaning of fuzzy concepts. Models based on first-order logic, or those based on classical probability theories, do not allow us to manipulate the inappropriately named common-sense knowledge. The reasons for this are as follows.

Knowledge derived from common sense is lexically imprecise.

Knowledge derived from common sense is of a noncategorical nature.

The characteristics of the fuzzy sets examined in the previous paragraphs give us clues as to the procedure to follow if what we require is the application of knowledge representation models and reasoning models, based on fuzzy logic(s). Thus (73)

1. In fuzzy logic, categorical reasoning is a special case of approximate reasoning.
2. In fuzzy logic, it is all a question of degree.
3. Any fuzzy system can be "fuzzified."
4. In fuzzy logic, knowledge should be interpreted as a collection of "fuzzy restrictions" placed on a collection of variables.
5. In fuzzy logic, reasoning problems and therefore inferential processes should be interpreted as "propagations" of the fuzzy restrictions mentioned previously.

Although the theoretical bases for fuzzy logic are quite clear, applications of the latter to systems of an inferential nature is problematic. Even at the time of writing, these difficulties have not been entirely overcome. Nevertheless, it appears that fuzzy-system theories applied to control problems, in place of more conventional approaches, are coming up with solutions that are both brilliant and elegant.

FURTHER READING

The fields of medical KBs and terminology are rapidly developing. There is no single comprehensive source of information available, but the professional reader is advised to scrutinize the following journals and organizations for updated information.

M.D. Computing, published by Springer Verlag (New York and Berlin), reports on research in the field of medical informatics. Of special interest to the clinician is also the journal *Experts Systems with Applications*, published by Pergamon Press (New York).

IEEE Expert appears four times a year and presents the latest on AI and expert systems. Contact P.O. Box 3014, Los Alamitos, CA 90720. This journal is for those interested in technical details on intelligent systems and their applications. IEEE (P.O. Box 1331, Piscataway NJ 08855-1331) also publishes a number of related journals, for examples, on knowledge engineering, fuzzy logic, ANNs, and multimedia.

The NLM releases news bulletins and provides information on UMLS and contracts for cooperation.

Obviously, annual meetings form the forum for presentation of the latest developments:

IMIA conferences are well known, besides the world congress organized every four years. IMIA publishes a *Yearbook of Medical Informatics*. It offers the pearls of medical informatics since it covers influential papers from 100 journals in the field. Contact Schattauer Publishers, P.O. Box 104545, 70040 Stuttgart, Germany.

USABILITY OF EXPERT SYSTEMS

In the development of medical software systems in general, and in medical KBSs in particular, there is currently a gradual shift in philosophy, from a "system-directed" one, in which internal architecture and functions set the pace for development, towards a user-centred philosophy (referred to as user-centered design, or UCD), in which the user is implicated in design aspects. However, one of the main obstacles to this new approach to design is the lack of suitable tools. Consequently, greater effort is required from the software engineering community in the field of man–machine interaction or human–computer Interaction (HCI). A wide range of techniques are currently available for analyzing the usability of computerised intelligent systems. (74) The fact that so many techniques have been developed is due to the fact that, to date, no single method will ensure that a system is usable. In fact, the use of several approaches and an overall analysis of results are generally recommended. In an attempt to organize and facilitate the learning of usability analysis techniques, a number of authors have classified these in terms of hierarchical models.

Of particular interest are the classifications drawn up by Ivory and Hearst (75), Adelman and Riedel (76), and Preece (77). The simplest of these is the Adelman and Riedel classification, consisting of three main categories, namely, heuristic methods, subjective methods, and empirical methods. These are described in turn as follows:

