## **Communications of the Association for Information Systems**

#### Volume 17

Article 40

June 2006

# Information Systems and Health Care IX: Accessing Tacit Knowledge and Linking It to the Peer-Reviewed Literature

Michael Shepherd Dalhousie University, shepherd@cs.dal.ca

Syed Sibte Raza Abidi Dalhousie University

Qigang Gao Dalhousie University

Zhixin Chen Dalhousie University

Quifen Qi Dalhousie University

See next page for additional authors

Follow this and additional works at: https://aisel.aisnet.org/cais

#### **Recommended** Citation

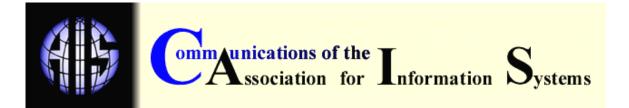
Shepherd, Michael; Abidi, Syed Sibte Raza; Gao, Qigang; Chen, Zhixin; Qi, Quifen; and Finley, G. Allen (2006) "Information Systems and Health Care IX: Accessing Tacit Knowledge and Linking It to the Peer-Reviewed Literature," *Communications of the Association for Information Systems*: Vol. 17, Article 40. DOI: 10.17705/1CAIS.01740 Available at: https://aisel.aisnet.org/cais/vol17/iss1/40

This material is brought to you by the AIS Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in Communications of the Association for Information Systems by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

# Information Systems and Health Care IX: Accessing Tacit Knowledge and Linking It to the Peer-Reviewed Literature

#### Authors

Michael Shepherd, Syed Sibte Raza Abidi, Qigang Gao, Zhixin Chen, Quifen Qi, and G. Allen Finley



## INFORMATION SYSTEMS AND HEALTH CARE IX: ACCESSING TACIT KNOWLEDGE AND LINKING IT TO THE PEER-REVIEWED LITERATURE

Michael Shepherd Syed Sibte Raza Abidi Qigang Gao Zhixin Chen Qiufen Qi Faculty of Computer Science Dalhousie University <u>shepherd@cs.dal.ca</u>

G. Allen Finley Departments of Anaesthesia & Psychology Dalhousie University

#### ABSTRACT

Clinical decision-making can be improved if healthcare practitioners are able to leverage both the tacit and explicit modalities of healthcare knowledge, yet at present there do not exist knowledge management systems that support any active and direct mapping between these two knowledge modalities. In this paper, we present a healthcare knowledge-mapping framework that maps (a) the tacit knowledge captured in terms of email-based discussions between pediatric pain practitioners through a Pediatric Pain Mailing List (PPML), to (b) explicit knowledge represented in terms of peer-reviewed healthcare literature available at PubMed. We report our knowledge mapping strategy that involves methods to establish discussion threads, organize the discussion threads in terms of topic-specific taxonomy, formulate an optimal search query based on the content of a discussion thread, submit the search query to PubMed and finally to retrieve and present the search results to the user.

Keywords: tacit knowledge, explicit knowledge, knowledge transformation, clinical decision-making

#### I. INTRODUCTION

Knowledge Management (KM) is a conscious and coordinated strategy to manage the knowledge assets within an enterprise in order to generate knowledge-mediated services that aim to improve outcomes, innovation and communications. In the realm of healthcare, knowledge management involves the identification, acquisition, sharing and operationalization of healthcare knowledge to bridge the perceived knowledge gaps in order to provide knowledge-mediated clinical decision support, patient education and policy planning. Functionally speaking, healthcare knowledge management strategies and methods aim to establish a knowledge-centric healthcare delivery environment, whereby healthcare stakeholders are able to seek and share both tacit and explicit healthcare knowledge resources. This presents an interesting set of challenges because (i)

Accessing Tacit Knowledge And Linking It To The Peer-Reviewed Literature by M. Shepherd, S.S.R. Abidi, Q. Gao, Z. Chen, Q. Qi, and G.A. Finley

healthcare knowledge is quite dynamic—the volume of additions, revisions and critiques to healthcare knowledge grows at such a rapid rate that it is becoming increasingly difficult to identify, access and consume new knowledge on a continuous basis; (ii) the generation and utilization of healthcare knowledge is contingent on the prevailing clinical context—i.e. healthcare knowledge for cardiovascular diseases is differentiated along the contexts of diagnosis, therapy, prognosis, rehabilitation and so on; and (iii) healthcare knowledge exists in a variety of modalities—i.e. tacit, explicit, experiential and social—that seamlessly interrelate with each other during a healthcare process.

Typically, healthcare knowledge is characterized along the lines of explicit and tacit knowledge: explicit healthcare knowledge encapsulates evidence whereas tacit knowledge represents the clinical experiences of healthcare practitioners [Wyatt, 2001]. This theoretical distinction between healthcare knowledge prevails in existing knowledge management systems that are constrained by the use of a single knowledge modality when providing decision-support. Typically, knowledge management systems deal with either some variation of explicit knowledge represented as either documents, guidelines/workflows, symbolic rules or tacit knowledge represented either as cases, scenarios or peer discussions [Abidi et al., 2004; Curran-Smith et al., 2005; Montani and Bellazzi, 2002; Rossille et al., 2005]. This approach is quite limiting as it provides only a single perspective to the problem's understanding and its solution; furthermore the said approach is guite contrary to the human decision-making process which involves an active introspection, selection and interaction between all knowledge modalities around a particular topic in order to derive decisions. We know that in practice, healthcare practitioners are guite competent in mapping tacit knowledge to explicit knowledge (and vice versa), as per the prevailing clinical context, in order to derive clinical decision/recommendation, to complement or critique existing а knowledge/understanding in order to make better clinical decisions. For instance, when addressing a clinical problem, practitioners may initially refer to the best evidence (i.e. explicit knowledge) available in terms of medical literature and then subsequently refer to corresponding clinical experiences of domain experts (i.e. tacit knowledge) to appraise, validate or confirm their understanding.

