Spring 3-27-2012

Knowledge Diffusion via AUTrated Organizational CARTography [AUTOCART]

Mounir Kehal
ESC Rennes School of Business, mounir.kehal@esc-rennes.fr

Mark Gregory
ESC Rennes School of Business, markrogergregory@hotmail.com

Follow this and additional works at: http://aisel.aisnet.org/ukais2012

Recommended Citation
http://aisel.aisnet.org/ukais2012/23
Knowledge Diffusion via AUTomated Organizational CARTography [AUTOCART]
Mounir Kehal, Mark Gregory
ESC Rennes School of Business, Rennes, France
Email: mounir.kehal@esc-rennes.fr; mark.gregory@esc-rennes.fr

Abstract
The Post-Globalisation era has thrust businesses everywhere into new and different competitive situations where knowledgeable, effective and efficient behaviour has come to provide the competitive and comparative edge. Enterprises have turned to Knowledge Management to elaborate a systematic approach to develop and sustain the Intellectual Capital needed to succeed. To be able to do that, you have to be able to visualize your organization as based on knowledge and knowledge flows presented in a graphical and visual framework, referred to in this article as automated organizational cartography. This builds upon the ability to actively classify existing organizational content evolving from and within data feeds in an algorithmic manner. The article firstly reviews certain definitions and classifications of knowledge management, representing a wide range of views from mechanistic (systematic, data driven) to a more socially (psychologically, cognitive/metadata driven) orientated. The paper reports more elaborate continuum models for knowledge acquisition and reasoning purposes and suggests that they can be used for effectively representing the domain of information that an end user may contain in their decision making process for utilization of available organizational intellectual resources.

Keywords: Knowledge Maps, Organizational Cartography, Content Classification, Semantic Relevance, Semantic Fit, Unsupervised learning, Kohonen Networks, Self-Organizing Maps (SOMs)
1 Introduction

Business organisations invest massively in the application of information and communications technology (ICT). In particular, they aim to provide support for the work systems (Alter 2003; Alter 2006; Alter 2008) which embody their processes and activities by means of information systems in which people and organisations use ICT to manage their data as they exploit their knowledge to make well-informed decisions which they enact in their own actions and those of others. As Ray Paul suggests, information systems are IT in use (Paul 2010).

Some information systems exist specifically to encourage the use and reuse of organisational knowledge. Over the last decade and longer, a wide range of business improvement philosophies, approaches and methodologies have been developed. This development has been largely based on various combinations of business practices, management perspectives, and subject related research. Examples of these approaches are numerous and include organizational learning, the learning organization, total quality management (TQM), business process re-engineering (BPR) and earlier on quality circles (QCs). In more recent times Knowledge Management (KM) has started to emerge as an area of interest both in academia and in the corporate world. The literature reveals a rapidly increasing body of knowledge relating to KM. Knowledge management cross-links many different disciplines and areas of interest to academics and organizational practitioners. Thus knowledge is increasingly regarded as a fundamental factor of or input to economic activity, on a parity with land, labour, capital (OECD 1996).

In their ground-breaking book The Knowledge Creating Company (1995), (Nonaka & Takeuchi 1995) posited a model of how organisational knowledge, categorised as explicit or tacit, is created through four conversion processes, these processes being tacit to explicit (externalisation), explicit to tacit (internalisation), tacit to tacit (socialisation), and explicit to explicit (combination). Key to this model is the authors’ assertion that none are individually sufficient. All must be present to fuel one another. However, many important questions and issues arise in regards to KM. Might KM be a temporary eccentricity promising yet more false dawns in regard to organization development and management learning? Or can ICT-based information systems concretise and thus assist effective knowledge management?
2 Knowledge Management Literature Synopsis

While subject matter definitions can be helpful in clarifying the scope and depth of the subject under consideration, they can also be notoriously difficult to articulate. Definitions can often result in unwarranted simplistic reductionist formulations. When the subject that is being considered is in the domain of management, the difficulty is compounded even further due to the extremely diverse nature of the field. So what is knowledge management? And what is the knowledge which it is claimed should be managed?

