Knowledge in IT – absolutely crucial, mostly ignored

L. Wolfgang Bibel

October 18, 2006

Knowledge has always been of central importance for man and mankind. Man's ability to create and keep knowledge in his memory and share it with others is among the important features which distinguish man from animals. Communication, education, social and political human life all are unthinkable without knowledge. Problem solving, man's most important activity for survival, relies to a large extent on our available knowledge, be it on the personal, the social, the national or even on the global level. So there is no doubt that knowledge deserves the highest level of priority in all our considerations.

It is generally agreed that we have entered a new period in the history of mankind characterized by an explosion in the amount of information being processed. Consequently we are speaking of the *information society* or *knowledge society* or of the *information age*. As with previous historical periods also this one was triggered by technological change. The invention of the universal computers in the middle of the last century has enabled a technological development of an unprecedented magnitude and speed. The new technological basis established by this development is collectively called *information technology* or *IT* for short.

The exponential growth of IT is really unimaginable. It can be illustrated with many hard figures. For instance the speed of processors, the basic modules in computational devices, doubles every 18 months according to what is called Moore's law. It did so for more than five decades and there is no indication whatsoever that this exponential speedup will slow down in the near future. Similar – in fact still higher – growth rates are experienced with the amounts of information which can be stored on memory devices and which can be transmitted through information transport channels, to mention just two further ones. The amazing observation thereby is that these technologically improved devices are available for even less money rather than more in comparison to the price for their older versions.

Especially since the introduction of the technology underlying the world wide web (WWW) about a decade ago IT has really become part of every-day-life of the majority of people in industrial countries like Germany. In many cases we are not aware any more of the many computational devices which accompany our behavior. For instance driving a higher class car activates more than a hundred of processors built into it. For this reason we are talking now of the phenomenon of *ubiquitous computing*.

So both knowledge and IT pervade our daily lifes. Is this mere coincidence or have they something to do with each other? They do, of course. For instance, there is a huge amount of knowledge coded into the technological devices of IT. For instance, a technological protocol realizing the transport of my email from one node to the next, consists of a cleverly coded piece of sophisticated knowledge. But so does my coffee machine although no one would associate any special relationship between coffee machines and knowledge. So let us rather ignore the knowledge coded into devices for the purpose of this talk.

Further, if I share some knowledge with a friend through email, IT realizes the transportation. But the same is done by a telephone switching system which simply connects telephones and has no access to the knowledge in the conversations transported through it. Finally, IT in the form of the WWW makes knowledge scattered all over the world readily available for access anywhere in the world. It is this fact that IT processes huge amounts of knowledge every day worldwide which has led to the recent introduction of new fields within IT termed *knowledge management* (or similarly). Knowledge management systems (KMS) under study and development in these fields manage the storage, the retrieval and the control of the flow of knowledge coded in some way (eg. within a language called XML which structures natural text, pictures and so forth with sophisticated markings). But is it really "knowledge" that is managed by a KMS? Is it not rather strings of symbols or pixels the semantics of which is totally inaccessable to the KMS?! As we see there is actually no knowledge in a KMS. There is indeed no knowledge in most of IT except in a form which makes it inaccessible for the systems involved.

This observation and the unfortunate consequences of this fact are the topic of this talk. Our main thesis is that the revolution triggered by IT will not really take place and contribute to the betterment of mankind unless knowledge is activated explicitly within the IT devices comparable to the way it is activated in humans. In the remainder of this talk I will make more precise what knowledge is, how it can be represented in IT devices and how those devices enhanced in this kind would support and better society in genuine ways.

Knowledge is a term denoting a psychological concept. Humans are embedded into their environment similarly as animals are. Both can move around and survive. But only humans can form a model of the environment which they can even share with others. For instance I could describe you my garden to a point where you would have a vivid imagination of how it looks like and, in passing by, would even recognize it among other gardens. This model in my brain is knowledge, more specifically empirical or scenic knowledge in this particular case. Similarly I know how to drive a bike and could explain details about this activity to you. From scenic and factual knowledge we abstract (or induce) generic knowledge like "*birds can fly*" or even abstract (or conceptual) knowledge like Pythagoras' theorem or some protocol algorithm. A key feature of all this knowledge is that it is built on top of a complex structure of concepts.

Knowledge is an ingenious product of evolution. Out of the zillions of bits of information continously hitting our sensors the brain compresses this unstructured mass into comparably small pieces. Over more than a hundred thousand years the pieces have been selected such that the chances for survival improved.

