Special Languages and Shared Knowledge

Rafif Al-Sayed and Khurshid Ahmad
University of Surrey, Guildford, UK
r.sayed@eim.surrey.ac.uk
k.ahmad@eim.surrey.ac.uk

Abstract: The transfer of knowledge between groups of individuals of different levels of expertise and orientation is discussed with reference to the manner in which knowledge is disseminated using the specialist language of a given domain. A prototype system that allows access to knowledge at these different levels, through the automatic construction of keyword indexes, is outlined. The controversial relationship between knowledge and language is the basis of arguments in this paper.

Keywords: Knowledge management, knowledge sharing, knowledge diffusion, best practice, terminology management, health care.

1. Introduction

The transfer of knowledge within an organisation, across organisations, between an individual and an organisation, and between individuals is facilitated through a number of *sign* systems. Such systems include natural languages, mathematical equations, subject specific notations, and other conventions including graphical conventions. The term *facilitation* is a broad term, however, the key to facilitation is a common consensus on the meanings of words of natural language, kinds of mathematical equations, and agreement on notations and conventions. So, in some respects, the transfer of knowledge requires a consensus amongst organisations and individuals.

Much knowledge management literature has focused on the "sharing" of know-how and expertise through protocols devised by managers (Nonaka and Takeuchi 1995, Davenport and Probst 2002) or the focussed discussion of problems related to the sociology of organisations (Scarbrough 1996). Some have even looked at this problem from a cybernetic point of view in terms of *feedback* and *control* systems (Morgan 1996). Management Studies, sociology, and cybernetic models address fairly high-level conceptual issues. However, the surface form of knowledge, the trace of knowledge left behind on a document, whether paper or electronic, is amongst the few discernible forms of knowledge. We will focus on how this trace is transferred.

The long-standing controversy about the relationship between knowledge and language (see Baker and Hacker on Wittgenstein 1988) notwithstanding, it is almost universally true that the development of a subject or the development of a subdomain within a subject discipline invariably leads to the appropriation of certain words from the everyday natural languages of the emergent subject or subdomain workers. Words are given specialist interpretation; words like *energy, mass* and *force* existed in the English language prior to Isaac Newton. However after Newton propounded his theory relating to the material nature of being, these three words assumed a more specialist meaning and spawned a whole new discipline, i.e physics. Physicists, initially called *natural philosophers*, started discussing different kinds of forces, different sources of energy and problems relating to the metrication and instrumentation of quantities related to energy, mass and force. No journal of physics, standard textbooks or encyclopaedias of physics will accept an alternative term for these concepts. There is no obvious coercion but there is a consensus. The consensus is brought about partly through patronage, for instance having a degree in physics will allow one to write a doctoral dissertation or indeed obtain a job in various physics establishments but one has to speak and write in the specialist language of physics. Much the same is true of other disciplines.

We mentioned the development of subdomains within a specialism. Sometimes the subdomain relates specifically to the application of principles and empirical results related to the parent domain. In our times, *gene therapy* is a good example of such a transfer. Starting from the rather abstract concept of the molecular basis of animal or plant life, originally a theoretical and experimental enterprise variously called *biochemistry* and *molecular biology*, one sees the development of industrial methods and instrumentation for extracting and harvesting so-called genetic material – an enterprise now called *genetic engineering*. From genetic engineering the notion developed that some genetic material can malfunction giving rise to sickness of various organs within an organism; by replacing the defective genetic material, the organ will recover - hence *gene therapy*. Each of these different subjects i.e. nuclear biology and gene therapy has its own vocabulary and, indeed, writing styles for the discussion of theories and the reportage of experimental results.

Consensus relating to terminology, and elements of other sign systems, is used to show a commitment to certain concepts within a particular domain. This commitment is, in one sense, philosophical, for example Newton's notion of the material being of nature is a philosophical commitment to materialism articulated through words of the English language which were given specialist meaning. The commitment also relates to the basis of methods and techniques of the new science of the material being – physics – in that Newton chose differential calculus over algebra or geometry to describe the movement of material beings. A series of graphical conventions were adopted for displaying the results of experimental observations and tabulation protocols were set up to show the relationship between two or more variables. There is a third sense of this commitment which relates to the structure of knowledge – also referred to as epistemological commitment – in that Newton argued about the primacy of the three concepts, mass, force and energy, and emphasised that the other physical concepts could be derived from these three. The umbrella term for different kinds of commitment adopted by a domain community at a given time in their genesis relates to the existence of that community and of the ideas propounded by the community. This umbrella term is *ontology* – the study of the existence of being: the commitments could be called different kinds of ontological commitments.

In this paper, we discuss some of the challenges and opportunities related to sharing knowledge between experts and practitioners within a specialist domain and the sharing between the two groups and the potential end-users of the knowledge of the domain or those upon whom the knowledge will have an impact. The case in point here is that of breast cancer therapy. This is an extensively researched topic involving major laboratories and academic departments working on cancer treatment. The results of their deliberations are published in learned journals, written in a formal style for peer-to-peer communication – if you are not an expert or aspiring to be one in oncology or radiation therapy, for example, learned papers in these disciplines will mean very little to you. The knowledge of the experts is refined, related to the knowledge of other experts, and then passed on to the practitioners including cancer therapists working in hospitals, some having close links with the laboratories/departments, and nurses specialising in cancer therapy together with technicians involved in the operation of complex radiotherapy machines, various imaging devices, and/or highly toxic drug treatments. This refined and correlated knowledge is documented in a peer-to-operative language and practitioners themselves write some of the documents. Another important development in recent times has been that of digital libraries and documentation archives that can be accessed through the Internet. Nowadays, the Internet is the first place people go to seek clarification and knowledge related to complex topics; sometimes cancer patients, especially those who have just been diagnosed or about to receive (novel) therapy, tend to consult the Internet. Major cancer charity organisations have devised documents in a language which is more accessible to this new audience. These documents are written in an operative/expert-to-lay person language.

We report on the development of an information spider: a computer program that can allow access to a range of documents, for example learned papers, practice manuals, and fact sheets. The spider not only allows access but helps in creating a text archive and in extracting terms from documents for indexing purposes as well.

2. Shared concepts, terminology and knowledge spirals

Early literature on knowledge management focused on sharing knowledge related to industrial innovation: there are two well-cited examples of this genre of sharing. The first relates to the development of new product lines by persuading researchers, product designers, manufacturing and sales personnel to work together across departmental and status boundaries (Nonaka and Takeuchi 1995:95-123). The second example relates to the sharing of 'local innovation' in the design of usable technology by sharing the knowledge of the end-users of the products (Seely-Brown 1998). Both of these classic examples describe how large organisations used brainstorming methods, and software systems for co-designing and for cross levelling the knowledge within the organisations.

