

GÖRÜŞLER / OPINIONS

On Knowledge Organization Bilginin Düzenlenmesi Üzerine

Alan GILCHRIST*

Abstract

It is argued that because knowledge is abstract and every person has a unique perception of his environment and the properties and behaviour of its components, it follows that those people engaged in Knowledge Organization (and less directly Knowledge Representation) must base their work on physical records, which we may call carriers of information, or messages. The products based on analysis of these messages can then be considered as models of knowledge. Models are created in order to reduce complexity and to gain a clearer understanding of aspects of the world around us, but they must be continuously tested and revised in a working environment. The testing of the products of Knowledge Organization is often carried out by information scientists in their provision of information retrieval, whereas while the products of Knowledge Representation also rely on Knowledge Organization, they may be considered, to some extent, to be self-testing. It follows that much can be gained by a closer collaboration between those engaged in Knowledge Organization, Knowledge Representation and various other information professionals engaged in delivering information to end users.

Keywords: Knowledge organization, Representation, Knowledge models, Information retrieval, Knowledge organization systems

Öz

İleri sürüldüğü gibi, bilgi soyut bir kavram olduğuna göre, her kişi kendi çevresini, bu çevrenin özelliklerini ve bütünün parçalarını oluşturan davranışları, kendine özgü ve benzeri olmayan bir biçimde algılar. Sonuç olarak, Bilginin Düzenlenmesi (Knowledge Organization), ve doğrudan ilgili olmasa da, Bilginin Simgelenmesi (Knowledge Representation) alanlarında çalışanlar, bu uğraşlarını fiziki belgelere dayamak zorundadır: biz bu belgelere bilgi (enformasyon) ya da ileti (mesaj) taşıyıcıları adını veririz. Bu iletilerin analizleri sonucunda elde edilen ürünler, bilgi modelleri olarak kabul edilir. Modeller bizim için karmaşık durumları basitleştirmek ve çevremizdeki dünyanın farklı görünümünü daha iyi anlamamızı sağlamak için yapılır; modeller ise aynı zamanda çalışma ortamlarında sürekli olarak denenmeli ve yenilenmelidir.

Bilginin Düzenlenmesi sonucunda elde edilen verilerin sınanması, genelde bilgi bilimciler tarafından bilgiye erişimi sağlama sürecinde gerçekleşir; Bilginin Simgelenmesi sonucunda elde edilen veriler her ne kadar Bilginin Düzenlenmesi temeline dayansa da; bir noktaya kadar

* Cura Consortium and Metataxis Ltd. (alangilchrist77@gmail.com)

bunların kendi kendilerini sınadıkları kabul edilir. Bunun bir sonucu olarak denebilir ki, Bilginin Düzenlenmesi, Bilginin Simgelenmesi ve bilgiyle ilgili değişik alanlarda çalışan ve kullanıcılara bilgi sağlayan değişik bilgi bilim uzmanları arasında daha yakın işbirliği olmalıdır.

Anahtar sözcükler: *Bilginin düzenlenmesi, Bilginin simgelenmesi, Bilgi modelleri, Bilgiye erişim, Bilgi düzenleme sistemleri*

Introduction

The term Knowledge Organization, as discussed in this paper, was coined by the German classification expert Ingetraut Dahlberg as quoted by Peter Ohly, "...the science of structuring and systematically arranging of knowledge units (concepts) according to their inherent knowledge elements (characteristics) and the application of concepts and clusters of concepts ordered by this way for the assignment of the worthwhile contents of referents (objects/subjects) of all kinds" (Ohly, 2002).

This is now a working definition of the German International Society of Knowledge Organization (www.isko.org), founded in 1989 and now with national Chapters in Brazil, Canada + United States, People's Republic of China, France, Germany + Austria + Switzerland, India, Iran, Italy, Maghreb (Tunisia + Algeria + Morocco), Poland, Spain and the United Kingdom. The UK Chapter (www.iskouk.org) was formed relatively recently in 2007; has run two international conferences in London and has held afternoon meetings on a variety of topics including Interoperability, Linked Data (the future of Knowledge Organization on the Web). Geospatial information, and reviews of Knowledge Organization in the Legal, Health and Cultural heritage sectors, each of which has attracted upwards of 100 attendees.