Early authors in the field were often imprecise in their terminology. Hedlund used ‘knowledge’ and ‘information’ interchangeably and although he acknowledged that they should be distinguished, his use amounts to treating them as being identical (Hedlund 1994). (Myers 1996) referred to organizational knowledge as ‘processed information’. Nonaka and his colleagues describe knowledge as ‘a meaningful set of information that constitutes a justified true belief and/or an embodied technical skill’ (Nonaka et al. 1996). For Dejarnett, knowledge management starts with knowledge creation, which is followed by knowledge interpretation, knowledge dissemination and use, and knowledge retention and refinement (Dejarnett 1996). Thus one could consider Knowledge Management as a framework providing the ability to utilize the available knowledge resources effectively, and in a timely manner, for organizational benefit and advantage. As such it can be evident in organizational processes, the combination of data and information sources, the processing capacity of IT solutions, people, and the creation and innovative sharing of knowledge throughout the organization.

(Gregory & Descubes 2011) observe that the question of what knowledge is and how it can be represented or managed in an ICT-based artefact forming part of an information system is very far from trivial and indeed remains controversial. In their paper they broadly favour the view of (Kettinger & Li 2010), which is that information is the joint function of data and knowledge. Information, representing a status of conditional readiness for an action, is generated from the interaction between the states measured in data and their relationship with future states predicted in knowledge. Different forms of information system (IS) are conceptualized as the embodiments of knowledge domains capable of processing specific categories of data into information for business operations and decision-making. Kettinger and Li
conclude that the production of information from data needs knowledge, and when knowledge varies, so does information. They name their approach the KBI theory, the knowledge-based information theory.

However, (Gregory & Descubes 2011) suggest that even this approach is inadequate. The meaning of information has also to be considered. Thus viewing information not as processed data but rather as ‘data plus meaning’, (Mingers 1995) distinguishes four levels of information: symbolic empirics, syntactics, semantics and pragmatics. Meaning is generated from the information carried by signs. Information is objective, but inaccessible to humans, who exist exclusively in a world of meaning. Meaning is inter-subjective — that is, based on shared agreement and understanding — rather than purely subjective. Information and information processing systems exist within the wider context of meaning or sense-making (Weick et al. 2005).

Process is a crucial element inevitably omitted in any view of data, information and knowledge as static concepts.

In this paper we accept that certain kinds of knowledge can be represented on a computer and can be processed within a computerised information system. Thus the work of the programmer or developer is at least in part to represent certain business rules or process activities in the form of a computer program. To that extent, we can say that a computer program embodies or encapsulates a particular kind of explicit knowledge. This present paper is written within the conventions of automated processing of knowledge represented as text; thus it is not knowledge per se that is managed, but the traces of knowledge represented as natural language text and, as we shall see, subjected to a process which visually maps semantic inferences which illustrate and convey knowledge.

Andreas Abecker and Stefan Decker, in (Abecker & Decker 1999) and elsewhere, distinguish between two basic categories of IT contributions to knowledge management according to their main focus and approach. These are:

- The **process-centred view** which understands knowledge management as a social communication process which can be improved by means of ICT techniques such as groupware, computer-supported cooperative work CSCW and workflow management
• The product-centred view which focuses on knowledge represented as documents, that is, the results of an explicit knowledge explication process which seeks to formalise knowledge as a tangible resource.

In passing, Abecker and Decker note the serious deficiencies of the expert (rule-explication) systems of which such high hopes were held in the late eighties and early nineties. They suggest more modest goals focussed on supporting users as they perform arbitrary processes in which they exploit the knowledge already represented within an organisation. Part of their conclusion is that organisations should construct what they call Organisational Memory Information Systems OMIS, defined at (Abecker & Decker 1999):

“An Organizational Memory Information System (OMIS) integrates basic techniques into a computer system which within the enterprises’ business activities continuously gathers, actualizes, and structures knowledge and information and provides it in different operative tasks in a context-sensitive, purposeful and active manner in order to improve cooperative, knowledge-intensive work processes.”

They distinguish:

- Knowledge acquisition and maintenance
- Knowledge integration
- Knowledge retrieval

They looked forward to what were then to be next generation systems which would integrate product and process views via a knowledge flow mechanism on the basis of some knowledge cartography, an essentially visual categorisation.