Knowledge sharing in more recent literature stresses more indirect interaction between the constituent members of a (geographically distributed) organisation. For instance, organisations keen on their staff sharing 'best practices' typically use a document repository – for example reports of past successful/failed projects, employee, product, and service profiles (e.g. the so-called Yellow Pages) – and tools for inputting and extracting knowledge from such repositories (Davenport and Probst 2002). The range of knowledge sharing systems includes document management systems, systems that manage documents which have been selected and annotated by experts for the use of others (Gibbert, Jonczyk and Völpel 2000), to the ambitiously-titled *intelligent systems* (Fisher and Ostwald 2001).

Knowledge sharing within a community is a more recent phenomenon and appears to be supported by public-sector organisations. For example, the US National Cancer Institute, a US government agency, is 'cross levelling' knowledge across the sub-communities of cancer researchers, cancer-care professionals, and the public at large (Cancer 2003). Again, a document repository is at the heart of the National Cancer Institute's system. The repository comprises newsletters, fact-files, journal papers, application notes for care workers, information specific to cancer for the public at large, and a glossary of terms.

2.1 Intra-organisational knowledge sharing and exchange

Classical knowledge sharing models suggest that the knowledge transfer/sharing process involves the conversion of *tacit* knowledge into *explicit* knowledge and vice versa. En route there are processes that help share explicit and implicit knowledge without conversion. These models focus largely on how knowledge is shared within an organisation or intraorganisationally. The sharing of knowledge within an organisation at one level should be part of the natural functioning of the organisation. At another level there are a number of bottlenecks prohibiting this transfer including physical problems of disseminating information, social problems related to prestige and power, and linguistic problems of sharing knowledge across different levels and kinds of expertise. As we show later, interorganisational transfer of knowledge can pose equally severe challenges.

The terms *implicit* and *explicit* knowledge are ambiguous and subject to much philosophical debate. For Nonaka and Takeuchi (1995) the conversion of knowledge from implicit to explicit and finally to implicit is the basis of knowledge creation. Choi and Lee (2002) have observed a close relationship between the management strategies of Korean enterprises and the knowledge conversion modes suggested in Nonaka and Takeuchi.

Generally, explicit knowledge is formalised consensually, and is articulated in the language of a specialist domain through texts. These texts are either informative (learned texts) or instructive (instruction manuals). Implicit knowledge is articulated mainly through the spoken word and is suffused with metaphors, similes, and analogies. Implicit knowledge is largely informal and idiosyncratic of individuals. Documents like inter-office memos, product catalogues, advertisements for goods and services, comprise both implicit and explicit knowledge.

The knowledge conversion process involves a close interaction between, and understanding amongst, the key players - the *knowledge crew* of an organisation: these include the experts, professional workers, including production/marketing/sales staff, researchers and design engineers, the end-users of the artefacts created by the experts and professional workers. The artefacts may include goods and services.

There are four modes of knowledge conversion, according to Nonaka and Takeuchi (1995:71-73), and we discuss these modes with reference to the exchange of terminology and concepts amongst the crew during each of the modes:

(i) In the SOCIALISATION mode the crew works on an informal basis: verbal exchanges enable the crew to understand each other's vocabulary.

(ii) *SOCIALISATION* is followed by *EXTERNALISATION*. Here, an inventory of novel, revised, and abolished concepts is produced in a written document;

(iii) *SOCIALISATION* and *EXTERNALISATION* produce fragmented knowledge. The knowledge crew then tends to fuse concepts and terminology in the so-called *COMBINATION* mode. The fusion is implicit in the development of new methods of working or new products.

(iv) Once the method and products are established, the crew internalises the operational details, sometimes improving on it and at other times jettisoning some of the new knowledge. This is the INTERNALISATION *mode of knowledge transfer. This ultimately leads to*

SOCIALISATION, EXTERNALISATION *and* COMBINATION.

The articulated public and consensual development of a shared conceptual system and its vocabulary is more vivid in a loosely-organised setting, e.g. systems for sharing best practice, than in the high-pressured setting as encountered in the creation of a new type of automobile, home bakery (Nonaka and Takeuchi 1995), or smarter and non-intrusive photocopiers (Seely-Brown 1998) where an organisation explicitly plans for a targeted change.

Best practice is shared across an organisation and the recipients of collated/created knowledge are not as well defined as may be the case for design and production engineers sharing the ideas of an architect (product/services) and a marketing expert. Recent developments in knowledge creation are broad-spectrum. This we discuss next.

2.2 Inter-organisational knowledge sharing and exchange

Mergers and acquisitions (M&A) between organisations present a major challenge to knowledge management in that M&A precipitate lasting changes in the participating organisations, and the acquiring organisation undergoes changes when it takes over the other organisation. The example of Siemens' Information and Communication Mobile (ICM) segment is quite apt here (Kalpers *et al* 2002).

There are a number of tasks that involve the workers in the two (or more) organisations during a merger and acquisition: Kalpers *et al* describe the workers as a *Business Community*: 'a [geographically and organizationally distributed] group of people who share existing knowledge, create new knowledge, and help one another on the basis of a common interest in a business-related topic' (2002:197). The *Business Community* 'was designed as socio-technical system' for facilitating the 'combination of knowledge and the creation of new knowledge' (ibid:198). The five main activities of the *Business Community* suggest that the exchange of knowledge is primarily through social interaction and quadri-modal as per Nonaka and Takeuchi (Table 1).

Key Activities of the Business Community	SOC	EXT	COMB	INT
Sharing regular events: face-to-face and phone conference	✔			
Urgent request forum: Discussion forum with email and Net-meeting sessions	✔	✔		
Information-platform process for knowledge packages and project information		✔	✔	
Merger and Acquisition (M&A) process improvement work-shops			✔	✔
Disseminating information related to M&A projects through information brokering and debriefing	✔			✔

Table 1: Activities of the Business Community and knowledge conversion modes.

The technical component of the *Business Community* is an information system that helps in the storage, annotation and retrieval of documents. Kalpers and colleagues talk about *K(knowledge) Packs*: clearly formatted structures for encapsulating meta-level and summarised contents of documents. The documents can be classified in different facets: (i) according to the type of change – *merger, acquisition, divestment*; (ii) according to the relevant business process – *human resources, logistics, product design*; (iii) according to M&A processes and phases - *monitoring, evaluation, integration/post closing*; (iv) according to IT topics - *data, applications, infrastructure, security*; and (v) according to the organisational structure of Siemens – *group-wide, business-unit wide, region-wide*. K-Packs range from informative (contacts, project documentation, laws, contracts) to instructive documents (checklists, documents templates, lessons learnt/annotated histories).