Knowledge

The word knowledge has become popular in a number of areas including content management and organizational communication. Thus we have, in addition to Knowledge Organization (KO) and Knowledge Representation (KR), Knowledge Management, and Knowledge Engineering, but what do we, or can, mean by this word Knowledge? The study of Knowledge *per se* is Epistemology: "the theory or science of the method or grounds of knowledge" (Knowledge, 2012) This is a very ancient branch of philosophy stretching back at least as far as the ancient Greeks, and it was Plato who pronounced the formula "Knowledge is justified true belief" which lasted for some time before each of its component words was closely scrutinized by later philosophers till the "truth" of the formula was virtually demolished. Much of the subsequent debate has concentrated on such questions as "What does it mean 'to know'"; What can we mean by the term 'scientific knowledge' ?; Are there limits to what we can know? This last question inevitably leads to the thought that doubt must come after perceived knowledge, which is elaborated in the concept called "fallibilism", defined as "the philosophical principle that human beings could be wrong about their beliefs,

expectations, or their understanding of the world”, and that “any claim justified today may need to be revised or withdrawn in light of new evidence, new arguments, and new experiences” (Fallibilism, 2012). In this respect, the American physicist Lawrence Krauss has been quoted as saying: “Uncertainty is a central component of what makes science successful”. This constant state of flux concerning what we think we know forces us to consider what we can mean by the word knowledge in a practical and every day sense; and the only way we, in the information professions, can deal with it is to recognize that knowledge is held in the individual brain, and that shared knowledge is a consensus (inevitably a compromise) of shared perceptions. It follows that, strictly speaking, Knowledge Management is an impossibility (being an activity that rather focuses on managing the environment in which people communicate.) Knowledge Engineering is the term which seems to have replaced the earlier ‘Artificial Intelligence’, though with perhaps a slightly wider content, while KR is closer to KO and will be touched on briefly later in this paper. All these terms are using the word ‘Knowledge’ loosely, and by the same token we must also be careful what we mean by KO. Knowledge is abstract and shared either orally (and this is where Knowledge Management attempts to facilitate such transfer), or through written records which is the domain of the information sciences (IS). Nonaka and Takeuchi (1995), have convincingly made this distinction in their book *The knowledge creating company* in which they describe “tacit knowledge” as that which is stored in the brain and “explicit knowledge” as that which is recorded, (which we may fairly call ‘information’). They also, in what has become known as the SECI Model, show the four transformations between tacit and explicit knowledge:

- ◇ Tacit to Tacit – **Socialization**; a classical component of Knowledge Management
- ◇ Tacit to Explicit – **Externalization**: transfer of knowledge from minds to formal records
- ◇ Explicit to Explicit – **Combination**: creation of new records through collation
- ◇ Explicit to Tacit – **Internalization**: assimilation of knowledge from records.

Information

Though it has just been claimed that we can consider explicit knowledge to be the same as information in the normal sense of the word, we should continue to be careful with our choice of words and remember that one ‘item’ of information assimilated by one person does not have exactly the same meaning as it might have for another. We might, then, argue that the term ‘Information Management’ is also misleading; and, indeed, Miller (2012), has argued that “information has no intrinsic meaning”, distinguishing between meaning and ‘messages’, whether oral or written. Not only may interpretations vary according to the different mental ‘knowledge stores’ of the two individuals, but context and use also come into play, and as Myers and Myers say “...words don’t have meanings. There is no direct relationship between the thing you are talking about and the words you use. Only as these words are related through the thoughts of a person do they have meaning. Meaning is not in the object or in the

symbol but in the interaction of these through the human [communication process]" (Myers and Myers, 1998).

