Is KS=(D+I+S+K)*E + KM?

Juan Llorens
Computer Science Department
Universidad Carlos III de Madrid
Leganés, Madrid, Spain
llorens@inf.uc3m.es

Rubén Prieto-Díaz
Commonwealth Information Security Center
James Madison University, Harrisonburg, VA
prietodiaz@cisat.jmu.edu

Abstract

Is it time for a science of knowledge? In this paper we argue that yes, it is time and define it as:

KS = (D+I+S+K)*E + KM

Where KS is Knowledge Science, DE is Domain Engineering, IE is Information Engineering, SE is Software Engineering and KE is Knowledge Engineering, and that Knowledge Management (KM) moderates these disciplines. We try to clarify the apparent state of confusion in the current landscape where KS, KM and KE have different meanings to different audiences.

After clarifying the difference between KE and KM and arguing that KE and KM are complementary we explain why we selected the word *knowledge* instead of information and decided on the word *science* in KS. The core of the paper explains the term (D+I+S+K)*E in our proposed definition and why each of these disciplines contribute to KS. One of our conclusions is that KS deals with knowledge about knowledge, or about the knowledge needed to increase, expand or further acquire more knowledge, and that KS can be applied to all the existing sciences, covering their needs of methodologies, techniques and tools.

Is it time for a Science of Knowledge?

We argue that yes, it is time to create Knowledge Science (KS) and define it as the merging of several disciplines.

KS = (D+I+S+K)*E + KM

We propose defining KS as the merging of Domain Engineering (DE), Information Engineering (IE), Software Engineering (SE) and Knowledge Engineering (KE), all of them "moderated" by Knowledge Management (KM).

Our first task in explaining the above equation is to clarify the difference between KE and KM. When facing problems where computers are required to deal with knowledge, engineers must choose whether to follow KM or KE: are these fields different? If so, which one should I use? We believe both are needed if we want to deal effectively with knowledge.

The concept of KE can be traced back to the late 1950's [1], when Artificial Intelligence (AI) brought it up to the computer science scene. AI experts wanted to emphasize that "general-computer-based-problem-solving" was not any longer an illusion but a possible reality. State spaces from game theory, production rules, semantic networks, frames, and neural networks have all contributed to materialize the kernel of knowledge engineering.

Although it seems clear what scientists understand as KE, why is it different from KM? AI engineers, after all, model and manage knowledge in pursuing their AI engineering solutions. Some clarification is in order. We need to address the following two questions: What is engineering? And what is knowledge?

According to the Accreditation Board for Engineering and Technology (ABET)[1] *"Engineering is the profession in which a knowledge of the mathematical and natural sciences, gained by study, experience, and practice, is applied with judgment to develop ways to utilize, economically, the materials and forces of nature for the benefit of mankind"*. The Merriam-Webster Dictionary[2] defines engineering as *"the application of science and mathematics by which the properties of matter and the sources of energy in nature are made useful to people"*. Although other definitions can be found[3], most converge to the same concepts.

It seems from these definitions of engineering that what AI engineers were doing was managing rather than engineering knowledge. But before drawing any conclusions, knowledge needs to be defined.

According to the Merriam-Webster Dictionary and the Cambridge Dictionary of English[4], knowledge can be defined as *"understanding of or information about a subject which has been obtained by experience or study, and which is either in a person's mind or possessed by people generally"*, "*acquaintance with or understanding of a science, art, or technique*" and "*the fact or condition of having information or of being learned*". The Free On-Line dictionary of computing[5] provides a somewhat more specific definition of knowledge. *"The objects, concepts and relationships that are assumed to exist in some area of interest"*. To better understand these definitions it is necessary to define information.

Information is defined by the Merriam-Webster Dictionary as *"the equivalent of or the capacity of something to perform organizational work, the difference between two forms of organization or between two states of uncertainty before and after a message has been received, but also the degree to which one variable of a system depends on or is constrained by [..] another."*

Can it be concluded from these definitions that knowledge is learned information? Or that knowledge is a particular kind of information? If this is the understanding, then should AI's KE be called Information Management or Knowledge Management? It seems that we have further blurred rather than clarified the difference between KE and KM.