At this point science has little evidence for how knowledge is actually produced in the brain and how it is stored and processed there. We do have however ample evidence of a phenomenological kind. Namely, evolution has associated with knowledge the phenomenon of language (in a broad sense of the word which besides natural language includes formal languages, expressions of art, gesture and so forth). So as we study language, we study knowledge at the same time. Language refers to knowledge internal to the brain which in turn refers to the respective scenes in the world outside. This transitive relationship we call the semantics of language which is learned in the childhood by behavioral imitation.

Besides knowledge evolution holds a further principle of compression in store. If you recognize an object as a bird you automatically know many things about the object without further information flow. The mechanism behind this automatism we call logic. Like knowledge it is a product of evolution enhancing survival chances because the amount of information exchanges could thus be further reduced without sacrificing common understanding. Logic is a fundamental aspect of knowledge; in other words, there is no knowledge without logic (and, of course, no logic without knowledge). Like with language there is a syntactic and a semantic side of logic. The syntactic side is directly accessible for us through a study of language. This study has been going on in philosophy with moderate success for more than two thousand years, starting with Aristotle at the latest.

The IT revolution has speeded up these studies of language and logic to an unprecedented extent. Characterized by a totally new formal and experimental approach to the study of mind and cognition this led to the foundation of the new discipline of Intellectics which comprises (the more technical part of) Artificial Intelligence and Cognitive Science(s). For the first time in human history it is now possible, for example, to represent a huge body of human knowledge in such precise terms as to become processable by a computer. The CYC system, for example, developed under the leadership of Doug Lenat already comprises close to a million of chunks of knowledge.

Logic plays an important role in these activities, although in modifications which sometimes go so far that logicians from mathematics might not even recognize the logical structure anymore. This is because issues of representation have turned out to be of utmost importance for the computational processes involved, an issue not recognized before in logic. It is also because the logic considered more or less fixed until recently might not exactly model the logic involved in human reasoning. For instance, it is still not exactly clear how to cope in the standard logics with the nonmonotonic phenomena involved in human reasoning.

The enterprise to formalize a great deal of human knowledge has also made it clear how important an adequate and precise basis of concepts is. Intellecticians started therefore to build large ontologies of concepts with a degree of precision which obviously is a necessity for computer processing. Vague philosophical definitions like "*Wissen ist die Teilhabe am Sosein eines Seienden*" (Max Scheler, 1874–1928) from now on should disappear in the paper basket of the scientific enterprise. We all have to plead guilty in this respect. The difference between philosophers and intellecticians is, however, that the latter are backed in their vagueness by precise codings running on machines. Schr"oder whose death a hundred years ago we commemorate today is among the pioneers whose efforts towards mathematical precision of concepts contributed to the advances achieved so far.

A standard but misleading argument against a far-reaching formalization of knowledge is that all our concepts are intrinsically vague and carried by the context, and because of that nature can never be made fully precise. While human concepts are of course vague, we still can define them rigorously *along with the kind of vagueness* adhered to them and by making the respective context explicit.

A major obstacle for the progress of computational logic is the Babylon of logical formalisms which entered the stage in the last decades and mostly model one and the same logic. For instance in my course on Automated Deduction I show the students tens of different formalisms just for propositional logic which all model exactly the same basic idea of modus ponens type inference. Since formalisms are not easily accessible to humans the consequence is that two researchers working with different formalisms are often unable to communicate. So they reinvent old computational ideas again and again. For that reason I have become an advocate for the promotion of a suitable canonical form.

Apart from this communicational problem there are still challenging real technical problems in the representation of knowledge and its computational treatment. I already mentioned the phenomena of nonmonotonicity around human's ability to "jump to conclusions" where systems tackle hard computational problems instead. A fully satisfactory integration of knowledge about dynamic processes within the description-oriented and static logics without creating the so-called frame problem still remains another challenge. Similarly with the complex context issues. Also the deep relationship between proving on the one hand and computing on the other still waits for its full exploitation for the synthesis of programs from specifications. But these and other challenge problems are no excuse at all to keep us from implementing knowledge within IT devices explicitly. Why would that be desirable in the first place?

We have seen above that similarly as in a telephone switching system there is no knowledge explicitly built into a KMS. Why should this be ok for the telephone system, but not for the KMS? The telephone system connects people and does not interact with their communication in any meaningful way. If I "enter" the internet, my communication partners are systems no more people. Above we have pointed out the fundamental importance for communication of knowledge expressed in (natural) language and of logic. Communication in the internet lacks all these three fundamentals nor are they part of other IT devices (with a few exceptions). So the IT community pretends to provide natural interfaces to their systems which however lack the fundamental ingrediences for their realization. Also IT is penetrating vital areas of human and social life but still ignores those fundamentals which have formed the structure of this life for thousands of years.