If, then, we should consider that our only approach to knowledge is through the recorded message there are some important implications. First, that messages are issued about very many and diverse sorts of information. So, in principle, exponents of KO and IS (principally librarianship, information science or documentation) should be concerned with facilitating access to records of 'many and diverse sorts of information', and should be prepared to process all messages where there is a demand for such messages to be retrieved. Second, it is not for us to dictate what are good or bad messages (though we may well have informed opinions about them). Floridi (2010), has proposed a tree structure of information concepts which includes "untrue information", which can be either Misinformation (unintentional) or Disinformation (intentional). Librarians and information scientists process messages to facilitate their retrieval, and only indirectly to provide access to knowledge, and consequently are bound to deal with 'untrue information' as well as "true information". When does true information become untrue – and vice versa? Are we not concerned, for example, with the history of science and what arguments were once put forward for geocentrism? Are we not obliged to record the latest findings from bio archaeological evidence correcting earlier notions of *Homo Sapiens*, Neanderthal Man and their interaction? We know that Beijing was once called Peking and that Slovenia was part of Yugoslavia, but we must update our maps and gazetteers, while maintaining the previous names for the records. The third point arising from the obligation to deal with messages is that, if we accept KO to be largely about the semantics of Structured Vocabularies (SV) and IS to be about the delivery of information, then it is suggested that there is an increasingly large overlap between the two activities, which implication will be discussed later in this paper.

Models

Returning to the concept of fallibilism, defined earlier, we may acknowledge that the only way to rise above the tricky ground of knowledge, information and meaning is to think in terms of 'approximations' and to create models which we create as approximations representing issues and problems that are too complex to understand in their entirety. Neil Gershenfeld, another American physicist, has been quoted as saying that "the most common misunderstanding about science is that scientists seek and find truth. They don't – they make and test models". Are not all 'mentefacts'* in some sense models? And if this is true, then our SV can also be considered as models 'approximating' to the contents of messages for which they are designed. Some support for this view may be found in the concept of 'pragmatic epistemology' enunciated by Heylighen, who says: "This philosophy still dominates most present work in cognitive science and artificial intelligence. According to pragmatic epistemology, knowledge consists of models that

* A word coined by the classificationist Barbara Kyle as an amalgamation of mental and artefact.

attempt to represent the environment in such a way as to maximally simplify problem-solving" (Heylighen, 1993). In this context Heylighen is thinking of KR which, as was said earlier in this paper, is similar in some respects to KO but works at a more detailed level. Talking of KR, Davis *et al.* Davis, Shrobe and Szolovits (1993), argue that "Any intelligent entity that wishes to reason about its world encounters an important, inescapable fact: reasoning is a process that goes on internally, while most things it wishes to reason about exist only externally". They then go on to discuss the components of KR, being a defined area of discourse, a purpose in creating a model of it, an SV of concepts and complex relationships and an inference engine operated by particular software. KR, then, has a distinct relationship with KO (or can even be seen as an applied form of KO) in that both use complex SVs. A major and important difference is that KR works most effectively in relatively closed systems with specific purposes, such as suggesting treatment on the basis of medical diagnoses. The challenge for the Semantic Web is to vastly extend such logical reasoning and, we should note here, to ensure that the underpinning and necessary SVs are not only available, but are as well-formed and maintainable as possible. A second difference that leads on from the first is that the vocabularies supporting KR are "special" as in the early use of the term to distinguish such schemes (usually specific to individual organizations) from the larger universal schemes such as the Dewey Decimal Classification or AGROVOC. A further distinction is that the special and universal schemes used in 'bibliographic' retrieval systems are in some sense predictive, especially where flexibility of co-ordination of concepts is supported by a structure created by facet analysis. Where the two converge is in the development of what we may call an ontology. In KR this is the vocabulary component that is manipulated by a logic layer such as OWL and mounted on the Internet using a language such as RDF. This ontology will usually have genus/species relationships (designated as Broader Terms (BTG)/Narrower Terms (NTG) in thesauri, and as 'isa' in KR); and whole/part relationships (designated as BTP/NTP in thesauri and 'is part of' in KR). A major difference is that the KR also has Related Terms which, unlike most thesauri, are specifically defined and bi-directional (for example Irbersartan *treats* Hypertension and Hypertension *is treated by* Irbersartan), where the traditional thesaurus uses an undefined RT. However, with currently available software it is now possible to create thesauri for information retrieval applications incorporating not only hierarchies and relationships with an admittedly limited range of definitions but to add notes and sortable codes to each term. These 'enriched' thesauri, sometimes called ontologies, can be extended to perform as terminology repositories to facilitate interoperability within, and even between, the various information systems of organizations. (Note, however, that closely defined RTs may be mostly specific to an organization or application, because whereas hierarchies must be either generic/specific, whole/part or instancial, thesaurus standards do not list universal RT types, other than as guidelines drawn from facet analysis).