It is necessary then to provide another viewpoint by examining the present understanding of KM. In the last decades people like Drucker [2], Strassman [3] and Senge [4] introduced the concept of KM. A definition of KM as understood today may require

[1] See http://www.abet.org
[2] See http://www.m-w.com
[3] See http://civil.engr.siu.edu/intro/definitions.htm
[4] See http://dictionary.cambridge.org
[5] See http://wombat.doc.ic.ac.uk/foldoc/index.html

more text than it is allowed in this article. Partitioning the definition, however, may help.

We see two main orientations, the academic and the commercial. A search in the web[6] regarding Knowledge Management will lead the user to several commercial companies interested in selling services where their definitions of KM, if any, strongly depend on their provided services. We selected some in addition to others from academia.

From academia, the AIAI of the University of Edinburgh[7] describes: "*Knowledge management involves the identification and analysis of available and required knowledge assets and knowledge asset related processes, and the subsequent planning and control of actions to develop both the assets and the processes so as to fulfill organisational objectives*" where "*knowledge assets are the knowledge regarding markets, products, technologies and organisations, that a business owns or needs to own and which enable its business processes to generate profits, add value, etc. Knowledge management is not only about managing these knowledge assets but managing the processes that act upon the assets.*" W. Newman, from the Johns Hopkins University[8], describes KM as "*...an integrated approach to identifying, capturing, retrieving, sharing and evaluating an enterprise's information assets. These information assets may include databases, documents, policies, procedures, as well as the uncaptured tacit expertise and experience stored in individuals' heads." It also includes content management, best illustrated in the "traditional or industrial age library," as well as process flows in automated environments*".

From the commercial point of view, IT Consultancy[9] defines KM as *"...a conscious strategy of getting the right knowledge to the right people at the right time and helping people share and put information into action in ways that strive to improve organizational performance. Knowledge management is a complex process that must be supported by a strong foundation of enablers. The enablers for KM are strategy and leadership, culture, measurement, and technology. Each of these must be designed and managed in alignment with the other and in support of the process*

From the previous definitions we can understand that, even if both KE and KM certainly manage knowledge, they are focusing on different and complementary aspects of it.

KE intends to "dominate" knowledge in order to allow computers help humans solving problems [5]. KM is more of an enveloping discipline that deals with several areas applying a variety of techniques, with the main intention to "administratively manage" knowledge. It is this complementary viewpoint that we want to emphasize in our explanation of the proposed concept of KS.

Knowledge Science

It is important first to reassess our choice of terms. Why knowledge and not information? If we accept that knowledge is learned information, we suggest that one way to transform information into knowledge in computer science is by normalizing information words into normalized terms (e.g. all possible lexical representations for the concept of vehicle should be normalized to the term "Vehicle" which could represent that concept). By normalizing, and assuming the hypothesis that for every term there is one concept, a computer could be able to map terms into concepts. Therefore, if all the information stored in a computer regarding a particular subject is controlled by just one unique "subject identifier" called a normalized term, we could think that the computer handles learned information: knowledge.

Why science? According to the Merriam-Webster Dictionary, science is "*the state of knowing: knowledge as distinguished from ignorance or misunderstanding*", and for the Cambridge Dictionary of English science is "*knowledge obtained from the systematic study of the structure and behaviour of the physical world, involving experimentation and measurement and the development of theories to describe the results of these activities*". Since science is related to knowledge or its acquisition, KS then deals with knowledge about knowledge, or about the knowledge needed to increase, expand or further acquire more knowledge.