Clinical decision-making can be significantly improved if healthcare practitioners are able to simultaneously leverage both the experience of healthcare experts and the corresponding published evidence [Abidi et al., 2005]. The potential linking of tacit knowledge to explicit knowledge (i.e., published literature) will have a significant impact on healthcare practitioners when dealing with atypical and ill-represented/researched medical conditions [Abidi et al., 2004]. Yet, at present there are no mechanisms whereby these two vital modalities of medical knowledge can concurrently be presented to healthcare practitioners as 'holistic' knowledge. Notwithstanding the complexity of healthcare knowledge and the associated healthcare delivery practices and processes, we argue that innovative knowledge management strategies need to be developed to provide healthcare topic, such that healthcare practitioners are able to seamlessly navigate between these two modalities in order to retrieve all the knowledge pertinent to the clinical context. Current knowledge around a specific topic to corresponding explicit knowledge during an active problem solving activity (or vice versa).

In this paper, we present our healthcare knowledge-linking framework for mapping tacit and explicit healthcare knowledge modalities pertaining to the management of pediatric pain [Abidi et al., 2004]. The tacit knowledge is captured in terms of email-based discussions between pediatric pain practitioners through a Pediatric Pain Mailing List (PPML). The explicit knowledge manifests in terms of peer-reviewed healthcare literature available at PubMed, which is an online healthcare literature repository. We have developed a four-stage knowledge mapping strategy that allows linking the tacit knowledge in a discussion thread to corresponding published articles (i.e. explicit best evidence) in order to provide a multi-perspective outlook to the pediatric pain issue. The linking is achieved by (1) cleaning the content and normalizing the terminology in the email-based discussions; (2) linking the emails around a similar topic as a discussion thread and then organizing the discussion threads based on a taxonomy of clinical concepts (or topics); (3)

generating an optimal literature search query from the context and content of the discussion thread; and (4) submitting the search query to PubMed to retrieve healthcare articles from PubMed. In practice, through our knowledge linking framework a pediatric pain practitioner can (a) navigate through the various discussion threads—each discussion thread comprises a continuous sequence of individual email messages around a specific topic—about pediatric pain topics to leverage the experiences of a community of pediatric pain practitioners; and (b) choose to retrieve the corresponding published literature around the discussion topic by simply selecting the discussion thread. It is our contention that the potential linking of tacit knowledge (i.e. PPML) to explicit knowledge (i.e. published literature) will help pediatric pain practitioners to deal with atypical and ill-represented/researched medical conditions for which both experience and evidence is simultaneously required.

We discuss the knowledge management strategies developed in this research project to achieve the abovementioned tacit-explicit knowledge linking. The key task of tacit knowledge organization was pursued through machine learning and ontology-based methods, and we present a comparative analysis of these methods. Finally, we present a working example to demonstrate the mapping of tacit knowledge within a discussion thread to medical articles at PubMed.

Section 2 of this paper describes the tacit knowledge artifact. Section 3 describes our knowledge mapping strategy. Section 4 discusses methods to clean the tacit knowledge and transform it to discussion threads. Sections 5 and 6 describe approaches for the organization of the threads. Section 7 presents methods to link to PubMed. Section 8 illustrates a working example of linking a discussion thread with medical literature. Section 9 presents discussion and future research.

#### II. PEDIATRIC PAIN MAILING LIST: A KNOWLEDGE ARTIFACT

The existence of pediatric pain has been recognized widely in the past fifteen years. Although knowledge about pediatric pain has accumulated over this period, it is still under-treated. Many pediatric pain problems are relatively rare and it is impossible for even a pediatric pain specialist to have personal experience with all possible symptoms and syndromes. In addition, clinical research has been limited and there is little in the medical and scientific literature to support clinical decision-making. Therefore, the Pediatric Pain Mailing List (PPML) has become an important valuable information resource for clinicians, researchers and patients.

The PPML is an international forum for clinical discussion on pediatric pain between pediatric pain professionals [Bodenreider and McCray, 2003]. PPM is a list server that permits informal email-based communication whereby individuals can post/reply to e-mail messages that are then sent to all the subscribers and the messages themselves are archived at the central server. At present there are over 700 subscribers including clinicians, researchers, academics and editors of major current texts in pediatric pain. Currently, the PPML archive contains more than 10,000 messages covering various pediatric pain topics. The message content includes clinical problems, research problems or proposals, successful experiences, suggestions and administrative aspects of children's pain management and prevention. The messages are stored in a raw e-mail format and that makes it difficult to retrieve knowledge about a specific topic [Abidi et al., 2004].

In practice, PPML promotes the articulation of problem-specific tacit knowledge and clinical experience of pediatric pain specialists, which might not otherwise be shared or made explicit via published medical literature. The workflow of PPML is as follows: (1) a practitioner initiates discussion on a specific pediatric pain issue; (2) multiple pediatric pain practitioners discuss the issue—a debate involving experts ensues during which they share experiences, relate theory to practice and collaboratively suggest a solution; (3) expert's tacit knowledge explicated during the discussion is disseminated to various pediatric pain practitioners. We believe that such discussions capture the accumulated experience and clinical judgment of many expert practitioners that can impact the delivery of healthcare [Chen et al., 2006]. From a knowledge management perspective, based on Nonaka's knowledge creation model [Nonaka, 1994], the

PPML workflow features the process of externalization—i.e. the identification, elicitation and codification of tacit knowledge such that it becomes explicit knowledge—of the cumulative tacit knowledge of a vibrant community of knowledge workers that informally share their experience and expertise around pediatric pain management.

Although the PPML is a validated repository of tacit healthcare knowledge on pediatric pain management, the encapsulated tacit knowledge is not fully exploited because the knowledge is not (a) organized along distinct topics; (b) accessible through a simple 'query-based' interface; and (c) linked with additional knowledge resources, such as medical literature repositories. In our research we aim to apply knowledge management techniques to address the abovementioned limitations so that the PPML becomes an online knowledge resource for the pediatric pain community.