Though the following point is not central to the argument in this paper, it could be argued that we, whether working in the areas of KO or IS, must no longer be concerned only with the models of SVs which, incidentally, have called on, for example General Systems Theory and the Theory of Integrative Levels in the approach to building bibliographic classifications. We must now also concern ourselves more thoroughly with those models, mentioned above; semantic maps as used in ontologies, as well as models that we might use in that part of IS called information architecture, such as enterprise and information architecture models and domain models used to describe the environments in which information retrieval operates. We should also be conversant with modelling techniques such as mind maps and the Unified Modelling Language, which has been used to model a complete set of all the possible components of a thesaurus and their relationships to each other.

In essence then, KO is about the analysis of explicit knowledge, and such analysis may not be confined to the support of conventional information retrieval which have grown and expanded in recent years as electronic information becomes more manipulable. These alternative applications of KO were well illustrated in a number of interesting papers given at the second Biennial Conference of ISKO-UK in July 2011. For example, Campbell (2012) showed how application of Farradane's relational indexing to parallel texts in the languages of health professionals and lay writers discovered significant, but not immediately recognizable, ambiguities between the two texts. In another paper, Lambe (2012) argued that in the current complex state of scientific research involving multidisciplinary teams, KO had a role to play, as an aid to sense-making in the pursuit of scientific discovery, by identifying and recording new boundaries and overlaps between apparently disparate subject areas. Finally, Petras (2012) discussed the deployment of Knowledge Organization Systems (KOS) in seven distinct applications of information retrieval systems.

Testing Models

Mention was made earlier of 'pragmatic epistemology', which must be an offshoot of a branch of philosophy called 'Pragmatism' which, proposed by the American philosopher Charles Sanders Peirce and elaborated by his compatriots William James and John Dewey, is defined as "A method of understanding facts and events in terms of cause and effect, and of inferring practical lessons or conclusions from this process" (Oxford English Dictionary, 2012). The testing of models created by KO for KR, KOS and information retrieval must, to a large extent be pragmatic: a process of inference rather than reliance on circumstantial detail. There are many more types of models than those discussed above, but all share the property of being representations of complexity so that we can better understand and deal with the world around us. But, because they are approximations and because the things that we model are constantly changing, we must also continuously test our models and revise them so as to maintain some