1. Heuristic methods are based on the opinions of usability experts. These experts analyze the different system interfaces and determine strengths and weaknesses from an end-user perspective. Heuristic analysis techniques can be classified in one of two main groups:
   - Analytical techniques, in which evaluators collect information on the usability of a system by constructing formal models that represent the system being assessed for usability. One of the most relevant analytical techniques in our context is the GOMS analysis (78).
   - Heuristic inspections, in which evaluators apply criteria of a heuristic nature to the identification of possible usability problems. Specifically this consists of a systematic analysis of the system's user interface by usability experts. The ultimate aim of the inspection is to identify usability problems in relation to design that can be resolved in the iterative development phase (79).
2. Subjective methods are based on the opinions of the system users. They analyze operational prototypes of the product in the development phase and voice their opinions on the usability of these prototypes. These methods can be classified in one of two main groups: survey methods and observational methods.
   - The aim of the survey methods is to compile data on the opinions of users after these have used the system, by using interviews, questionnaires and/or user feedback.
   - Observational methods, on the other hand, are based on obtaining user opinions as these use the system rather than after they use the system. The techniques used in this case are direct observation, indirect observation, verbal protocols, post-event protocols, and constructive interaction. Of the techniques described above, the only approach that can be completely automated - in view of the current state of technology - is the development of closed questionnaires. A particularly popular approach to evaluating closed questionnaires is MAUT (multi-attribute utility Theory) analysis (80). An alternative technique is AHP (Analytic Hierarchy Process) (81).
3. Empirical methods are based on the actions of the system users. The approach is based on obtaining objective data on practical hands-on use of the system. Empirical methods are based on an analysis of the actions of users of a system. The process generally consists of logging a user's interaction with the system while he/she is performing a series of tasks. This log is then analyzed to produce a series of measurements (such as number of errors, time required to perform a task, etc.). Conclusions on system usability can then be drawn by performing a detailed analysis of these measurements.

KBS HYBRID ARCHITECTURES

During the last years, there has been an increasing tendency in developing expert systems that are progressively more complex. In order to be able to cope with the increasing complexity, expert systems have incorporated several intelligent techniques that try to take advantage of the complimentary characteristics of symbolic systems, neural networks, fuzzy systems, or genetic algorithms, among others. These systems are called hybrid intelligent systems, and combine intelligent techniques as well as conventional computing techniques to achieve a higher level of machine intelligence. Hybrid systems help to:

1. Improve the available techniques, integrating several of them so as to conceal the problems that each of them present. For example, neural networks are good at learning, but can not do high level reasoning. On the other hand, symbolic expert systems are good at high level reasoning, but more limited in learning.
2. Find solutions for complex tasks. Most application domains present several subtasks with different characteristics. For example, the logic and static components can be adequately managed by expert symbolic systems, while other components that are dynamic, fuzzy or poorly understood could be managed, for example, by neural networks.
3. Implementing multifunctional systems. In this case, the goal is to create a system that can exhibit multiple capacities for information processing in a unique architecture. That is, there is only one system, but it tries to emulate the behaviour of different processing techniques. One example of this is the use of neural networks for symbolic processing.

Depending on factors such as functionality, processing architecture and communication requirements, three basic types of architectures for hybrid systems can be distinguished (82):

1. Expert systems with function replacement, in which a principal function of a given technique is replaced by another intelligent processing technique. The aim for replacement is either increasing the execution speed or enhancing reliability. An example of this type of hybrid system could be the replacement of the backpropagation weight changing mechanism of a neural network with genetic algorithm operators.
2. Intercommunicative hybrids, which are independent, self-contained intelligent processing modules that exchange information and perform separate functions to generate solutions. It is used when a problem can be divided in sub-problems, each of which can be resolved using a different technique, such as neural networks, symbolic systems, etc. An example could be a diagnosis system in which an expert system realizes inferences, and calls neural networks when needed to analyze data top obtain patterns (83).
3. Polymorphic hybrids, which are systems that use a single processing architecture to achieve the functionality of different intelligent processing techniques. An example is a neural network that tries to perform symbolic tasks such as step-wise inferencing (84).

METHODOLOGIES

Nowadays, the perspective of knowledge transfer (eliciting knowledge from the expert and translate it to a tool using some kind of software methodology) has been substituted by the perspective of knowledge modelling. This has been achieved using model-based methodologies that approach the complex problem of knowledge engineering by constructing different aspect models of the human knowledge involved in some complex domain. There are several methodologies available, such as CommonKADS (85), MIKE (86), Protégé-II (87), etc. All these approach knowledge acquisition and modelling from a structural point of view and try to palliate the knowledge acquisition bottleneck.