#### III. PEDIATRIC PAIN KNOWLEDGE MAPPING STRATEGY

The broad objective of the KM research program is to transform the PPML into a pediatric pain knowledge resource—i.e. to identify, organize the tacit knowledge of pediatric pain practitioners inherent within PPML discussions and then to map it to corresponding explicit knowledge at PubMed.

From a KM perspective, within the PPML there is no accommodation for the process of socialization in transforming tacit knowledge to explicit knowledge as there is no means for synchronous communication as might be found in chat rooms or supported by groupware. The PPML does, however, support the process of externalization as tacit knowledge is elicited as people ask questions through the list server and those with tacit knowledge respond, asynchronously. The tacit knowledge captured is that of the responder, i.e., what the responder knows from experience. The post-reply dialog explicates this tacit knowledge in electronic terms. Yet, the explicated tacit knowledge accumulated in the PPML archives was very difficult to access, as it was not organized in any manner other than chronologically. This made it difficult to provide the subscribers with information and knowledge discovery from the archive. Therefore, we developed methods to organize these discussion threads (analogous to the process of combination) and provide users with access to these threads (analogous to the process of internalization).

In this section we highlight the main elements of our knowledge mapping strategy:

- 1. Cleaning the PPML by removing emails not related to pediatric pain management.
- **2.** *Establishing* a discourse for the email messages by threading all emails around a particular topic/issue within a given timeframe. This is to realize a discussion thread that organizes all related emails along the progression of a meaningful discussion between multiple participants.
- **3.** Organizing the discussion threads in a topic-specific hierarchy. This involves the realization of a taxonomy of discussion topics together with the set of discussion threads associated with each discussion topic. We experiment with two different approaches—i.e. a text clustering approach vs. an ontology based approach—for organizing the discussion threads. The approach offering both better organizational structure and ease of search query generation will be selected.
- 4. Identifying medically salient terms from a given discussion thread to formulate a search query for querying PubMed for medical articles [Abidi et al., 2005]. This requires parsing a topic-specific discussion thread to firstly identify constituent medical terms and then to filter out less significant terms. It may be noted that a small yet focused set of terms is desired to query PubMed in order to achieve a high impact retrieval result. In our work, we plan to perform term filtering at the higher level of semantically-motivated concepts as opposed to linguistic terms. The outcome is a small yet focused set of medical terms that

best represent the selected discussion thread, and will serve as the query content for retrieving corresponding articles from PubMed.

- **5.** *Providing* users the ability to specify literature search criterion by allowing them to select pre-defined search filters available at PubMed.
- **6.** *Formulating* the search query by combining both query content and the user's preferences. The search query is to be automatically submitted to PubMed and the retrieved articles are captured and organized for presentation to the user.
- **7.** *Presenting* the results of the knowledge mapping exercise—i.e. medical articles retrieved from PubMed—to the user.

In the following section we discuss in detail the technical methods deployed to pursue the proposed knowledge mapping strategy (as depicted in Figure 1).

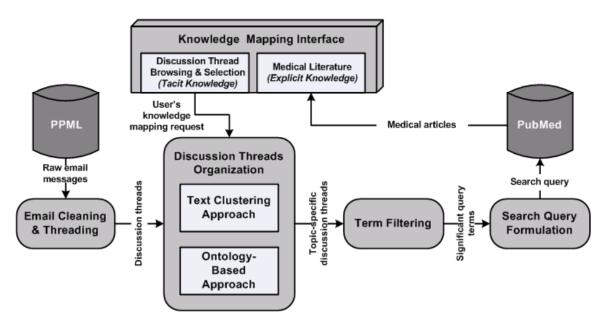


Figure 1: Schematic of Our Knowledge Mapping Strategy, Highlighting the Main Tasks and Their Input and Output.

#### IV. CLEANING AND THREADING THE EMAIL MESSAGES

As a first step towards organizing the PPML in terms of meaningful pediatric pain topics, we performed the following steps: (a) Filtering out 'noisy messages; (b) Establishing a discussion thread around a specific pediatric pain topic, where each discussion thread may comprise multiple topic-specific emails.

Filtering the 'noisy' messages involved removing duplicate messages as identified by subject and date/time stamps; automatic responses generated by "vacation" mail programs; junk e-mails containing announcements, promotions, news items; and personal correspondences. Initially, the cleaning of PPML was done manually and later we developed programs that used pattern matching methods to identify noisy emails.

As the PPML originally started simply as an e-mail server without the facility to create "topics" for discussion, the resulting collection of emails were recorded as isolated communiqués between two parties. However, in many cases a pediatric pain topic resulted in a multitude of emails being

exchanged between the participants. We believe that it is the sequence of emails, as opposed to a single email, that encapsulates the tacit knowledge of the discussion participants around the specific pediatric pain topic. Hence, we identified and established discussion threads—i.e. identifying the originating email postings and the subsequent chain of e-mail responses to those previous postings. The emails were threaded based on time stamps and subject headings. Those messages that had a blank subject field were processed based on the included original messages to which they had replied. Lewis and Knowles [1997] found that threading based on subject field content and on included original messages gives reasonably high thread coherence.

Prior to cleaning and threading the PPML messages, we had 6939 messages from the years 1993 to 1999. After cleaning and threading process, the PPML was reduced to 4033 messages that were organized in 1289 unique discussion threads.

A sample thread is shown in the Appendix, with all identifying information removed. Note the expression of tacit information in the responses to the stated question. We suggest that the threading of the e-mail messages is synonymous to the process of externalization, i.e., the transformation of tacit knowledge into explicit knowledge. And, these discussion threads serve as explicit knowledge artifacts that are linked to explicit knowledge resources as described in the following section.

#### V. THREAD ORGANIZATION: TEXT CLUSTERING APPROACH

At the conclusion of stage 1, we generated a collection a discussion threads covering a variety of pediatric pain topics; some topics may have resulted in more than one discussion thread. For efficient viewing of the discussion threads it was deemed important that the discussion threads be systematically organized based on their content—the organization scheme was required to not only group the discussion threads belonging to the same topic but to additionally identify the salient topics and sub-topics manifested within the discussion threads.