This multi-faceted information platform is called an *information spider* or an *infospider.* There is a team of authors and editors involved in providing potentially 'reusable knowledge' to this document repository. According to Kalpers *et al* 'a sophisticated search engine allows the user to keyword-search (*sic*) the K-Packs …[and there are facilities] to browse the most popular and often used K-Packs' (2002:201). The initial evaluation of the Siemens' M&A *Knowledge Exchange* (MAKE) appears to be encouraging. What interests us is how the M&A experts built up the knowledge of the mergers and acquisitions business.

3. Special language and knowledge sharing

The different modes of knowledge conversion help in the articulation, explanation, revision, and acceptance/rejection of key concepts within a group with diverse interests: the players in the group ensure that the terminology they use in articulation and explanation of concepts is clearly understood by others. The group interaction helps the group in achieving a shared understanding of concepts by sharing the terminology of each other. There is anecdotal/case study evidence in Nonaka and Takeuchi suggesting that 'speaking a common language and having discussions can assemble the power of the group. This is a vital point, even though it takes time to develop a common language' (1995:99). The development of the understanding of the vocabulary of a specialism is discussed under the rubric of languages for special purpose (LSP) (Sager, Dungworth and MacDonald 1980; Schröder 1991): this subject has an active constituency in Northern Europe and North America as evidenced by academic journals (e.g. Fachsprache). The use of LSP in shaping specialist written knowledge is a subject of debate in pure and applied linguistics (Halliday and Martin 1993; Bazerman 1988). One major area of research in LSP is the growing gulf between language used by experts and by the layperson

3.1 Knowledge exchange and LSP terminology

Any specialist language is a part of the natural language of the authors of specialist texts: 'Scientific English may be distinctive, but it is still a kind of English, likewise scientific Chinese is a kind of Chinese' (Halliday and Martin 1993:4). Pejorative remarks that equate specialist talk with obfuscating jargon notwithstanding, specialist languages are an excellent example of parsimony that hallmarks human cognition: a small set of keywords is used to represent a large body of knowledge, or, more specifically, these keywords usually comprise a significant proportion of specialist texts. This parsimony is essential for reducing ambiguity and increasing precision. An even smaller set of single words is used by the community as their (specialist) signature: physicists will write around and about mass, energy, force, time and space, biologists around and about life forms, evolution, heredity, and environment for instance.

The role of shared terminology in knowledge creation is perceptible in the MAKE system. Each K-Pack has associated keywords and MAKE has access to a search engine that presumably makes use of the keywords. Human editors append the keywords to the documents. The editors make a judgement about the suitability of the keywords for a given document and assume that a potential user will be familiar with the keywords. This is a time-consuming and expensive process.

In the following, we outline a method for automatically extracting candidate single word terms and compound terms, for automatically identifying relationships between terms based solely on the behaviour of the candidates in relation to other terms and words used in everyday discourse, the so-called general language discourse. Our method is domain-independent and relies only on a representative but random sample of texts used in a given specialism – cancer care for example – together with a sample of texts used in general language.

3.2 A text-based method for identifying shared knowledge

The introduction, usage, and obsolescence of words in a language is complex and creative. Language experts, particularly lexicographers, have advanced a plausible explanation in relation to the birth, currency, and death of words: they argue that the *frequency* of a word generally correlates with its acceptability by the language community (Quirk *et al* 1985). The frequency is computed by examining a collection of written texts (or speech fragments) randomly sampled from a *universe* of texts. Such sampling is essential especially since the language system is open-ended.

Corpus linguistics is a branch of linguistics where the emphasis is on the use of systematically organised text collections – text corpora or text corpus (singular) – as a starting point of linguistic description or as a means of verifying hypotheses about a language. Machine-readable versions of such collections have been developed for major languages of the world. One major beneficiary of corpus linguistics is lexicography – and many individual dictionary publishers have their own in-house corpora.

The British National Corpus (BNC) of 20th century English language comprises over 100 million words including written text (c. 90%) and speech fragments (10%) (Aston& Barnard 1998). The written component comprises 3,209 texts published mainly between 1975-1993: two-thirds of the texts belong to imaginative genres (novels, literary magazines), the arts, world affairs and leisure, and the other third to natural, pure, applied and social sciences. There are approximately 250,000 unique words including plurals of nouns and verbs in different tenses. Some of the words are used in most texts and most frequently - 6% of the BNC is the word *the* (6 million instances) - and yet others are used rarely; the word *cancer* is used 949 times in the BNC, *neutron* appears 247 times and *radionuclide* 40 times. Words like 'the' and other determiners (*a, an*), conjunctions (*and, but*), and prepositions (*in, on*) are the most frequent and comprise a quarter of the BNC. These are called *closed-class* words as English-language users seldom invent new determiners or prepositions.

Words belonging to the open-class category*, nouns, adjectives, adverbs*, are not as frequent. Indeed, amongst the 100 most frequent words in the BNC comprising about half the words in the corpus there are only two nouns, *time* and *people*.

3.2.1 Language-related and subject-related signatures

Recall that a specialist writing about his or her domain of specialist knowledge writes in a form of natural language. A specialist document typically has two signatures. The first signature signifies the natural language of the document and the second signifies the special domain.

A corpus-based analysis of a number of individual subject domains, ranging from subjects as diverse as nuclear physics to dance studies, philosophy of science to sewer engineering, theoretical linguistics to cancer research, suggests the existence of the two signatures (Ahmad 2001 and references therein). A corpus was created for each domain usually by keying in a subject name on a search engine and selecting texts of different genres: journal papers, text books, advertisements for goods and services, conference announcements specifically dealing with topics in the domain. The corpora varied from 150,000 words to 750,000 words.

The language-related signature of an English LSP shows itself in the distribution of closed-class words. This distribution is the same as that of the British National Corpus: the first 10 most frequent words in almost each of the domains included determiners, prepositions, and conjunctions. The subject related signature of an LSP is reflected in the profusion of open-class words, mainly nouns, in the 100 most frequent words: in some disciplines as many as 30 nouns comprise the 100 most frequent words and in others about 10 or so.

The most frequent nouns refer to a small group of concepts in the domain: in nuclear physics the 100 most frequent words include the names of key objects of study in nuclear physics - the *atomic nucleus*, constituent particles of the nucleus, protons and neutrons - and key concepts in physics - *energy*, *force* and *mass*. In linguistics, the 100 most frequent words include the names of the grammatical categories or words, *noun*, *verb*, *adjective*, together with important theoretical notions of *transformation, structure* and *grammar*.

The subject-related signature discussed above refers to single words. Specialist language differs more sharply from general language in the usage of compound words, containing as many as six single words. It turns out that the most frequent single words, *nucleus* and *nuclear,* are the key ingredients of many of the most frequent compound terms in nuclear physics, i.e., *nuclear structure* and *nuclear reaction*, *target nucleus*, stable/unstable *nucleus.*

3.2.2 Automatic identification of terms

It is the profusion of subject-related nouns that distinguishes a special language text from a text written in general language. For example, for one instance of the term *nucleus* in the BNC there may be as many as 300 instances in a typical nuclear physics corpus – the ratio rising to over 5000 for the plural *nuclei*.