To practice KS we need techniques, methods and tools that support the activities of increasing our knowledge about knowledge. These activities, which include: identify, represent, analyze, update, collect, organize, share, use and adapt knowledge, must be covered by KS. We argue therefore that KE, SE, IE, and DE support these activities, and that KM catalyzes and moderates their interaction in addition to providing the medium where KS can be applied.

Knowledge Engineering (KE)

In Knowledge Science we must use elicitation techniques to identify knowledge, we have to provide representation techniques for storing expertise, we need to apply specific algorithms and methods for solving problems, we need to apply data mining techniques for analyzing knowledge, and we still have to provide expert systems to use knowledge. KE provides all of the above.

Software Engineering (SE)

It is impossible to understand knowledge representation without the modern software representation methodologies, like UML[10], OPEN, etc., especially when trying to represent software artifacts. UML provides a standard way to represent the knowledge structure, dynamics and architecture, that can be applied not only to information systems design, but to model any artifact, either material or abstract. SE also provides technology for knowledge update: without change control policies it is not possible to conceive a holistic approach to organization-based knowledge management. SE provides knowledge reuse techniques; essential for uniform knowledge organization. They cover, from organization-based best practices (maturity models, quality assurance, software and information economic models) to all the human factors involved (leading implications, new roles in the organization, incentives to reuse etc.). Finally, adapting knowledge is essentially the same as reusing knowledge. KFR-KWR (knowledge for reuse / knowledge with reuse) techniques, therefore, become relevant for knowledge scientists.

---

[6] At present time, almost year 2002

[7] See http://www.aiai.ed.ac.uk/~alm/kamlnks.html

[8] See Wilda B. Newman The Johns Hopkins University http://www.infotoday.com/cil2000/presentations/newman.ppt

[9] See http://www.it-consultancy.com/extern/apqc.html

[10] See http://www.uml.org and OMG Unified Modeling Language (2001), Specifications V 1.4. Semantics. Accessible at: http://www.omg.org/technology/documents/formal/uml.htm

Information Engineering (IE)

In modeling and representing knowledge into computers we are creating simplified representations of complex and abstract concepts. This is however, a theoretical utopia because it is impossible to represent information using concepts as they mean abstract ideas. To get around this problem, IE experts suggest a mapping between concepts and terms, using the latter as knowledge descriptors. Furthermore, IE provides advanced models such as Vector Space, Boolean and Probabilistic to represent knowledge. KS needs the IE approach as a solution framework to representation problems.

In addition IE has been trying to identify relevant information in textual documents for many years. Its models, techniques and algorithms (tf-idf, n-Grams, Kohonen's auto-organized maps, co-wording algorithms, etc.) can be used in KS for identifying certain kinds of knowledge. IE also provides indexing technology for human text artifacts that can be used by KS for analyzing knowledge. Recent application of Natural Language Processing (NLP) to text analysis in the areas of language translation and conversation generation is certainly improving (see http://www.cs.columbia.edu/~acl/). KS also needs information retrieval techniques for finding knowledge (relevance, pertinence, precision, recall, noise, silence, etc.). Librarians as knowledge classifiers are also essential to KS. They are in charge of organizing incoming knowledge into the representation systems.

Domain Engineering (DE)

Knowledge identification and extraction are two key activities when computer systems deal with knowledge. Knowledge identification is a key practice in DE. Domain engineers use the most advanced techniques for identifying general models or design patterns. Discovering frameworks, structural or behavioral patterns, static semantic networks or possible product lines certainly need domain analysis methods and techniques as well as computer systems that support them. Identifying commonalities and variabilities that represent a domain is at the core of knowledge identification, and form the basis for knowledge reuse. DE makes organizations aware of the tasks needed for knowledge identification and domain engineers take care of the continuous creation of knowledge for reuse.

DE also is behind knowledge representation fundamentals. Although IE provides KS with the support for concept representation using terms, DE provides some of the techniques that make possible such representations. Semantic noise must be filtered out for a computer to address knowledge using terms [6]. Domain analysis, which is a component discipline of DE, uses a filtering technique based on applying a restrictive one-to-many mapping between concepts and terms. This constraint implies the application of a controlled vocabulary representing a particular domain, where the possibility for a term to represent different concepts is strictly limited. The generation of a domain vocabulary is a key outcome of domain analysis.