In our work, we pursued full content-driven organization of the discussion threads. Two different approaches were explored: (i) text clustering using different machine learning methods; and (ii) ontological classification using UMLS semantic types. In the following section we discuss both these approaches in detail and compare the results with classifications performed by human experts.

Text clustering is an unsupervised learning process of grouping documents into clusters so that the documents within a cluster have high similarity with one another, but are very dissimilar to the documents in the other clusters [Han and Kamber, 2001]. It was initially used as a way of improving the efficiency of best match searching in information retrieval systems [Willett, 1988]. More recently, it has been used in browsing a large collection of documents [Cutting et al., 1992] and in organizing the results returned by a search engine to help users find relevant documents more quickly [Zamir, 1997]. Document clustering has also been used to automatically generate hierarchical clusters of documents [Koller and Sahami, 1997]. Two recent papers [Berkhin, 2002; Jain et al.; 1999] offer a comprehensive survey of different clustering algorithms and applications.

In our work, two different clustering approaches were investigated. The first approach is based on the k-means clustering algorithm and the second approach is based on Self-Organizing Maps (SOM). In both approaches, each thread was treated as though it was a contiguous document and was represented by a weighted term vector. The original messages that were embedded in the reply messages are removed. The terms from the remaining text were checked against an augmented list of stop words [WAIS, 2004]. If a word was not on the stop list, it was matched against a synonym dictionary manually created by a pediatric pain specialist to ensure terminology standardization. These terms were then stemmed using Porter's algorithm [Porter, 2004]. The stemmed terms were assigned weights based on their tf.idf (term frequency – inverse document frequency) values [Korfhage, 1997], so that the resultant vector for each thread consists of weighted stemmed terms.

#### K-MEANS CLUSTERING APPROACH

The k-means text clustering algorithm [Sebastiani, 2002; Steinback et al., 2000] is a top-down or divisive algorithm that partitions the dataset into a non-hierarchical set of k clusters. Repeated application of the k-means algorithm produces a hierarchy of arbitrary breadth and depth. Assuming that the dataset of threads is the initial cluster, the algorithm is repeated until all the clusters at the bottom level of the hierarchy contain k or few threads.

In our experiments we set k equal to 2, 4, 6, 8 or 10, and generated clustering hierarchies of different breadths and depths. In general, the larger the value of k, the broader and shallower is the resulting cluster hierarchy because the algorithm tends to subdivide each cluster into k subclusters at each level. Figure 2 is an example snapshot of the resulting system, showing the root (Pediatric pain) and the first level of sub-clusters. Note that the user has chosen to view the hierarchy generated with k=6. Thus, the hierarchy shown has 6 sub-clusters at each level. Each cluster is represented by the fifteen highest ranked terms in the discussion thread vectors in that cluster. A single thread may be viewed by traversing to the bottom of the hierarchy and clicking on a thread.

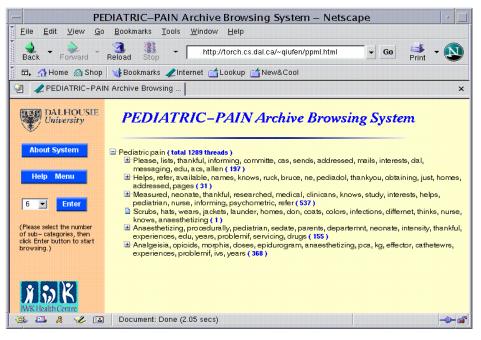


Figure 2. Root and Top Level of Clusters for k=6

#### SELF-ORGANIZING MAP APPROACH

The SOM [Kohonen, 1982] model is an unsupervised clustering method to represent multidimensional data in lower dimensional spaces-typically two-dimensional maps. The SOM employs an unsupervised learning algorithm that eventually realizes clusters based on the affinity between the data points whilst preserving the topological relationships among the data.

In our work, we first used Principal Component Analysis (PCA) to map the 4111 term vectors into a smaller vector space, prior to the application of the SOM algorithm. This reduced the number of features describing each thread significantly. The eigenvalues dropped off quickly and approached zero and stabilized at the 150<sup>th</sup> eignvalue, indicating that almost all of the information can be captured by the first 150 eigenvectors or features. This was verified by the experiments described below in which a wide range of the number of PCA features was used and the best results occurred for the first 150 features.

We experimented with three different variations of the SOM approach:

- 1. Each individual cell in the resulting SOM was treated as an individual cluster. We performed a number of experiments, whereby in each experiment we varied the number of features from 100 to 4111 (all features), the number of cells in the SOM and the dimensions of the SOM from as small as a 2\*2 to the biggest map comprising 8\*6 units.
- 2. A hierarchical SOM was developed using the Growing Hierarchical SOM algorithm [Dittenbach et al., 2002], again treating each resulting cell as a separate cluster. We performed a number of experiments by varying the number of features, the number of layers or depth of the hierarchy, the number of sub-trees or branches at each level.
- 3. Extending the first SOM approach by treating each cell as a unit and clustering these units into a fewer number of larger clusters. Each cell was represented by the centroid of the threads assigned to that cell and then the k-means clustering algorithm was applied to these units. This version of the k-means algorithm created a single layer of clusters, not a hierarchical clustering. The number of features was fixed at 150 but the size and dimensionality of the SOM maps were varied. In all, 31 different maps were created. Once each SOM was created, the k-means algorithm was applied to that map. The value of k was varied from 1 through 20 and, for each value multiple iterations were run with k randomly selected units as seeds for the clusters. The Davies-Bouldin Index [Stein et al., 2003] was used to determine the highest quality clustering for each value of k. The Davies-Bouldin Index is based on the density of each cluster and distance between clusters on the assumption that a good cluster structure would have clusters in which the elements of each cluster are close together and in which the clusters themselves are well separated.