The ratio of the relative frequency of a word in a specialist corpus and in a general language corpus may suggest whether or not the word is a term. As closed-class words have a similar distribution in the two corpora, the ratio of relative frequencies of these words in the two corpora, one specialist and the other general language, is generally around unity. But the ratio of the relative frequency of subject-related nouns within a specialist text (corpus) to that in the BNC is generally greater than 1 and indicates a candidate term. This ratio is sometimes called the *weirdness ratio*. The computation of weirdness is the first step in automatic extraction.

3.2.3 Subject-related signatures and knowledge sharing

One example of knowledge sharing is the emergence of an applied science or engineering science around a theoretical subject. The example of nuclear physics (NP) will illustrate this point. The systematic use of nuclear radiation in medicine and agriculture is discussed in the radiation physics (RP) literature. RP is based on key concepts in nuclear physics: concepts that help explain naturally radioactive elements, or unstable elements that emit nuclear radiation, or concepts that describe how stable elements can be made unstable, or radioactive, by *bombarding* or *irradiating* these elements with other radiation. The controlled use of emitted radiation is used in radiation therapy or diagnosis. Nuclear (reactor) engineering is a branch of engineering based on the theoretical concepts of nuclear fission in nuclear physics.

The applied sciences and engineering are regulated by law to ensure the safety and well being of humans whilst promoting the use of potentially lethal artefacts like nuclear radiation. Radiation protection/safety has emerged as a discipline following the extensive use of radiation physics.

In order to be autonomous disciplines, both radiation physics and radiation protection have to have their own concepts and associated terminology, a terminology that manifests itself as subject-related signatures. A three-way comparison between the three subjects will show the influences of the parent and the progeny's own identity. We have created three corpora to study these influences and identity: theoretical nuclear physics (151 texts comprising 444,540 words, published between 1970-1999), radiation physics (91 texts, comprising 286,676 words, published between 2001-2003), and radiation safety (16 texts, comprising 127704 words, published in 2003). The texts are written in American and British English and are drawn from journals, textbooks, public announcements and advertisements.

Table 2 shows the ten most frequent single words in each of the corpora: nuclear physics and radiation physics 'share' two key terms: *energy* and *neutron*; radiation physics and radiation safety 'share' the terms *dose* and *radiation*. The other eight terms show the autonomy of the disciplines.

Table 2: Subject-related signatures in three disciplines in physics

Nuclear Physics		**Radiation Physics**		**Radiation Safety**	
N= 444540		N= 286676		N= 127704	
Term	*f/N*	Term	*f/N*	Term	*f/N*
energy	0.57%	dose	0.79%	mutation	0.91%
nucleus	0.52%	neutron	0.41%	dose	0.75%
neutron	0.41%	beam	0.40%	disease	0.60%
nucleon	0.35%	radiation	0.33%	gene	0.59%
nuclear	0.32%	energy	0.30%	radiation	0.57%
potential	0.32%	system	0.27%	risk	0.47%
target	0.25%	treatment	0.24%	rate	0.45%
scattering	0.24%	image	0.22%	exposure	0.32%
interaction	0.21%	rays	0.22%	cancer	0.31%
mass	0.20%	detector	0.19%	radionuclide	0.30%
TOTAL	3.390%		3.356%		5.254%

Let us now compare the distribution of five of the most frequent terms in each of our corpora and in the BNC (see Table 3). What one sees in the distributions is that the term energy is used 43 and 23 times more frequently in the NP and RP corpora respectively than in the BNC; more demonstrably, the term dose is used 337 and 291 times more in the RP and RS corpora respectively than in the BNC, and the term neutron is used 790, 1379 and 54 times more in NP, RP and RS corpora respectively than in the BNC. The term nucleon, the weirdest in the three corpora, is used only in our nuclear physics corpus.

Table 3: Weirdness ratio for the most frequent open-class words in the three corpora

Nuclear Physics		**Radiation Physics**		**Radiation Safety**	
N=	444540	N=	286676	N=	127704
Term	$f_{NucPhys}/f_{BNC}$	Term	$f_{RadPhys}/f_{BNC}$	Term	$f_{RadSafets}/f_{BNC}$
energy	43	dose	337	mutation	629
nucleus	535	neutron	790	dose	291
neutron	790	beam	218	disease	50
nucleon	6402	radiation	125	gene	309
nuclear	39	energy	23	radiation	409

The 10 subject-related signature terms help (in Table 2) in the formation of compound terms and illustrate the linguistic parsimony and linguistic productivity of specialist writers. The term *nucleus* is used as a head word for two frequent compound terms, *target nucleus* and *halo nucleus*, and the neologism *nucleon* acts as a modifier for the most frequent compound in our nuclear physics corpus, *nucleon-nucleon amplitude*. In radiation physics *neutron* is used as a head word for the frequently occurring *thermal neutron*, or as a modifier in *neutron-capture therapy* and the other noun in the noun-noun compound *neutron fluence*. *Radiation* acts as a dominant constituent in the radiation safety corpus, as a modifier in *radiation exposure* and *radiation dose*, in its derivative form *radiological protection*, and as a head word in *ionizing radiation*.

Table 4: Most frequent compound terms in the three corpora. Terms in italics are neologisms

Nuclear Physics	Radiation Physics	Radiation Safety
nucleon-nucleon amplitude	dose distribution	radiation exposure
neutron star	thermal neutron	congenital abnormalities
nuclear physics	neutron capture therapy	Multi-factorial disease
angular distribution	radiation therapy	ionising radiation
target nucleus	*neutron fluence*	air concentration
halo nucleus	spatial resolution	genetic disease
nuclear reaction	fluorescence reabsorption	transfer coefficient
nuclear structure	maximum dose	radiological protection
angular momentum	intensity matrix	breast cancer
radioactive beam	radiation physics	radiation dose

The theoretical notion of a structured and composite nucleus, and interaction between the constituents of two nucleons (as in n-n amplitude), shows the physico-philosophical bias of the subject and that of the terms. In radiation physics, the term dose (or the energy of the radiation), and its control, dominate the discussion and show the applied physics/engineering bias of the subject. Radiation safety deals with exposure to the risk of nuclear radiation – hence the most frequent terms radiation exposure, radiation dose and the current interest in breast cancer dominate the discussion in the RS corpus demonstrating the ethico-legal aspect aspects of the subject.

We have attempted to describe how knowledge sharing can be monitored using a text and terminology management system by identifying the subject-related signature of specialist subjects, and particularly how the sharing of terminology across disciplines indicates the sharing of concepts. The explication of knowledge in nuclear physics resulted in the development of radiation physics, and explication of radiation physics knowledge led to the domain of radiation safety. Each of the two explications have led to the internalisation of knowledge which when explicated has its own terminology.