The last methodology, Protégé, allows also for the definition of ontologies. Ontologies make possible knowledge sharing and reuse, playing a major role in supporting information exchange across various networks. An ontology describes the concepts and relationships that are important in a particular domain, providing a vocabulary for that domain as well as a computerized specification of the meaning of terms used in the vocabulary. Ontologies range from taxonomies and classifications, database schemas, to fully axiomatized theories. In recent years, ontologies have been adopted in many business and scientific communities as a way to share, reuse and process domain knowledge. Ontologies are now central to many applications such as scientific knowledge portals, information management and integration systems, electronic commerce, and semantic web services. Ontologies applied to the World Wide Web are creating the Semantic Web (88). There are several groups that attempt to implement content management infrastructure and support the management of the vast amount of knowledge encoded in clinical systems. These ontologies and rules are served up through applications and services to support guided observation capture, guided ordering, and guided interpretation of clinical data. Workflow portals leveraging this knowledge include the electronic health record for care-givers and consumers, quality performance management, and clinical research.

BIBLIOGRAPHY

1. G. L. Solomon, M. Dechter, Are patients pleased with computer use in the examination room? *J. Fam. Pract.*, **41**: 241–244, 1995.
2. P. L. M. Kerkhof, M. P. van Dieijen-Visser (eds.), *Laboratory Data and Patient Care*. New York: Plenum Publishing, 1988.
3. J. F. Sowa, *Conceptual Graphs, Information Processing in Mind and Machine*. Reading, MA: Addison-Wesley, 1984.

[1] The ICD is a widely accepted system that organizes all possible medical diagnoses. The tenth version has been translated for worldwide application. Developed by the Commission on Professional and Hospital Activities, Ann Arbor, Michigan.

[2] 8600 Rockville Pike, Bethesda, MD 20894. The NLM releases news bulletins and provides information on UMLS and contracts for cooperation.

[3] The UMLS is a project initiated by the NLM and distributed on CD-ROM. Cooperation with parties to implement the system within their own environment is encouraged but requires a contract (25).