2.1 Knowledge Category Models

Such types of model categorize knowledge into discrete elements. For instance, Nonaka’s model is an attempt at giving a high level conceptual representation of KM and essentially considers KM as knowledge creation process. Figure 1 shows Nonaka’s knowledge management model reflecting knowledge conversion and dissemination modes (Nonaka & Takeuchi 1995):

Figure 1 Nonaka and Takeuchi’s Knowledge Management model

<table>
<thead>
<tr>
<th>Tacit</th>
<th>To</th>
<th>Explicit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socialization</td>
<td>Externalisation</td>
<td></td>
</tr>
<tr>
<td>Internalisation</td>
<td>Combination</td>
<td></td>
</tr>
</tbody>
</table>
As can be observed from the figure above, knowledge is seen to be composed of two constituents, Tacit and Explicit. Tacit Knowledge is defined as non-verbalized, intuitive, and unarticulated. Explicit or articulated knowledge is specified as being formally structured in writing or some other explicitly defined form. However, is it appropriate to solely categorize knowledge in such a way? Another approach is the concept of P and Q knowledge recalled by (McLaughlin & Thorpe 1993) in their paper on Action Learning. This concept was originally suggested by Reg Revans in a book more recently republished as (Revans 1998). Revans suggested that

\[ L = P + Q \]

where L is learning, P is programming (or programmed knowledge with simulations) and Q is questioning to create insight into what people see, hear or feel. Tacit knowledge does not exactly map onto Q, neither does explicit knowledge exactly map onto P. Thus P and Q represent a different categorization, or taxonomical approach with regards to knowledge.

Hence, from an analytical and critical standpoint, Nonaka and Takeuchi’s categorization of knowledge can perhaps be seen as limited and uni-dimensional in its approach. However, their model assumes tacit knowledge can be transferred through a process of socialization into tacit knowledge and that tacit knowledge can become explicit knowledge through a process of externalisation. The model also assumes that explicit knowledge can be transferred into tacit knowledge through a process of internalisation, and that explicit knowledge can be transferred to explicit knowledge through a process of combination. Accordingly, the knowledge transforming processes are:

- Socialization, everyday comradeship
- Externalisation, formalizing a body and framework for knowledge
- Internalisation, translating theory into practice
- Combination, combining existing theories

Figure 2 shows Boisot’s model (Boisot 1987), which considers knowledge as either codified or uncodified, and as diffused or undiffused, within an organization. Boisot
uses the term ‘codified’ to refer to knowledge that can be readily prepared for transmission purposes (e.g. specialised data). The term ‘Uncodified’ refers to knowledge that cannot be easily prepared for transmission purposes (e.g. experience). The term ‘diffused’ refers to knowledge that is readily shared while ‘Undiffused’ refers to knowledge that is not readily shared.

Figure 2 Boisot’s Knowledge Category Matrix (Boisot 1987)

<table>
<thead>
<tr>
<th>Codified</th>
<th>Public Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proprietary Knowledge</td>
<td></td>
</tr>
<tr>
<td>Uncodified</td>
<td>Diffused</td>
</tr>
<tr>
<td>Personal Knowledge</td>
<td>Common Sense</td>
</tr>
</tbody>
</table>

Boisot tackles the issue of what knowledge is and how it can be represented in his article (Boisot & Canals 2004); see (Gregory & Descubes 2011) for a discussion.

In his doctoral research, the first-named author of this present paper aimed to approach this problem from a computational modelling perspective – see Figure 3:

Figure 3 Three-Tier Knowledge Lifecycle (Source: authors)

Following Figure 3, we believe that knowledge creation undergoes a nested set of computerized processes [explicit] and accompanying human practices [tacit], these allowing interlinkages and cross levelling to yield knowledge considered as the highest level of awareness available for an object of concern. The aim is rather to acquire automatically, represent visually, and reason collectively on textual content. Thus, a computationally mediated tool was conceived and built. Named AUTOCART, AUTomated Organizational CARTography, it supports knowledge evolution studies, knowledge sharing and corresponding flow representation.
## 3 Organizational Cartography and knowledge mapping

According to the Oxford English Dictionary, *Cartography* is the drawing of charts or maps. Our aim was to generate cartograms representing stored content attained from specialist data feeds. Hence, the generated map is one on which information is presented in a visual or illustrative form. The map reflects content by expressing it in terms of its textual constituents, sources of data feeds, relations and dependencies, and so forth as pertinent to its effective visualization. Figure 4 represents the characteristics by which ‘information in context’, a metaphor for knowledge, is dealt in the process of its acquisition.