#### **EVALUATION OF CLUSTERING APPROACHES**

We compared the quality of the different clustering outputs against the clustering performed by two experts in the field of pediatric pain. One hundred threads were selected randomly from the dataset. These threads were presented to the two experts and they were asked to independently partition these threads into a flat (one-level) set of clusters. As can be seen, the resulting sets of clusters are quite different. Expert 1, a psychologist, generated 13 clusters (Table 1) and Expert 2, a medical doctor, generated 17 clusters (Table 2). The class labels the experts assigned are also quite different.

Cluster ID	Cluster Label	Number of Threads
1	adverse effects of a treatment or medication, monitoring requirements	6
2	advice on medications or treatment technique for a particular condition	33
3	announcement of a publication or event	9
4	assessment methods for a particular condition	2
5	availability and benefits of a nonstandard drug compound	3
6	availability and validation of a particular assessment tool	11
7	contact information or other information about a specific person	3
8	dosage or adverse effects or technique for a medication or other treatment	9
9	information on a condition: description, etiology, prognosis	4
10	job description, posting of job or fellowship	3
11	Miscellaneous	3
12	other newsgroups and listservs	2
13	policies, guidelines, protocols, algorithms, quality assurance, supervision, competency	12

Cluster ID	Cluster Label	Number of Threads
1	Assessment	10
2	Musculoskeletal	2
3	sedation & procedures	4
4	oral drugs	5
5	miscellaneous /irrelevant/out-of-date	21
6	neuropathic pain	9
7	regional analgesia	10
8	postoperative pain	5
9	intravenous opioids	11
10	Psychology	5
11	visceral pain, bowel function, etc	2
12	topical analgesia, EMLA	2
13	everyday pain	1
14	NMDA antagonists, ketamine	1
15	Resources	8
16	Administration	3
17	Burns	1

Table 2. Clusters and Labels Created by Expert 2 (Medical Doctor)

We measured inter-rater reliability using an information theoretic measure of redundancy [Tague and Shepherd, 1978; Young, 1971]. The Redundancy(X, Y) is the proportion of uncertainty about X that is removed by knowing Y, and is given by Redundancy(X, Y) = (H(X) - H(X|Y)) / H(X). In this instance, X and Y represent the two sets of clusters generated by the experts. The measure is asymmetrical and the calculated redundancy measures are:

R(Expert-1, Expert-2) = 0.51

R(Expert-2, Expert-1) = 0.44

The inter-rater reliability between the two expert-generated sets of clusters is therefore noted to be between 44 and 51 percent, indicating that the identified clusters have marginal similarity. As the experts are from different but related professions, one would expect such modest agreements.

We used the F-measure to determine the effectiveness of hierarchical clustering [Steinback et all, 2000]. The F-measure is a traditional effectiveness measure in information retrieval area, and was recently introduced to document clustering [Larsen and Aone, 1999]. For each humanly generated set of clusters, an F-measure value was calculated for each cluster and an overall F-measure was then calculated as a measure of the hierarchical clustering as compared to that expert's clustering [Qi et al., 2005]. For this experiment, the k-means clustering was done on the same 100 threads that the human experts clustered. An overall F-measure was calculated for each clustering that reflects the quality of the resulting hierarchy, i.e., the closeness between the resulting hierarchy and manually generated clusters. It ranges from 0 to 1 where 1 is the best quality.

#### **EVALUATION OF K-MEANS CLUSTERING**

This evaluation was conducted as described above for all five values of k for the hierarchical clusters. For any one value of k, different clusters may be generated, depending on which threads are randomly selected to represent the initial centroids of the clusters being generated. Therefore,

for each value of k, the algorithm was run ten times with different threads randomly selected to act as initial centroids, resulting in ten different hierarchical classifications.

The best results (highest overall F-values) were found for k=6. The results of the paired-samples t tests (p=0.05) show that there was no significant difference between the two sets of manually generated clusters on the overall F-measures. The average overall F-value for expert 1 is 0.47 and for expert 2 it is 0.48. In other words, the k-means clustering algorithm generated yet a different set of clusters and there is no significant difference between the inter-rater reliability between the k-means clustering and expert 1 and the k-means clustering and expert 2.

#### **EVALUATION OF SOM CLUSTERING**

We applied the F-measure to the results of all three of the SOM approaches to compare it against the two experts. Table 3 shows the experimental results for the best SOM for each experiment. For the general SOM, the results were not as good as the results obtained from the k-means approach. For the GHSOM, although the results were better than for the first SOM approach, the results were only approximately equivalent (but not quite) to the average F-measure results for the k-means approach from the previous research. The SOM-k yielded the lowest F-measures of all the approaches, even though the Davies-Bouldin index was applied and the resulting number of clusters was similar to the number of clusters created by the domain experts (13 and 17, respectively). In summary, none of the above SOM-based approaches provided F-measures as good as the average k-means clustering from the original research.

Method	Features	Map Size	Number of	Best F-Measure (SOM)		Average F-Measure (k-means)	
Wethou	Teatures		Clusters	Expert 1	Expert 2	Expert 1	Expert 2
SOM	150	8*6	48	0.2968	0.4043		
GHSOM	500	5 layers, 2*2	53	0.4235	0.4466	0.47	0.48
SOM-k	150	10*5	13	0.2783	0.3896		

Table 3. Comparison of Best SOM-Based Results with k-Means, k=6

#### CONCLUSION: TEXT CLUSTERING APPROACH

Our experiments with different text clustering methods led to the conclusion that the kind of 'information content' contained within the email messages could not be reliably clustered without due consideration to the inherent semantics of the discussions. For the different experimental setups, the inter-rater reliability between the two experts and between the various clustering methodologies and each of the experts was always found to be too low to use the emergent clusters with any confidence. We believe that one reason for the low agreement between the automatic text clustering methods and the domain experts is that the domain experts were taking into account the overall context of the email messages, as opposed to just the content which in essence was what the text clustering methods were working on, in classifying the email messages.

In summary, we conclude that text clustering methods—these methods approach classification at the 'term' level—are not suitable for organizing the discussion threads in an elegant and efficient manner. Rather, guided by the classification process of the domain experts we proceed with a knowledge-driven approach, whereby a medical ontology is used to identify salient concepts within the discussion threads and its organization is pursued at the higher 'concept' level.