The results in nuclear physics and related disciplines have been replicated in the transfer of knowledge in theoretical solid state physics to electron device engineering (Al-Thubaity and Ahmad 2003); in knowledge transfer from civil engineering to environmental planning systems (Ahmad and Miles 2001); and in a study of how concepts in cognitive psychology and structuralism found their way in theoretical linguistics (Ahmad 2002).

In the next section we discuss how the automatic extraction of terminology for identifying the subject-related signature of a domain, and for identifying its impact on its application/applied domain, can be used to build an information spider semi-automatically. Such a method will facilitate the automatic annotation of key terms for each of the documents and the stronger and weaker cross-referencing between the parent and progeny domains.

Our chosen domain is cancer care where experts are attempting to share their knowledge with professional workers, including therapists, nurses, and radiation workers, and where both experts and professionals are attempting to do the same with increasingly Internet-aware actual or potential cancer patients. Ours is a corpus-based study.

4. Monitoring and documenting change and differences: A health infospider

Health-care is an all-pervasive domain where advances in medicine and the concomitant costs respectively encourage and discourage the use of new knowledge. In this domain documentation is the 'main means of communication between care providers' (Ruch et al 1999) and the effective healthcare delivery systems have become increasingly dependent on accurate and detailed clinical information based on best practices (Chute, Cohn and Campbell 1998).

Knowledge of advances and best practice can be shared and refined by formal knowledge dissemination outlets, for example journal papers, workshops and seminars, and through learning-by-doing during encounters with patients. The Internet facilitates sharing of scientific results either through digital journals or through research notes posted on secure websites relating to drug trials, for example. The widespread use of the Internet has led to potential and actual patients, or their friends and relatives, going online for information after receiving news that the patient is or might be suffering from cancer.

Health-care knowledge has to be shared between many organisations and increasingly that knowledge has to be shared with an open-ended audience. In health-care or its sub-domain cancer care, as in any other specialist domain, terminology management is of the essence: including new terms and expunging old ones. Maintainers of controlled medical vocabularies recognize that such vocabularies are not static (Cimino 1996).

The US National Cancer Institute (NCI) is attempting to provide up-to-date online information on cancer to two groups: health-care professionals and patients. The NCI website provides a facility for searching the contents of its document base; there is also a glossary of cancer terms. The website is organised and is accessible according to different facets: users can look at individual types of cancer, at different types of treatments, and at the results of studies being carried out. Information for professionals is generally in the form of an extended abstract or summary about a specific topic together with an extensive bibliography. References to published journal articles in the bibliography of a given extended abstract are generally hyperlinked to the abstract of the cited article. Information for patients is provided without extensive references to journal articles and is mainly in the form of fact sheets: highlights of a recent diagnostic or therapeutic discovery, of a long-term study and other useful information. In addition to the US NCI, and other national cancer charities like Cancer Research UK, pharmaceutical companies also provide information about their drugs as fact sheets.

4.1 Building a cancer infospider

In order to ascertain the subject-related signature of the language used by experts for cancer-care professionals and for addressing laypersons, especially patients, we have created three text corpora. We are not considering the parent discipline - cancer research - rather focusing on its three progenies to determine the extent to which knowledge is shared between the three progenies by measuring terminological commonalities. In order to illustrate our ideas we have focused on aspects of diagnosis (specifically the breast cancer gene), therapy and after-care of breast cancer patients.

The breast-cancer expert corpus comprised 300 texts, abstracts, and full papers (114,394 words). The texts were collected by navigating medical journals and websites (such as the breast-cancer research and nature.org web sites) using the keyword *breast cancer gene* (abbreviated as *brca1* and *brca2*). The breast cancer care professional corpus, comprising 1,000 texts (226,464 words) was built by collecting texts from the US National Cancer Institute, US National Library of Medicine, and the Journal of American Medical Association. The keyword used to collect the texts was *breast cancer*. The cancer-patient corpus, comprising 800 texts (464,000 words) was collected by mainly focusing on texts made available by cancer charities – the American Cancer Society, Cancer Research UK, Alliance of Breast Cancer Organisations, and the California-based Bay Area Tumor Institute. (Recall that US NCI website has two sub-sites - one for professionals and the other for patients.)

The subject-related signature of each of the corpora was compared to the British National Corpus. The terms *breast* and *cancer* dominate the three corpora and comprise 3.26 % of the expert corpus 3.3% of the professional corpus and 5% of the patient corpus. The word *women* dominates the three corpora and was among the most frequent words, but the term *patient* acted as a dominant constituent in the professional and patient corpora. The key differences in the corpora perhaps indicate the extent to which the experts think they are ready to share their current knowledge with professionals and patients. One can detect some differences in the most frequently used words in the these corpora – the experts have found new breast cancer genes, so new that they have not been given names, rather they are referred to as *brca1* and *brca2* and *mutations;* the rather high frequency in the professional corpus of these acronyms, as compared to the patient corpus, suggests that experts are almost ready to share this knowledge with the professionals.

Of the established knowledge, the terms *(breast) surgery, mastectomy* that are preceded (or followed) by *biopsy* and *radiation*, occur more frequently in the patient corpus than in the professional, while *biopsy* is an not frequently used in the expert corpus. Comparison with the BNC is also instructive:

the comparison of the use of the 14 most highest frequent terms in each of the three corpora with the frequency of the terms in the BNC show how weird these terms are: even the familiar word *family* is used 63 times (expert corpus), 4 times more frequently than the BNC. There are certain terms that are used 5000 times more in our corpora than in the BNC - *tamoxifen* and *ovarian* in the expert corpus, *tamoxifen* in the professional corpus and *mastectomy* in the patient corpus. (See Table. 5)

Table 5: The contrastive distribution of scientific terms in the expert, professional and patient corpora compared to the BNC. Terms in bold provide a subject- related signature.