4. P. W. Moorman *et al.*, A model for structured data entry based on explicit descriptional knowledge. *Meth. Inf. Med.*, **33**: 454–463, 1994.
5. P. L. M. Kerkhof, Knowledge base systems and medical terminology. *Automedica*, **14**: 47–54, 1992.
6. A. L. Rector *et al.*, Medical-concept models andl medical records: An approach based on GALEN and PEN&PAD *J. Am. Med. Informatics Assoc.*, **2**: 19–35, 1995.
7. P. L. M. Kerkhof, *Woordenboek der geneeskunde N/E & E/N* (Dictionary of Medicine) 2nd printing. Houten, The Netherlands: Bohn Stafleu Van Loghum, 1996.
8. B. G. Firkin, J. A. Whitworth, *Eponyms*. Park Ridge, NJ: Parthenon Publication Group, 1987.
9. E. A. Murphy, *The Logic of Medicine*. Baltimore: Johns Hopkins University Press, 1976.
10. V. L. Yu, Conceptual obstacles in computerized medical diagnosis. *J. Med. Philos.*, **8**: 67–83, 1983.
11. M. C. Weinstein, H. V. Fineberg, *Clinical Decision Analysis*. Philadelphia: W. B. Saunders, 1980.
12. S. A. Eraker, P. E. Politser, How decisions are reached; physician and patient. *Ann. Intern. Med.*, **97**: 262–268, 1982.
13. A. J. Finkel (ed.), *Current Medical Information and Terminology*, 5th ed. Chicago: American Medical Association, 1981.
14. MeSH, *Medical Subject Headings*. Bethesda: National Library of Medicine, 1993.
15. P. L. M. Kerkhof, Dreams and realities of computer-assisted diagnosis systems in medicine (Editorial). *Automedica*, 8: 123–134, 1987.
16. J. W. Hurst, The art and science of presenting a patient's problems. *Arch. Intern. Med.*, **128**: 463–465, 1971.
17. J. Verby, J. Verby, *How to Talk to Doctors*. New York: Arco Publ., 1977.
18. C. Du V Florey, European Commission *EPILEX*, ISBN 92-826-6211-X, Luxembourg, 1993.
19. R. Baud *et al.*, Modelling for natural language understanding. Symposium on Computer Applications in Medical Care, Washington DC, in *SCAMC*, **93**: 289–293, 1994.
20. A. Kong *et al.*, How medical professionals evaluate expressions of probability. *N. Engl. J. Med.*, **315**: 740–744, 1986.
21. P. L. M. Kerkhof *et al.*, The Medwise diagnostic module as a consultant. *Expert Syst. Applic.*, **6**: 433–440, 1993.
22. F. T. de Dombal, *Diagnosis of Acute Abdominal Pain*. Edinburgh: Churchill Livingstone, 1980.
23. SNOMED International, College of American Pathologists, Northfield, IL, 1993.
24. *Physicians' Current Procedural Terminology* (CPT), 4th ed. Chicago: American Medical Association, 1992.
25. D. A. Lindberg, B. L. Humphreys, A. T. McCray, The unified medical language system. *Methods Inf. Med.*, **32**: 281–291, 1993.
26. E. R. Gabrieli, A new American electronic medical nomenclature. *Automedica*, **14**: 23–30, 1992.
27. J. Chisholm, Read clinical classification. *Brit. Med. J.* **300**: 1092, 1990.
28. J. Wyatt, Computer-based knowledge systems. *Lancet*, **338**: 1431–1436, 1991.
29. W. G. Baxt, Use of an artificial neural network for the diagnosis of myocardial infarction. *Ann. Intern. Med.*, **115**: 483–488, 1991.
30. M. S. Blois, *Information and Medicine*, Berkeley: University of California Press, 1984.
31. G. O. Barnett *et al.*, DXplain, an evolving diagnostic decision-support *J. Am. Med. Assoc.*, **258**: 67–74, 1987.
32. B. Middleton *et al.*, Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base, part II—Evaluation of diagnostic performance. *Meth. Inform. Med.*, **30**: 15–22, 1991.
33. H. S. Waxman, W. E. Worley, Computer-assisted adult medical diagnosis, subject review and evaluation of a new microcomputer-based system. *Medicine (Baltimore)*, **69**: 125–136, 1990.
34. E. S. Berner *et al.*, Performance of four computer-based diagnostic systems. *N. Engl. J. Med.*, **330**: 1792–1796, 1994.
35. B. Bergeron, ILIAD, a diagnostic consultant and patient simulator. *M.D. Comput.*, **8**: 46–53, 1991.
36. P. Krause *et al.*, Can we formally specify a medical decision support system: *IEEE Expert*, **8**: 56–61, 1993.
37. C. W. Bishop, P. D. Ewing, Framemed, a prototypical medical knowledge base of unusual design. *M.D. Comput.*, **10**: 184–192, 1993.
38. STAT!-Ref. Available from Teton Data Systems, P.O. Box 3082, Jackson, WY.
39. MD-Challenger, 70 Belle Meade Cove, Eads, TN 38028.
40. Labsearch/286, software available from REMIND/1 Corp., P.O. Box 752, Nashua, NH 03061.
41. M. E. White, J. Lewkowicz, The CONSULTANT database for computer-assisted diagnosis and information management in veterinary medicine. *Automedica*, **8**: 135–140, 1987.
42. H. W. Griffith, M. R. Dambro, *The 5 Minute Clinical Consult*. Philadelphia: Lea & Febiger, 1997.
43. M. L. Buyse (ed.), *Birth Defects Encyclopedia*. Center for Birth Defects Information Services, Inc., 30 Springdale Ave., Dover, MA 02030. BD Fax service (508) 785–2347.
44. DiagnosisPro by MedTech USA, P.O. Box 67543, Los Angeles, CA 90067.
45. D. A. Waterman, *A Guide to Expert Systems*. Reading, MA: Addison-Wesley, 1986.
46. A. J. Gonzalez, D. Dankel, *The Engineering of Knowledge-Based Systems: Theory and Practice*. NJ: Prentice-Hall, 1993.
47. J. Giarratano, G. Riley, *Expert Systems: Principles and Programming*. PWS Publishing Co., 1994.
48. P. Jackson, *Introduction to Expert Systems*. Reading, MA: Addison-Wesley, 1993.
49. S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, Englewood Cliffs, NJ: Prentice-Hall, 1995.
50. U. Gupta (ed.), *Validating and Verifying Knowledge-Based Systems*. NJ: IEEE Computer Society Press, 1991.
51. R. M. O'Keefe, O. Balci, E. P. Smith, Validating Expert System Performance. *IEEE Expert*, **2**: 81–89, 1987.
52. R. M. O'Keefe, D. E. O'Leary, Expert system verification and validation: a survey and tutorial. *Artif. Intell. Rev.* **7**: 3–42, 1993.
53. D. E. O'Leary, Verifying and validating expert systems: A survey, in *Expert Systems in Business and Finance: Issues and Applications*, P. R. Watkins and L. B. Eliot (eds.), Chichester, UK: John Wiley & Sons Ltd., 1993, pp. 181–208.
54. J. P. Kassirer, A report card on computer-assisted diagnosis—the grade: C (editorial). *N. Engl. J. Med.*, **330**: 1824–1825, 1994.