#### VI. THREAD ORGANIZATION: ONTOLOGY BASED APPROACH

An ontology based approach for discussion thread organization aims to identify the salient concepts within the email-based discussions and then use a medical ontology—the medical ontology relates medical terms to a set of high-level concepts which in turn are grouped in terms of semantic types—to group the email messages based on the constituent concepts along the taxonomy defined by the medical ontology. In our work we use the Unified Medical Language System (UMLS) by the National Library of Medicine, featuring the MeSH terminology system and the UMLS semantic network, as the medical ontology for organizing the discussion threads.

Our ontology-based approach to classify the discussion threads involves the following tasks: (a) terminology normalization—i.e. all the medical terms within the discussion threads are identified and changed to the standard terminology of MeSH. This has dual advantages, firstly it ensures semantic interoperability between the synonymous terms used in different messages for the same purpose; and secondly it allows the development of efficient queries, comprising terms found in the discussion threads, to query PubMed; (b) conceptual abstraction—i.e. the medical terms are abstracted to higher conceptual levels whereby different terms may belong to the same semantic type. In fact, UMLS abstracts all medical terms to a mere 134 semantic types. In this case, the discussion threads can be classified based on their representative semantic types as opposed to their constituent terms; and (c) concept-level organization—i.e. the discussion threads are organized based on the pre-defined semantic structure offered in terms of the MeSH tree.

In addition to the organization of the discussion threads, the ontology approach also allows a mechanism to generate optimal search queries—whereby the query content is derived directly from a discussion thread—to link a discussion thread (manifesting tacit knowledge) with corresponding medical articles (manifesting explicit knowledge) at PubMed. Functionally speaking, when a medical practitioner viewing a pediatric pain related discussion thread is required to review corresponding evidence/studies, he would simply select the discussion thread and our knowledge management system will automatically generate a search query from the selected discussion thread and retrieve the relevant medical articles from PubMed. In this way, a practitioner may seamlessly navigate between the PPML and PubMed, and may thus be able to simultaneously leverage both past experiences (i.e. tacit knowledge) and evidence (i.e. explicit knowledge) pertaining to a pediatric pain situation. In the below discussion we discuss the steps to the organization of the discussion threads.

#### DISCUSSION THREAD PARSING

The first step is to parse the discussion threads to identify the constituent terms and then assign a semantic type to each identified term. We use a medical text parsing and term identification tool—called MMTx [MetaMap, 2005] provided by the National Library of Medicine to both identify the medical terms and then normalize them to MeSH terminology. The MMTx program performs a shallow syntactic analysis of raw sentences and separates them into phrases, identifies the medical concepts and assigns proper semantic types to them according to the UMLS semantic network. The final result from MMTx includes lists of term phrases and for each such phrase, the corresponding MeSH terms that were matched, the UMLS Semantic Types and a score out of 1000 indicating the goodness of the match.

#### **DISCUSSION THREAD TERM FILTERING**

The second step involves the filtering of medical terms that are deemed of less medical significance [Jain et al., 1999]—i.e. their presence may unnecessarily constrain the query. Our approach, therefore, was to design a term filtering mechanism to identify a small set of MeSH terms representative of a PPML thread. We designed three different term filters and the user has the functionality to chose a particular filter.

Our term filtering approach operates at the semantic/conceptual level as opposed to the term level [Haynes et al., 1994]. The UMLS semantic types associated with each MeSH term is used

as the basis for term filtering. The rationale is that working at a higher level of abstraction—i.e. the semantic level—we can (a) establish a medical context for the thread which can assist in subsequent search for corresponding literature; (b) characterize the entirety of medical terms into a small number of medical concepts—134 semantic types to be exact [National Library of Medicine, 2006]; and (c) design filtering rules that apply to broad semantic types as opposed to focused individual terms [Haynes et al., 1994]. We have developed the following term filters that can be chosen by the user

#### Filter 1: Filtering Based on Semantic Group

This filter makes use of the pre-defined semantic grouping of the UMLS semantic types—the 134 semantic types are distinguished into 15 semantic groups. In our work we have identified three main semantic groups that contain semantic types (and in turn medical concepts and terms) that are representative of the kind of discussions recorded in PPML.

Given a discussion thread, our approach is to retain only those terms that belong to the selected semantic groups and filter out the remaining terms as they are deemed to have nominal significance towards pediatric pain terms/concepts. Table 4 shows the list of semantic groups, the corresponding semantic types for each semantic group and whether terms belonging to the semantic group will be filtered or not.

In essence, we work with only three semantic groups—i.e. DISO, CHEM, ANAT. These groups are deemed relevant because: DISO contains the semantic types of concepts related to medical disorders; CHEM contains the semantic types of concepts related to chemicals & drug. These semantic types are quite pertinent to drug related discussions noted in PPML; and ANAT contains semantic types of concepts related to anatomy, thereby indicating the physical location of the problem.

#### Filter 2: Filtering based on MeSH Tree

This filter uses the pre-defined MeSH trees to filter out less significant terms from the discussion thread. Note that, there are 15 different MeSH trees where each tree pertaining to a specific medical aspect, such as disease, drugs, etc.

In our work, we have used the C (Diseases) and D (Chemical and Drugs) trees. The rationale for selecting these two trees is as follows: (a) The C tree contains disease related MeSH terminology. Since PPML largely deals with disorders and diseases, it was determined to keep the C tree to retain disease related terminology; and (b) The D tree contains MeSH terminology about chemicals and drugs.

As per our approach for filter 2, given a discussion thread we retain terms that belong to the C and D MeSH terms and filter out the remaining ones.

#### Filter 3: Filtering based on Mapping Score of Title

This filter exploits the MMTx mapping scores as the basis for filtering out terms. Note that, each phrase processed by MMTx is transformed to a MeSH term with a mapping score that reflects the goodness of the mapping of the original phrase to a MeSH term. The mapping score ranges between 0 - 1000, where the maximum mapping score of 1000 indicates a perfect mapping. We believe that MeSH terms with a high mapping score are good candidates for subsequent search for literature and hence should be retained. However, we believe that it is also important to consider the physical location of the term; terms that appear in the title are more representative of the context of the discussion as compared to terms in the body of the message. We have designed the following filtering rules that are applied in the sequence that they are presented:

For a term in the title, if mapping score = 1000 then retain the term.