![Figure 4 Knowledge Acquisition Spectrum (Source: authors)](image)

In Figure 4, Certainty, Internal, Interpretation and External are all knowledge states attained by means of extraction of contained-tacit and/or stored-explicit knowledge, with various possible values, states and roles, from data feeds and the outcomes of processing achieved by a mediated computation. Figure 5 below reflects the nature anticipated from such processing in a framework that models parameters by means of which knowledge may be viewed or represented. They are distinguished in the form of an [intangible object] action or thinking; and a [tangible object] archetype or human. An archetype is a universally understood symbol, term or pattern of behaviour; cf. (Wiig 1993).
This present paper reports on a tool which concentrates on knowledge cartography based on textual analysis. Such textual analysis is based essentially on mathematical analysis of histograms of the frequency of textual terms. The approach adopted by the first-named author of this paper in his doctoral research exploited Kohonen Nets. An alternative is latent semantic analysis, originally proposed and discussed by (Landauer et al. 1998) as implemented by (Rehder et al. 1998) and applied by (Wolfe et al. 1998). In fact, the implementation of Kohonen nets might itself depend on latent semantic indexing, as (Kohonen et al. 2000) acknowledges. (Kohonen et al. 2000) identify three possibilities to reduce the dimensionalities of the histogram vectors, without essentially losing accuracy in classification: 1) representation of the histogram vectors by their eigenvectors (which closely parallels latent semantic indexing as proposed by Landauer and his colleagues); 2) clustering of words into semantic categories and 3) reduction of the dimensionality of the histogram vectors by a random projection method, a computationally-lighter approach which Kohonen and his colleagues preferred for indexing massive document collections.

4 Systematic view of AUTOCART
The knowledge spectrum models covered above provided us with a framework for the development of AUTOCART, represented more abstractly in Figure 6 AUTOCART, Meta Level model, by use of dependency relationships and associations among processes and/or instances of objects. The Relationships and associations are stereotyped as <<refine>>, in accordance with the UML notation (Booch et al. 2005).
These dispositions of knowledge comprise parts of the Knowledge Constituents, which embody the ‘raw’ material of the organisation in question. Therefore, a UML generalisation relationship is used to depict the more specific kinds of knowledge elements in relation to the ‘whole’. Knowledge Constituents undergo some form of filtering, based on criteria derived from the document specification model and partly determined by the textual content. These functional processes are modelled in the next model, Figure 7, which focuses on functional requirements at a lower level of processing. Using this approach, AUTOCART assigns each knowledge element (texts in this case) to its textual category, primarily as determined in a pre-defined algorithmic manner by its textual contents. This is done using the principles of Kohonen Nets providing direction towards an automated inductive learning environment. This in turn makes possible alterations in terms of activation and threshold functions deterministic weights, which leads to toggling between
unsupervised and system-supervised learning for a networked representation of data. To establish the textual category is vital in classifying textual content and, along with characteristics such as links, directly added from the filtering process, Knowledge Nodes are formed. Consequently, the outcome of the AUTOCART Meta level model can be considered to be an entity embodying knowledge, possibly representative of peak levels of innovation feeding to actions, mediated by physical end users and presented in the form of archetypes. These specific types of knowledge representation are linked to the Knowledge Node using a generalisation relationship, following the notation of (Booch et al. 2005).

Figure 7 AUTOCART, process level
At process level, *Data Elements* within *Knowledge Constituents* are to be filtered and then accordingly classified. The overall aim is to surface the latent semantic structure of the *Knowledge Constituents*. The filtering process is primarily based on a document specification model – as incorporated in Figure 6 AUTOCART, Meta Level model – which is an aggregate of textual components. These can be identified as being the actual text of the document, annotation apparent in the document and the links present. The latter can be further specialised into *association Links* – pointing to and from related documents – and *classification Links*, including domain, project and user specific links, and other relationship links as a build-up of the data semantics is established based on semantics of content.

Once text components have been determined, each text is assigned a *Text Category*, driven by the coherent relationship between the document specification model and the textual content. In cases where the category is not readily determinable, a *Category Generator* is invoked, which assigns a category in an algorithmic manner.

Effectively, the process of textual categorisation and filtering results in assigning some kind of index to each textual input – in the form of data entries per document – in an attempt to reveal the latent semantic structure underlying the organisational knowledge elements. The associated data semantics form the core of the *Knowledge Nodes*, as the aim is to abstract the node’s content while retaining the original semantics.