True, IT had its romance with knowledge-based (or expert) systems in the seventies and eighties. Then, the companies wanted quick results in terms of profits which could of course not be delivered by the scientists within a few years given the research challenges involved and the machinery available then. A quarter of a century later there is now a huge stock of research results available which along with the spectacular technological progress in IT in general again raises the hopes for a promising realization of these old dreams.

Namely, we now could progressively refine the ontologies to cover any of the areas of scientific and general interest. On such a conceptual basis knowledge bases could be extended to cover all these areas. They would of course have to be maintained on a continuous basis and the knowledge in them would have to carry an identification of the responsible authorizing body. Also the technology is now there to associate with these huge bodies of knowledge effective logical reasoning systems (including deductive, inductive, abductive, analogical, fuzzy etc. reasoning modes). As a consequence the interaction with such systems would mimic the interaction among people as described above. However, in contrast to people, knowledge systems would not suffer from the many weaknesses experienced in our human handling of knowledge. These weaknesses are no problem in our daily personal interactions. But they become more and more intolerable in a global world where small errors and failures tend to have ever more fatal consequences.

In a recent book I have laid down the vision of the application of such systems to a number of areas in our societies. Let me mention two of these areas here in order to give an idea of the revolutionary benefits that might come out of such developments. One refers to public problem solving, the other to jurisprudence.

We all have experienced in university committees and elsewhere the frustrations generally associated with the public problem solving process. Not only in communities but also in states its inherent weakness leads regularly to irrational and faulty solutions even in cases where the underlying political attitudes are of no relevance for the matter. The simple reasons for this utmost unfortunate situation lies in the lack of relevant knowledge for those participating in the decision process and in the inferential mistakes made during their deliberations. In consequence convincing arguments remain unheard and are rather substituted by ideological statements which are repeated at every possible occasion, and winning are solutions which for various reasons might bear the promise of advantages for groups of interest.

Imagine we are faced with some such problem. Think of the organisation of the teaching schedule as to minimize time conflicts for students, or of coping with a daily traffic jam in some part of a city, or of the problem whether or not to provide permission for the construction and establishment of a new supermarket, or of what have you. Assume we first collect all knowledge pertaining to our problem as well as all opinions, assumptions and arguments of the people involved. It is not hard to imagine that from this body of knowledge the true available options for a solution could be extracted in an absolutely rational way, each along with the underlying assumptions and preferences. Analysed in this way the remaining rational choices, determined by the different assumptions and preferences, could now be made in a truly democratic way whereby we should use more sophisticated voting mechanisms like the STV system. The outcome would not only be guaranteed to a large extent to be free of logical mistakes but also would find the approval of the majority of participants in all details. This scenario is not wishful thinking but rather lies within the reach of current IT technology.

While we all agree that the constitutional state still provides the best political framework, it is clear that it has not produced a satisfactory level of justice in our states. Well-known problems with our legal system are: (i) the flood of laws, each with numerous modifications, so that no one has a clear idea what is legal and what not; (ii) the lack of public backing as many of the laws do in their details not reflect public opinion at all as the details mostly go through unnoticed by the members of parlament; (iii) the complexity of even the simplest legal cases so that no one can predict the outcome of such a case which in fact often differs substantially at the various levels of the judical process; (iv) the resulting duration of such processes which again bears injustice simply by the years-long delay in obtaining "justice".

Imagine again all legal concepts are formalized (or "reengineered") the way discussed above and laws are phrased at a level of precision as the knowledge implemented in CYC. Not only would the inconsistency of laws be detectable, but we could also solve legal cases with respect to their chances preliminarily in seconds or hours rather than in years. Even the making of laws should be supported by IT. In order to comply with the judgment of people these would offer their decisions in particular cases. Based on thousands of individual judgments inductive mechanisms could induce rules at an abstract level which in turn could be phrased in normal language. If the body of law would be constructed in such a systematic way, we could even reach a level of abstraction so that the law could be condensed to a few general principles with which the rest is logically connected by precisely defined operations of instantiation. So everybody would be able to know about his or her rights and obligations – in sharp contrast to the current situation. Again such a scenario is within the reach of current IT technology, the more so as the revolution of law in such a way would certainly take more than a decade for its implementation, enough time for further major breakthroughs in technology.

Note that I am not favoring here a full automation neither of the problem solving process in society nor of the law making and the jurisdiction. Rather this proposal envisions these activities without the substantial flaws inherent in their present versions. Given these beneficial prospects for such fundamental issues of our societies all disciplines with a share in the grand vision of Intellectics are called for their contributions towards these goals.

Some people are worried about this "way out" which amounts to exporting more and more human capabilities into systems. The global society has no other choice though, simply because any problem beyond the personal sphere has become too complex for men's brains to be feasible without technical support. Complementarily we need to intensify the "way in" as far as our personal lifes are concerned. But the discussion of this latter aspect would be the topic of a different talk.