55. A. Alonso-Betanzos *et al.*, FOETOS in clinical practice: A retrospective analysis of its performance, *Artif. Intell. Med.* **1**: 93–99, 1989.
56. V. Moret-Bonillo, *et al.*, The PATRICIA project: A semantic-based methodology for intelligent monitoring in the ICU, *IEEE Eng. Med. Biol.* **12**: 59–68, 1993.
57. A. Alonso-Betanzos, *et al.*, Computerized antenatal assessment: The NST-EXPERT project, *Automedica*, **14**: 3–22, 1992.
58. A. Alonso-Betanzos, *et al.*, The NST-EXPERT project: the need to evolve, *Artif. Intell. Med.*, **7**: 297–314, 1995.
59. C. Hernandez, *et al.*, Validation of the medical expert system RENOIR, *Comput. Biomed. Res.* **27**: 456–471, 1994.
60. V. Moret-Bonillo, E. Mosqueira-Rey, A. Alonso-Betanzos, Information analysis and validation of intelligent monitoring systems in intensive care units, *IEEE Trans. Inf. Technol. Biomed.* **1**: 87–99, 1997.
61. G. Guida, G. Mauri, Evaluating performance and quality of knowledge-based systems: Foundation and methodology, *IEEE Trans. Knowl. Data Eng.* **5**: 204–224, 1993.
62. K. Clarke, *et al.*, A methodology for evaluation of knowledge-based systems in medicine, *Artif. Intell. Med.*, **6**: 107–121, 1994.
63. R. S. Ledley, L. B. Lusted, Reasoning foundations of medical diagnosis, *Science*, **130**: 9–21, 1959.
64. A. R. Feinstein, The haze of Bayes: The aerial palaces of decision analysis and the computerized ouija board. *Clin. Pharma. Therap.* **21**: 482–496, 1980.
65. R. O. Duda, P. E. Hart, N. J. Nilsson, Subjective Bayesian methods for rule-based inference systems, Proc. Nat. Comput. Conf., pp. 1075–1082, 1976.
66. E. H. Shortliffe, B. G. Buchanan, A model of inexact reasoning in medicine, *Math. Biosci.*, **23**: 351–379, 1975.
67. J. F. Lemmer, Confidence Factors, Empiricism and the Dempster–Shafer Theory of Evidence, in *Uncertainty in Artificial Intelligence*, L. N. Kanal and J. F. Lemmer (eds.), Elsevier, 1986, pp. 117–126.
68. K. Ng, B. Abramson, Uncertainty management in expert systems, *IEEE Expert*, **5**: 29–48, 1990.
69. L. A. Zadeh, The role of fuzzy logic in the management of uncertainty in expert systems, in *Fuzzy Sets and Systems*, 1983, pp. 199–227.
70. H. J. Zimmerman, *Fuzzy Set Theory and Its Applications*, Dordrecht, The Netherlands: Nijhoff, 1987.
71. G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.
72. L. A. Zadeh, Fuzzy sets, *Inf. Contr.* **81**: 338–353, 1965.
73. L. A. Zadeh, Knowledge representation in fuzzy logic, *IEEE Trans. Knowl. Data Eng.* **1**: 89–100, 1989.
74. E. Mosqueira-Reyand V. Moret-Bonillo, Validation and Usability Analysis of Intelligent Systems: An Integrated Approach *The IPSI-BGB Transactions on Advanced Research*, vol. 1,pp. 37–47, 2005.
75. M. Ivory, M. Hearst, The State of the Art in Automated Usability Evaluation of User Interfaces. *ACM Computing Surveys*.Vol. 33, No. 4, 2001, pp. 173–197.
76. L. Adelman, S. L. Riedel, *Handbook for Evaluating Knowledge-Based Systems*, Kluwer Academic Publishers, Boston, 1997.
77. J. Preece, *A Guide to Usability: Human Factors in Computing*, Addison-Wesley, Reading, MA, 1993.
78. S. K. Card, T. P. Moran, A. Newell, *The Psychology of Human Computer Interaction*. Erlbaum, Hillsdale, NJ, 1983.