For a term in the title, if semantic type = Age group (T100) then retain the term.

For a term in the title, if semantic group = CHEM | DISO | ANAT AND (mapping score > 800) then retain the term.

For terms in the title, if semantic type = Diagnostic Function (T060) | Therapeutic or Preventive Procedure (T060 | Laboratory or Test Result (T034) AND (mapping score > 800) then retain the term.

Semantic Gro	oups	Semantic Types	Terms Retained
Activities & Behaviors	ACTI	Activity, Behavior, Event, Machine Activity	NO
Anatomy	ANAT	Anatomical structure, Body location	YES
Chemicals & Drugs	CHEM	Amino Acid, Antibiotic, Chemical	YES
Concepts & Ideas	CONC	Classification, Concept Entity	NO
Devices	DEVI	Medical Device, Research Device	NO
Disorders DISO		Acquired Abnormality, Disease	YES
Genes & Molecular Sequence	GENE	Amino Acid Sequence, Gene or Genome, Molecular Sequence	NO
Geographic Areas			NO
Living Beings	LIVB	Age group, Alga, Animal	NO *
Objects	OBJC	Entity, Food, Manufactured Object	NO
Occupations	OCCU	Biomedical Occupation	NO
Organization	ORGA	Organization, Professional Society	NO
Phenomena	PHEN	Biologic Function, Test Result	NO **
Physiology	PHYS	Cell Function, Clinical Attribute	NO
Procedures	PROC	Diagnostic procedure	NO ***

\* Except Age Group (T100)

\*\* Except Laboratory or Test Result (T034)

\*\*\* Except Diagnostic Procedures (T060), Therapeutic or Preventive Procedure(T061)

#### Filter 4: Filtering based on Mapping Score of Body

This filter works similar to that of filter 3, but is applied to the body of the messages. The filtering rules in order of precedence are:

For concepts in the message body, if semantic group =  $CHEM \mid DICO \text{ AND } (mapping \text{ score} = 1000)$  then retain the term.

For concepts in the message body, if semantic group = DICO AND (mapping score > 600) then retain the term.

#### **VII. PUBMED SEARCH QUERY GENERATION**

The previous processing steps yield a set of medical terms that constitute an optimal search query to retrieve related medical literature from PubMed. It may be noted that PubMed allows a few different search options, and in our work we allow the user to exploit the PubMed's search options in conjunction with the search query terms identified earlier. We allow users three search options to query PubMed.

#### **OPTION 1: DEFAULT SEARCH USING UMLS CONCEPTS**

This is the simplest procedure to query PubMed. The query terms identified earlier are used to generate a query, where the query length is restricted to 5 terms in order to ensure a focused query. Users can choose logical operators—i.e. AND and OR operators—with the query terms to generate a more systematic query.

#### **OPTION 2: SEARCH BY CLINICAL STUDY CATEGORY**

This search option leverages the pre-defined search query filters available at PubMed—i.e. classifying the search type as therapy, diagnosis, etiology and prognosis [Koller and Sahami, 1997; National Center for Biotechnology Information, 2005; SUMSearch, 2006]. Users can choose the search type, based on their clinical intention, and use the search query terms to find medical literature that belong to the search type for the specified terms. For each of the four questions, there are two strategies. One is specific search, which aims to exclude as many nonrelevant articles as possible, and another one is sensitive search, which aims to include as many relevant articles as possible.

#### **OPTION 3: SEARCH SYSTEMATIC REVIEWS**

This search option allows the user to search systematic reviews whereby the search attempts to find citations for systematic reviews, meta-analyses, reviews of clinical trials, evidence-based medicine, consensus development conference and guidelines [a3, 14, 28] for their topic of interest.

#### VIII. MAPPING TACIT TO EXPLICIT KNOWLEDGE: EXAMPLE

We provide a working example to illustrate the functionality of these methods to link the tacit encapsulated within knowledge discussion threads to explicit knowledge available at PubMed.

As a first step, the discussion threads were organized based on the UMLS semantic network (Figure 3) and the MeSH tree (Figure 4) [Bodenreider and McCaray, 2003]. Users can traverse through the hierarchical structure (of both representations) to select a thread of interest.

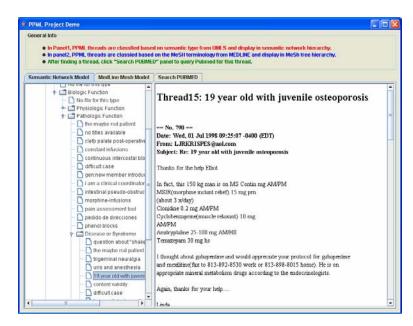


Figure 3. User Interface Showing the Organization of the Threads Based on UMLS Semantic Network

General Info  In Panel 1, IPPML threads are classified based on semantic type from UMLS and display in semantic network hierarchy. In panel 2, IPPML threads are classified based on the MeSH terminology from MEDLINE and display in MeSh tree hierarchy. After finding a thread, click "Search PUDMED" panel to query Pubmed for this thread.					
iemantic Network Model	MedLine Mesh Model	Search PUBMED			
Me8H Tree Hierarchy  Diseases[C]  Pain[C10]  Pinites <1>[C17]		Thread15: 19 year old with juvenile osteoporosis			
		== No. 790 == Date: Wed, 01 Jul 1998 09:25:07 -0400 (EDT) From: LJRKKISPES@ouLcom Subject: Re: J9 vear old with juvenile osteoporosis			
		Thanks for the help Elhot			
Mitochondrial Myopathies(C05.651)     Ghronic Childhood Arthritis(C05.550)     Ischemia[C14]     Pancreattis(C08)		In fact, this 150 kg man is on MS Contin mg AMPM MSIR(morphine instant relief) 15 mg pm (about 3 s/dky) Cloudine 0.2 mg AMPM			
Call anterior uveitis <2>[     Orinary Retention[C	12)	Cyclobenzaprine(muscle relaxant) 10 mg AM/PM			
Chemicals and Drugs(D)     Analytical, Diagnostic and Therapeutic Techniq     Psychiatry and Psychology(F)		Amitryptaline 25-100 mg AM/HS Temazepam 30 mg hs			
Biological Sciences(C)     Anthropology, Education, Sociology and Social     Persons(M)     Health Care(N)		I thought about gabapentine and would appreciate your protocol for gabapentine and mexilinine(fax to 813-892-8530 work or 813-898-8015 home). He is on oppropriate mineral metabolism drugs according to the endocrimologists			
- 🔄 Health Care(N) - 🗂 Geographic Locations(2	a	Again, thanks for your help			
		Linda			