As portrayed in Figure 7, the AUTOCART process level provides an architectural view of the processing needed for generation of Knowledge Nodes, mainly through links – obtained from the filtering process – and latent data semantics as determined by specification, categorization, and classification of the input data. Since the input data arises from designated data streams and from Web Technology (intranet) it can be described as *Network-enabled*. The resulting (intelligent) iMap is then a tool for distilling an enhanced representation of knowledge embedded within, from and by such data feeds.
Figure 8 demonstrates the modelling of knowledge nodes generation, after textual content have been categorized, following the practice of Kohonen Nets.

The text categories produced by AUTOCART, in a way illustrated in Figure 8, form the core of the knowledge nodes, accompanied by reference information such as extracted documented experience within the organisation, related communities of practice (Wenger 1998) and referenced expertise. This enhanced structure serves our purpose, which is not only the administration of electronically available information, but also a viable representation of the intellectual environment aiming to make information actionable and relevant within contexts of expertise coverage. Put simply, we aim to combine all valuable reference information in a framework to which everyone in a community of practice can relate to, effectively leveraging the organisational intellectual assets. These knowledge nodes would be of little value unless presented in an illustrative form. Therefore, it was chosen to generate cartograms to reflect knowledge instances comprising such nodes. Our approach has been heavily based on the concept of self-organising maps (SOM) introduced by (Kohonen 2001). Predefined text categories, either domain or project or user specific, play the role of input vectors while knowledge nodes correspond to neurons.
The main concept behind this analogy is to choose the winner topologically in the text categories space, according to its relevance for containment of the surrounding text categories. Figure 8 is representative of the intermediate step of this approach. In effect, spatial proximity – among knowledge nodes and predefined text categories – is taken to reflect semantic relevance and fit, a concept inherited from basic Information Retrieval models as in (Sparck Jones 1964). This is applied to each knowledge node provided as output of Figure 7 and used as input for the i-map depicted in Figure 8.

5 An application of AUTOCART: the case of SSTL

The doctoral research of the first-named author of this paper was reported in (Kehal 2006). The research centred on a case, that of SSTL, Surrey Satellite Technology Limited, a commercial spinoff of the University of Surrey. Almost all business organisations have to struggle to survive within a marketplace. Sometimes one of their main assets is the knowledge of certain highly motivated individuals who (appear to) share a common vision for the continuity of the organization. Satellite technology is a good example of that. From early pioneers to modern day mini/micro satellites and nanotechnologies, one can see a large element of risk at every stage in the development of a satellite technology, from inception to design phase, from design to delivery, from lessons learnt from failures to those learnt from successes, and from revisions to design and development of successful satellites. However, such knowledge creation and diffusion has sometimes been thought only to have manifested itself and been applied within large organizations and conglomerates. Organizations that manufacture, use, and maintain satellites depend on a continuous exchange of ideas, criticisms, and congratulations. One can view organisations ranging from the large (NASA) to the small (here SSTL Surrey Satellite Technology Limited) as forming a part of a class of knowledge-based organizations. Observational (questionnaire-based) and systematic (text corpus-based) studies – through case study elicitation experiments and analysis of specialist text - can support research in knowledge management. Through selective use of the previously stated approaches and also case study research methodology, we have been able to investigate how knowledge flows in a finite organisational setting and as modelled by analysis of specialist text. As shown in (Kehal 2006), we aimed to describe our understanding of
the nature of a specialist organization in a quantifiable manner. We have examined how knowledge flows and is adapted between commercial and research types of corpora. One of the major results deduced from the initial observational case study was that knowledge diffusion is paramount to the effectiveness of a knowledge-consuming research-focussed organization, supported by specialised and adapted information systems. This led us to investigate how knowledge diffusion takes place: in an empirical way. Our analysis shows that research papers (created within an educational institution) and commercial documents (created within a spin-off from such an higher education institution) can be distinguished rather on the basis of single word and compound terms. These two lexical signatures show the potential for identifying points of mutual interest in the diffusion of knowledge from the research institution to the commercialization process and thus to application(s) within a domain.

6 Conclusions and outlook

Technology enables data feeds linked to process content, which permeate utilization. Kohonen Nets, through unsupervised learning, may be able to behave correspondingly towards textual mapping and visual representation. We suggest that there is a need for combined unsupervised learning (which saves resources) and system supervised learning as threshold dynamism with parameters and a continuous cyclical behaviour to monitor newly generated textual categories.

7 References


OECD, 1996. The Knowledge-Based Economy.