79. J. Nielsen, Heuristic Evaluation, in: Nielsen, J. Mack, R. L. (Eds.). *Usability Inspection Methods*. John Wiley & Sons, New York, NY, pp 25–62 1994.
80. W. Edwards, J. R. Newman, Multiattribute Evaluation. Sage University Paper series on Quantitative Applications in the Social Sciences. Sage Publications, Newbury Park, CA, 1982.
81. T. L. Saaty, *The Analytic Hierarchy Process*, McGraw-Hill, New York, 1980.
82. S. Goonatilake, S. Khebbal, *Intelligent Hybrid Systems*. John Wiley and Sons, 1995.
83. B. Guijarro-Berdiñas, A. Alonso-Betanzos, O. Fontenla-Romero. Intelligent analysis and pattern recognition in cardiotocographic signals using a tightly coupled hybrid system. *Artificial Intelligence*, Vol. 136, N° 1, pp 1–27, 2002.
84. V. Aijanagadde, Venkat and L. Shastri. Rules and Variables in Neural nets. *Neural Computation*, Vol. 3, N° 1, pp. 121134, 1991.
85. G. Shreiber, H. Akkermans, A. Anjewierden, R. De Hoog, N. Shadbolt, W. Van de Velde and B. Wielinga. *Knowledge engineering and management*. The CommonKADS methodology. Third Edition. MIT Press, 2002.
86. J. Angele, S. Dekel, R. Perkuhn, and R. Struder. Developing knowledge-based systems with MIKE. *Journal of Automated Software Engineering*, Vol 5, n° 4, pp. 326–389, 1998.
87. Protégé. http://protege.stanford.edu/index.html.accessed November, 20th- 2006.
88. *IEEE Intelligent Systems.* Special issue on The Semantic Web: A brain for humankind. Vol 16, Issue 2, 2001.
89. M. S. Blois, *Information and Medicine*. Berkeley: Univ. California Press, 1984.
90. P. L. M. Kerkhof, M. P. van Dieijen-Visser (eds.), *Laboratory Data and Patient Care*. New York: Plenum Press, 1988.
91. E. A. Murphy, *The Logic of Medicine*. Baltimore: Johns Hopkins Univ. Press, 1976.
92. J. A. Reggia, S. Tuhrim (eds.), *Computer-Assisted Medical Decision Making*. New York: Springer, 1985.
93. E. H. Shortliffe, Computer programs to support clinical decision making. *J. Am. Med. Assoc.*, **258**: 61–66, 1987.
94. E. H. Shortliffe, L. E. Perrault (eds.), *Medical Informatics: Computer Applications in Healthcare*. Reading, MA: Addison-Wesley, 1990.
95. R. H. Taylor, S. Lavallee, G. C. Burdea, R. Mosges (eds.), Computer-Intergrated Surgery. Cambridge, MA: MIT Press, 1995.
96. B. T. Williams, *Computer Aids to Clinical Decisions*. Boca Raton, FL: CRC Press, 1982.
97. J. L. Bowen: Medical education; educational strategies to promote clinical diagnostic reasoning. *New Engl. J. Med.* **355**(2006) 2217–2225.
98. Alphabetical list of disease names and corresponding ICD-9-CM: http://www.dmi.columbia.edu/hripcsak/icd9/2indexa.html or: http://icd9cm.chrisendres.com/index.php?action=contents
99. Public Health Image Library: http://phil.cdc.gov/phil/whatsnew.asp
100. Software for PDA: http://www.epocrates.com/index.html
101. http://www.merckmedicus.com/pp/us/hcp/templates/tier2/PDAtools.jsp
102. http://www.skyscape.com/index/home.aspx

103. Differential Diagnosis:DXplain: http://www.lcs.mgh.harvard.edu/
104. Gideon: www.gideononline.com
105. Iliad: http://premieremedical.safeshopper.com/276/1904.htm
106. Scientific and Technical Acronyms, Symbols, and Abbreviations: http://www3.interscience.wiley.com/cgi-bin/mrwhome/104554766/HOME

PETER L. M. KERKHOF
Medwise Working Group
Maarssen, The Netherlands

AMPARO ALONSO-BETANZOS
VICENTE MORET-BONILLO
University of Coruna