Expert	f_{Exp}/N_E	f_{Exp}/f_{BNC}	Professional	f_{Prof}/N_P	f_{Prof}/f_{BNC}	Patient	f_{Pat}/N_{Pat}	f_{Pat}/f_{BNC}
N=114,394			N=226,464			N=464,000		
cancer	1.87%	443	cancer	1.41%	320	breast	2.19%	745
breast	1.39%	831	breast	1.25%	430	cancer	2.18%	465
brca1	1.37%	INF	women	0.64%	11	women	0.96%	15
brca2	0.71%	INF	risk	0.56%	43	treatment	0.61%	47
mutation	0.49%	1014	patient	0.53%	24	risk	0.47%	33
families	0.53%	63	treatment	0.27%	22	therapy	0.32%	153
risk	0.50%	41	therapy	0.23%	116	surgery	0.28%	100
ovarian	0.39%	7893	tamoxifen	0.21%	7149	chemotherapy	0.26%	969
gene	0.33%	148	chemotherapy	0.20%	757	cells	0.30%	23
carriers	0.33%	512	estrogen	0.20%	INF	lymph	0.29%	1316
women	0.23%	7	disease	0.20%	19	radiation	0.20%	108
dna	0.23%	68	brca1 & brca2	0.20%	INF	biopsy	0.18%	177
protein	0.22%	76	ovarian	0.19%	3687	mastectomy	0.16%	5360
tamoxifen	021%	7242	family	0.13%	4	tamoxifen	0.15%	5265

The notion of weirdness helps us to establish whether or not a word has been appropriated by the specialists in their general languages and turned into a term that, in turn, becomes part of the specialists' special language. Recall that weirdness is the ratio of the relative frequency of the term in a specialist corpus of texts and the relative frequency of the (source) word in the general language. Higher weirdness means that the word has been appropriated, and the key indicator of the appropriation is the (much) higher frequency of use in the specialist corpora than in the general language corpus.

Let us see whether we can extend the metaphor of weirdness when we compare the language of the experts with that of the professionals or when we compare the language of the professionals, or the experts, with that of the patients. If a term is much more widely used in the expert corpus than in the professional corpus then one might infer that the concepts/artefact denoted by the term are in a state of evolution and hence not used as extensively by the professionals as by the experts. Similarly, a weird use of a term in a professional corpus, when compared with the patient corpus, may suggest that the concept/artefact related to the term is either not important to the patient or the concept/artefact is still being matured by the professional community. Contrastingly, if a term has a weirdness of ONE when we compare its relative frequency in the expert corpus with that of either professional or patient corpus, then we might infer that the concept/artefact denoted by the term is quite well established amongst the professional and the patients.

A comparison of the distribution of 26 terms shows that terms like *brca1, brca2*, *mutation, carrier, chromosome, gene* are used over five times more in the expert corpus than in the professional corpus. The experts are less interested in *chemotherapy*, *carcinoma*, and *surgery*, as they use these terms 5, 14 and 16 times less than the equivalent use of the terms by the professionals. One way to illustrate the preference experts have for a term when compared to the professionals, and vice versa, is tabulate the logarithm of weirdness of the most weird terms for a professional when he or she reads an expert's texts: positive values of the logarithm of the ratio of the relative frequency of the same term in an expert's texts when compared to professional show preference use by experts. A negative value of the ratio shows the less frequent use of the term by the expert when compared to a professional.

Table 6a: The contrastive distribution of relative frequency of the terms in the experts and the professional corpus.

Words	$Log(r_{Expert}/r_{Professional})$	Words	$Log(r_{Expertl}/r_{Professionl})$
brca1	1.007	receptor	-0.08
tamoxifen	0.004	adjuvant	-0.24
chromosome	0.87	therapy	-0.63
brca2	0.85	chemotherapy	-0.69
carriers	0.84	diseases	-0.72
dna	0.82	clinical	-0.76
mutation	0.78	hormone	-1.09
gene	0.78	tumors	-1.09
protein	0.75	progestin	-1.15
germline	0.58	carcinoma	-1.15
susceptibility	0.39	metastatic	-1.15
ovarian	0.33	screening	-1.22
estrogen	0.01	surgery	-1.22

A comparison of the languages of the professionals and that used for patients shows similar disparity in the use of some of the terms (see Table 6b). Terms like *irradiation*, *ovarian* and the newly discovered *brca1* and *brca2* are used more in the professional corpus than in the patient corpus. Terms like *biopsy* and *mammogram* are used more extensively in the patient corpus than in the professional corpus. The inferences we may make are (a) professionals are involved in discussions about concepts/artefacts related to the terms they frequently use which are not yet common knowledge in the patient corpus and (b) having established concepts/artefacts some time ago, like mammograms, professionals are not actively involved in developing these concepts/artefacts further but these established concepts/artefacts are of considerable import to the patients.

Table 6b: The contrastive distribution of relative frequency of the terms in the professional and the patient corpus.

Words	$Log(r_{Professional}/r_{Patient})$	Words	$Log(r_{Professional}/r_{Patient})$
progestin	1.35	lump	-0.06
carriers	0.91	cancers	-0.13
irradiation	0.67	tumor	-0.14
ovarian	0.59	hormone	-0.16
postmenopausal	0.56	diagnosis	-0.19
patients	0.50	screening	-0.34
brca1 & brca2	0.47	mastectomy	-0.45
metastatic	0.39	symptoms	-0.55
adjuvant	0.35	nodes	-0.64
mutation	0.34	lymph	-0.82
tamoxifen	0.08	biopsy	-1.00
carcinoma	0.07	nipple	-1.22
genetic	0.06	mammogram	-1.22

Whilst we can readily compare the use of single words, the comparison of the frequency distribution of compound words in two different corpora is not as straightforward. One method of comparison can be the *rank* correlation of two compound words: the rank of a compound term refers to its frequency in a given corpus. If the order is the same in the corpora, then the correlation will be +1; if the order is reversed in the other then the correlation will be -1. If there is no correlation then the value of the correlation coefficient will be zero. The first comparison will be between expert and professional corpora. We chose the two most frequent words *brca1,* and *brca2* in the expert corpus that suppose sharing concepts with the professional corpus. Table 7 shows a comparison of ranks of compound terms in the expert corpus and the professional corpus. The dominant single term in the expert corpus is *brca* and it is the headword or modifier of many terms in the corpus. The correlation amongst the ranks of *brca*–based compounds in the two corpora is (coeff = 0.92) that is the relative rank-order of the compounds in the two corpora is the roughly the same.

Table 7: The rank-order correlation coefficient of compound terms based on *brca1* & *brca2* where $Rank_{Expert}$, $Rank_{Professional}$ are the rank-order of the compound terms in both expert and professional corpora.

Compound terms	$Rank_{Expert}$	$Rank_{Professional}$
brca1 & brca2 mutations	3	10
brca1 & brca2 genes	4	27
brca1 & brca2 protein	14	47
Correlation	0.92	

Similarly, *therapy* is a dominant term in the professional corpus and a root or stem of many compounds. However, the *therapy*-based compounds do not appear to have the same rank-order in the two corpora– the rank correlation is (coeff = 0.32) as Table 8 shows. What is important to point out is that some kinds of therapy such as *estrogen therapy* and *radiation therapy* were not discussed in the expert corpus at all, which supports the indication of weak relationship between the rank-order of the *therapy*-based compounds in the two corpora. On the other hand, the compound terms of cancer types that could be developed by having an inherited susceptibility or common genes such as *breast, ovarian, prostate* and *family history* indicate a relationship that could not be considered as a significant one between the two corpora (coeff =0.45).