Knowledge Management (KM)

KM's goal has been to create technology that allows organizations to use knowledge as a valuable corporate asset [7]. It is essential for KS to include the particular vision today's KM has in the organization aspects of knowledge: "Lets create a model to know what we do know". Certainly all this technology must be incorporated in the modern vision, as to use knowledge is one of the most important reasons why organizations should get into knowledge management[11]. KM is therefore the essential ingredient that puts KS into practice and keeps the scientific process cycle of observation, formulation, experimentation and validation alive.

We hope to have provided enough arguments and evidence to justify the definition of knowledge science as:

$$KS = (D+I+S+K)*E + KM.$$

Are we excluding anything?

Knowledge science, as we conceive it, is certainly multidisciplinary and of universal application. KS shares its basic principles with several other disciplines such as Philosophy, Logic, History of Science, Cognitive Science, Psychology, etc. In fact with all the sciences.

The universal application of KS is based on the general conception of science. The scientific process deals with the basic principles of observation, discovery, theory formulation, experimentation and verification. KS, as we propose it above, is intended to cover them and, therefore, it is applicable to any science.

The same techniques supported by the (DISK)E structure, for example, can be applied to any of the disciplines named above. Knowledge elicitation is used in modern Psychology; Cognitive Science uses information models to explain how the brain works; and Logic and Philosophy use ontologies as basic models for formalizing concepts [8].

In addition, every science in the science domain, requires support for engineering knowledge and for managing knowledge. KE and KM as discussed above can be directly applied to any scientific endeavour.

In our proposed equation we are not excluding other sciences but, on the contrary, we are proposing a KS that is general and applicable to any science.

Conclusions

Given the current landscape where knowledge science, knowledge management and knowledge engineering have different meanings to different audiences resulting in some state of confusion this paper attempts to clarify their meanings and proposes a definition for KS that integrates several areas that deal with knowledge and information.

An outcome of this analysis is the idea that KS is knowledge about knowledge, or about the knowledge needed to increase, expand or further acquire more knowledge.

We suggest, in addition, that knowledge engineering and knowledge management don't compete; they are complementary. And, together with other knowledge-based areas, they form a powerful framework for computing knowledge. This framework is what we call Knowledge Science and can be expressed by this equation:

$$KS = (D+I+S+K)*E + KM.$$

[11] Together with knowledge reuse.

Finally, KS, as it deals with knowledge about knowledge, can be applied to all the existing sciences, covering their needs of methodologies, techniques and tools.

References

[1] McCarthy, J. (1960). Recursive functions of symbolic expressions and their computation by machine, part I, *Communications of the ACM*, **3**(4):184-195, April, 1960.

[2] Drucker, P. F. (1973) *Management--Tasks, Responsibilities, Practices*, Harper and Row, New York.

[3] Strassman, P. (1985). *Information Payoff*, Free Press, New York

[4] Senge, P. (1990). *The fifth discipline: the art and practice of the learning organisation*. New York: Doubleday

[5] Lawton, G. (2001). Knowledge Management: Ready for Prime Time? *IEEE Computer,* **34**(2):12-14, February, 2001.

[6] Maron, M.E. Kuhns, J.L. (1960) On relevance, probabilistic indexing and information retrieval. *Journal of the Association for Computing machinery*, **7**(3):216--244, July 1960.

[7] Larson, L., Nidiffer, K., Rose, L., Small, R., and Stankosky, M. (2001) Knowledge Management: Insights from the Trenches, *IEEE Software,* **18**(6):66-68, November/December 2001

[8] Gruniger, M. & Lee, J. (2002) Ontology Applications and Design. *Communications of the ACM*, **45**(2):39-41, February, 2002.