understanding of our world. Scientists are constantly testing models and challenging hypotheses (just think about the amazing discoveries and hypotheses emanating from the experiments in the Large Hadron Collider on an almost daily basis) and subsequently describing these models and hypotheses in the learned journals. "The use of machines or other models to simulate processes such as perception, recognition and their papers may then be analysed and indexed by information scientists, using semantic models devised by 'Knowledge Organizers', and here we arrive at an important point. The models that we call SVs must, like all models, be tested; and the only practical way in which they can be tested is in their application, that is to say through their efficacy in relating to the written records for which they are designed. It is not too fanciful to suggest that such testing is an aspect of 'applied epistemology', defined as: learning, and selective recall, or the application of principles assumed to hold for human categorization, perception, storage, search, and so on, to the design of machines, machine programs, scanning, storage, and retrieval systems" (McGraw Hill, 2003). So we may now say that KO has, like many other disciplines, a pure and an applied component; the former being concerned with the theory and techniques of KO, the latter with the application of the complete mentefacts such as SVs. This indicates, not surprisingly, that there is a large overlap between the interests and activities of the two communities of KO and IS, and one which, perhaps, is likely to become more pronounced. There has been mention in this paper, for convenience, of the words indexing and tagging without further elaboration; but in an environment of distributed processing and full-text retrieval these words are now used more loosely and tagging is not confined to specialists. At the same time, information scientists are increasingly directly involved with the creation of 'special' SVs and interoperability between them. In other words they have become more concerned with KO at the application level, while other information scientists continue to help maintain, develop and map the big schemes such as Dewey, MeSH and EUROVOC.

Conclusion

In recognizing that knowledge is an abstract entity confined to the individual human brain, the proponents of KO and IS must both work pragmatically, working as they do with concepts, as concepts are also abstract and words have no intrinsic meaning. Information is also abstract and information scientists must work with 'messages' which are the attempts of knowledge creators to disseminate their opinions. According to the SECI model knowledge creators record their ideas as explicit knowledge (Externalisation) which is then analysed by others (Internalisation) so that they can 'describe' them by processes such as abstracting and indexing (a sort of Combination), making these new products available for intelligent information retrieval. And so the cycle continues with users reading the retrieved documents (Internalization) and creating new messages, sharing these with others in discussion or giving papers at conferences (Socialization) and recording them (Externalization). The eminent information scientist Robert Fairthorne once said "It is not the job of information scientists to give information,

but information about information". One might add that it is not the job of those concerned with KO to give knowledge, but information about knowledge; and this is achieved through the construction of models. Throughout these various information chains models are being continuously created and pragmatically tested, from which it follows that the efforts of those working in KO, KR and IS overlap to such an important degree that they should work closely together. Unfortunately, present institutional and educational structures are often not providing a proper effective support for such collaboration.

References

- Campbell, D. Grant. (2012, 4-5 July). Revisiting Farradane's relational indexing in a consumer health context. In *Proceedings of the ISKO(UK) Biennial Conference*. Bingley: Emerald Group Publishing.
- Davis, R., Shrobe, H. and Szolovits P. (1993). What is a KR? *AI Magazine*, 14(1), 17-33.
- Floridi, L. (2010). *Information. a very short introduction*. Oxford: Oxford University Press.
- Heylighen, F.(1993). *Epistemology, introduction*. Retrieved on March 8, 2012, from <http://pespmc1.vub.ac.be/EPISTEMI.html>
- Lambe, P. (2012, 4-5 July). Knowledge organization systems as enablers to the conduct of science. In *Proceedings of the ISKO(UK) Biennial Conference*. Bingley: Emerald Group Publishing.
- McGraw Hill (2003). *Dictionary of scientific and technical terms*. New York: McGraw Hill.
- Miller, F. J. (2012). I=0 (Information has no intrinsic meaning). *Information Research*, 8(1),
- Myers, G..E. and Myers, M.T. (1998). *The dynamics of human communications*. New York: McGraw Hill.
- Nonaka, I. and Takeuchi, H. (1995). *The knowledge creating company*. Oxford: OUP
- Ohly, P. (2012). *Knowledge organization 2.0 – a communicative paradigm?* Retrieved on March 6, 2012, from <http://www.iskoi.org/doc/web/ohly.pdf>
- Knowledge. (2012). *Oxford English Dictionary*. Retrieved on March 6, 2012 from www.oed.com
- Petras, V. (2012, 4-5 July). Knowledge organization systems and their consequences for information retrieval. In *Proceedings of the ISKO(UK) Biennial Conference*. Bingley: Emerald Group Publishing.
- Fallibilism. (2012). Wikipedia. Retrieved on March 7, 2012, from <http://en.wikipedia.org/wiki/Fallibilism>