Table 8: The rank-order correlation coefficient of compound terms based on *therapy* where $Rank_{Expert}$, $Rank_{Professional}$ are the rank-order of the compound terms in both expert and professional corpora.

Compound terms	$Rank_{Expert}$	$Rank_{Professional}$
endocrine therapy	37	26
hormone therapy	39	43
adjavant therapy	39	16
tamoxifen therapy	43	31
systemic therapy	43	37
Correlation	0.88	

Consequently, experts conducted deep research related to discovering or verifying the genes that prove the inherited element considering high risk - when having a family history - in developing such types of cancer as the order frequency of these terms was quite high in the expert corpus, while professionals are focused principally on breast cancer and its linkage to other types such as ovarian. Professionals concentrate on the application of such results in their practices, such as therapies, diagnosis and treatments. However, the feedback from professionals and practitioners to the experts is a vital element because innovation that does not have a good application might be obsolete, and a theory that is not put into practice might vanish.

The comparison of the *breast cancer*-based compound has shown a different distribution: the terms *breast cancer*, with *risk*, *patient, carcinoma, families, susceptibility, cells.* The correlation between the rank-order of these terms indicates a weak and negative relationship (coeff=-0.29) as the orders of *breast cancer patients* are roughly the same, while the terms *metastatic breast cancer* and *breast cancer susceptibility* have different order rank in these two corpora. The compound words related to breast cancer types and diagnosis have low rank in the expert corpus. And also the rank-order of *breast cancer families* and *susceptibility* is much higher than in the professional corpus as these concepts are related to other concepts such as the new discovered genes. And this can infer the negative weak relationship between these compound words (see Table 9).

Table 9: The rank-order correlation coefficient of compound terms based on *breast cancer* in the expert and professional corpora.

Compound terms	$Rank_{Expert}$	$Rank_{Professional}$
metastatic breast cancer	42	8
breast cancer patients	15	13
invasive breast cancer	42	20
breast cancer cells	42	31
breast cancer families	22	44

breast cancer susceptibility	25	47
Correlation	-0.29	

We will now discuss the extent of knowledge transfer between professionals and patients. We selected two frequent single terms – *therapy* and *breast* in the two corpora. The established concepts relating to the terms *chemo-*, *radio-*, *psycho-* and *cryo-therapy* have the same frequency order in the two corpora (correlation coefficient=0.87). However, in the order of more recent forms of therapy, for example, *hormone* and *estrogen* replacement to *breast conservation therapy*, the correlation is not quite the same (correlation coefficient =0.5). The frequency order of the terms in which *breast* is the modifier is anti-correlated (correlation coeff =-0.5): the order in the professional corpus is *breast carcinoma, b-tumors, b-tissue, b-reconstruction* and *breast implant*, but in the patient corpus *breast implant* had the top rank.

4.2 A prototype information spider and automatic indexing

We have created a knowledge-based system that was used for facilitating the 'search for reusable knowledge and to structure the knowledge' following the *infospider* of Kalpers *et al* (2002). Recall that MAKE-infospider depends crucially on the attachment of keywords to be stored in the system for subsequent recall. The indexing scheme depends on keywords and on the ability to identify and extract proper nouns. The system we have designed deals with cancer-related information produced by experts, professionals and patients in order to facilitate sharing best practice documents concerning this disease. In this system, the spider has six facets each of which represents a dimension or category: knowledge package or document (K-D) type, scope, process, audience orientation, sharing, and renewable ontology sharing. Each knowledge package is allocated to the meta-information contained within each 'leg'. An example of meta-information for a K-D document is displayed below:

Header Information
Title: Best Practices Of Cancer Diagnosis
K-Doc Type: best practice document
Author: The National Cancer Institute NCI
Publishers: www.Cancer.gov
Description:
Spider categories:
    Audience orientation: Health professional
    Established Terms: radiation therapy, chemotherapy, and hormone therapy, primary tumor
    Neologisms: estrogen-receptor, progesterone-receptor, HER2/neu gene amplification
    Scope: breast cancer
Abstract: **Breast cancer is commonly treated by various combinations of surgery, radiation therapy, chemotherapy, and hormone therapy. Prognosis and selection of therapy may be influenced by the age and menopausal status of the patient, stage of the disease, histologic and nuclear grade of the primary tumor, estrogen-receptor (ER) and progesterone-receptor (PR) status, measures of proliferative capacity, and HER2/neu gene amplification.**
***K-elements*:**
Full text view: \\Liberator\corpus\Breast_Cance: r\test2\1.txt
OriginalSource: sourcehttp://www.cancer.gov/cancerinfo/pdq/treatment/breast/healthprofessiona/
Search:
    Link to related K-Document : http://medline.cos.com/
Link to others search engine: http://www.breastcancercare.org.uk/Professionalresources.htm
Launch Search: *gene amplification*
General information
Date of publish: 10-10-2002
Total words: 5500 words
ID: number:2 Cancer Institute NCI Cancer.gov

The system can index, store and retrieve knowledge packs or document packs including best practice in health-care. The system can also summarise documents to produce an abstract with a summariser developed at the University of Surrey. Further, the system gives practitioners the opportunity to be engaged in communication concerning the K-D document by opening discussion or adding comments to the document in order to share their knowledge. This study has a potentially important impact on the management of the health-care workforce, and is therefore being conducted in conjunction with the University of Surrey's interdisciplinary Healthcare Workforce Research Centre.

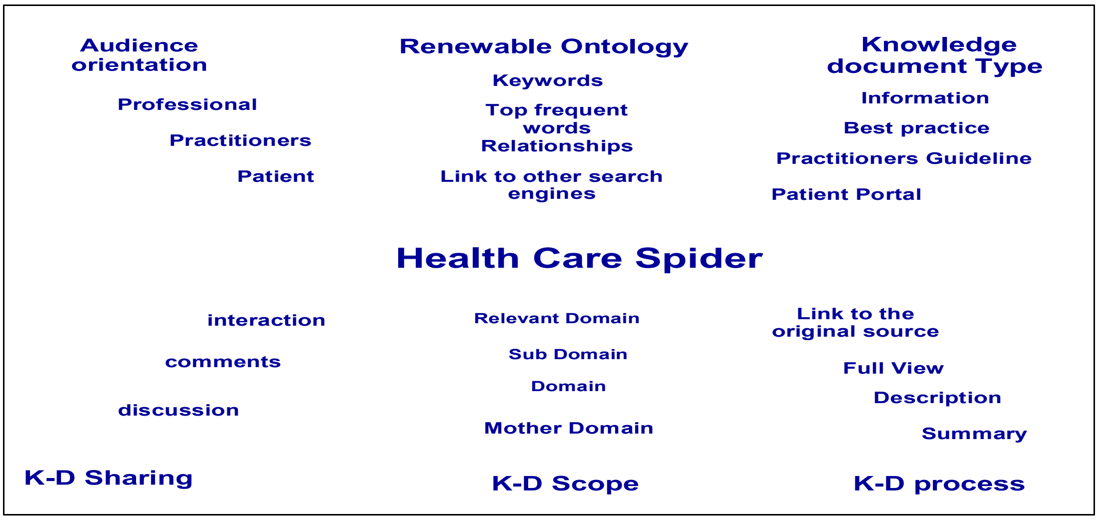

Figure 1: The Surrey Health-care Infospider

5. Conclusion

Knowledge sharing is facilitated through a number of different knowledge sharing or creation modes. We have argued that the successful completion of each of the modes manifests itself either through an understanding of terminology (for example the socialisation mode and internalisation mode) or through the production of documents as in externalisation and combination modes. The trace of knowledge of individuals and organisations, that is, written documents within the archives of a given domain, comprises much of the discernible knowledge of the domain. One of the major problems in knowledge sharing is the accessibility to documents within the archives, especially within a rapidly changing domain. For instance, terms used for indexing documents at an earlier stage of the evolution of the domain may become irrelevant to documents subsequently produced. Terms familiar to individuals at a given level of expertise may be quite opaque to individuals at a different level of expertise.

Terminology of a specialist domain emerges over time. The terminology in itself is a part of the wider language of everyday use with specialist meanings. A systematic extraction of these terms will obviate some of the challenges in accessing documents and, when accessed, understanding them. Our Infospider perhaps demonstrates the synergy between language and knowledge in domains as diverse as cancer therapy.

6. Acknowledgment

The computations reported here were carried out using System Quirk, a text and terminology management system that developed by University of Surrey to facilitate the creation and analysis of text corpora. Texts were captured by using the UK Universities Joint Academic Network and e-journal subscriptions of the University of Surrey. A number of public domain texts were also used

Thanks should also be addressed to the British Council in acknowledgment of their research scholarship. This research was supported by the EU co-funded project Generic Information based Decision Assistant GIDA IST-2000-31123, and SOCIS project GR/M89041.

References:

Al-Thubaity, A.B. and Ahmad, K. (2003). “Knowledge Maps as Lexical Signatures of Journal Papers and Patent Documents.” In Ebad Banissi *et al.* (eds.) *Proc. of 7th International Conference on Information Visualisation* (London, England, 16-18 July 2003). Los Alamitos: IEEE Computer Press. 582-588.

Ahmad, K. (2002). “Writing Linguistics: when I use a word it means what I choose it to mean.” In Manfred Klenner and Henriëtte Visser (eds.). *Computational Linguistics for the new millennium: divergence or synergy? Proceedings of the International Symposium held at the Ruprecht-Karls-Universität Heidelberg, 21-22 July 2000 Festschrift in honour of Peter Hellwig on the occasion of his 60th birthday*. Bern: Publishing Group Peter Lang. 15-38.

Ahmad, K. (2001). “The Role of Specialist Terminology in Artificial Intelligence and Knowledge Acquisition.” In S.-E. Wright & G. Budin (eds.) *Handbook of Terminology*

Management, Vol.2. Amsterdam: John Benjamin Publishers. 809-844.
Ahmad, K. and Miles, L. (2001). "Specialist Knowledge and its Management". *Journal of Hydroinformatics* 3(4**)** October 2001. 215-230.
Aston, G. and Barnard, L. (1998). "The BNC Handbook: Exploring the British National Corpus with SARA". Edinburgh: Edinburgh University Press.
Baker, G.P. and Hacker, P.M.S. (1988). "Wittgenstein Meaning and Understanding". Oxford: Basil Blackwell Ltd. 48-56.
Bazerman, C. (1998). "Shaping Written Knowledge: The Genre and Activity of The Experimental Article in Science". Madison: University of Wisconsin Press.
Cancer. (2003) [ONLINE] http://www.cancer.gov accessed 27 July 2003.
Choi, B. and Lee. H. (2002). "Knowledge Management Strategy and its Link to The Knowledge Creation Process." *Expert Systems with Applications,* 23(3). 173–187.
Chute, C., Cohn, S. and Campbell, J. (1998). "A Framework for Comprehensive Health Terminology Systems in The United States: development guidelines, criteria for selection, and public policy implications". *Journal of American Medical Association, (JAMA*) 5(6). 503–510.
Cimino, J.J. (1996). "Formal Descriptions and Adaptive Mechanisms for Changes in Controlled Medical Vocabularies". *Methods of Information in Medicine* 35(3). 202-210.
Davenport, T., and Probst, G. (2002). "Knowledge Management Case Book Siemens Best Practises." 2nd edition. Munich: Publicis Corporate Pub., and John Wiley & Sons.
Fisher, G. and Ostwald, J. (2001). "Knowledge Management: Problems, Promises, Realities, and Challenges", *IEEE Intelligent Systems*, 16(1). 62 .
Gibbert, M., Jonczyk, C., & Völpel, S. (2000). "ShareNet – The Next Generation Knowledge Management". In (Eds.) T. Davenport & G. Probst. 22-39.
Halliday, M.A.K. and Martin, J.R. (1993). "Writing Science – Literacy and Discursive Power". London and Washington: The Falconer Press.
Kalpers, S., Kastin, K., Petrikat, K., Scheon, S., and Spath, J. (2002). "How to Manage Company Dynamics: An approach for Mergers and Acquisitions Knowledge Exchange", In (Eds) T. Davenport and G. Probst. 187-206.
Morgan, G. (1996). "Images of Organization". 2nd edition, London: SAGE Publications.
Nonaka, I. and Takeuchi, H. (1995.) "The Knowledge-Creating Company". New York: Oxford University Press, Inc.
Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). "A Comprehensive Grammar of the English Language". London and New York: Longman.
Ruch, P., Wanger, J., Bouillon, P., Band, R., Rassinoux, A., Scherrer, J. (1999). "MEDTAG: Tag Semantic for Medical Document Indexing". *American Medical Informatics Association (AMIA) Annual Symposium*, November.
Sager, J.C., Dungworth, D. and MacDonald, P.F. (1980). "English Special Languages – Principles and practice in science and technology". Wiesbaden: Oscar Brandsetter Verlag–KB.
Seely-Brown, J. (1998). "Research that Reinvents The Corporation". *Harvard Business Review On Knowledge Management.* Boston: Harvard Business School Press. 153-180.
Schröder, H. (1991). (Ed.) "Subject-Oriented Texts – LSP and text Theory". Berlin and New York: Walter de Gruyter.
Scarbrough, H. (1996). "The Management of Expertise". *Macmillan Business.* London: Macmillan Business Press LTD. 83-89