Figure 4. User Interface Showing the Organization of the Threads Based on MeSH Tree

For the purposes of this discussion, we select a discussion thread—thread 15 that encapsulates a discussion pertaining to "19 year old with juvenile osteoporosis—and work it through the various stages. Thread 15 contains 5 email messages exchanged between 3 pediatric pain practitioners. The first part of the thread is shown in the right panel of Figures 3 and 4.

For generating the PubMed search query, we identify the constituent medical terms within both the discussion thread title and content, and a total of 151 unique UMLS concepts are identified for this thread. To filter out the less significant concepts we apply filter 3 (as described earlier) to the title of the thread to identify the candidate terms. Table 5 shows the output of filter 3, note that five UMLS concepts are identified and subsequent filtering based on their semantic group results in

only three concepts to be retained-i.e. Feline osteogenesis imperfecta, Adolescent and Osteoporosis.

Concept Name	Score	Semantic Group	Semantic Type	Retain
Year	694	CONC	Temporal Concept	No
Old	861	CONC	Temporal Concept	No
Feline osteogenesis imperfecta	1000	DISO	Disease or Syndrome	Yes
Adolescent	694	LIVB	Age Group	Yes
Osteoporosis	861	DISO	Disease or Syndrome	Yes

Table 5. The filtering process via filter 3

Finally, the three terms—i.e. Feline osteogenesis imperfecta, Adolescent and Osteoporosis retained by filter 3 are used as the search query for PubMed. The term "Feline" was subsequently filtered out as the data set refers to children, not cats. For this working example, we selected the basic PubMed search option—i.e. option 1 described above. Figure 5 shows the user interface for searching PubMed based on the selected discussion thread. The response of PubMed is illustrated in Figure 6, and it highlights medical articles related to search query derived from discussion thread 19.

emantic Network Model   MedLine Mesh Model	Search PUBMED
Default	
Default Search Use UMLS Concepts	Thread15: 19 year old with juvenile osteoporosis
the second se	
AND OR	
Custom	== No. 790 ==
Query	Date: Wed, 01 Jul 1998 09:25:07 -0400 (EDT)
Give your Query here	From: LJRKRISPES@aol.com
Example: (Head or Back)pain	Subject: Re: 19 year old with juvenile osteoporosis
	Thanks for the help Elliot
Option 1	Thanks for the help Emot
CATEGORY	In fact, this 150 kg man is on MS Contin mg AM/PM
etiology      therapy	MSIR(morphine instant relief) 15 mg prn
	(about 3 x/day)
🔾 diagnosis 🔷 prognosis	Clonidine 0.2 mg AM/PM
SCOPE	Cyclobenzaprine(muscle relaxant) 10 mg
	AM/PM
narrow, specific search	Amitryptaline 25-100 mg AM/HS
broad, sensitive search	Temazepam 30 mg hs
	I thought about gabapentine and would appreciate your protocol for gabapentine
Search By Clinical Study Category	and mexilitine(fax to 813-892-8530 work or 813-898-8015 home). He is on
	appropriate mineral metabolism drugs according to the endocrinologists.
Option 2	
Find Systematic Reviews	Again, thanks for your help

Figure 5. User Interface for Submitting Thread-Based Search Query to PubMed

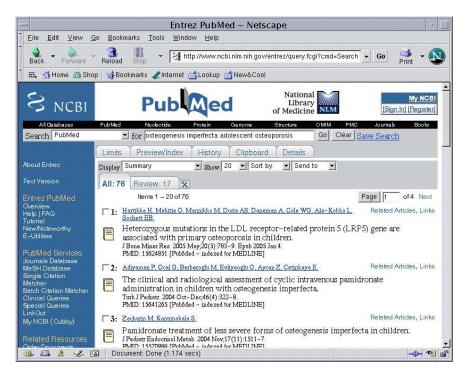


Figure 6. A Sample PubMed Search Query Result

#### IX. SUMMARY AND FUTURE WORK

From a healthcare KM perspective, there is a premium in linking the tacit knowledge of healthcare practitioners with profound explicit knowledge resources. Establishing such a linking leads to knowledge translation because it allows healthcare practitioners to (a) refer to peer moderated discussions on various important topics about pediatric-pain management; (b) refer to summaries of past discussions for just-in-time clinical evidence; and (c) refer to the published medical literature related to a discussion topic.

The objective of this research program is to provide a KM framework to transform the tacit knowledge in an e-mail archive into explicit knowledge. To this end, we developed a strategy and methods to:

- Organize the tacit knowledge inherent in the emails between healthcare practitioners into meaningful and comprehensive topic-specific discussion threads. These discussion threads could then we viewed through different perspectives as per the user's preferences, and users were able to browse and find relevant information within these discussion threads.
- Link the discussion threads through the UMLS and MeSH ontology to the peerreviewed literature in PubMed.

The featured work is the first step in developing a unique tacit-explicit knowledge gateway for healthcare practitioners to support evidence-based clinical decision-making in pediatric pain, which is regarded as a critical evolution in clinical practice.

#### ACKNOWLEDGEMENTS

We would like to thank Dr. Carl von Baeyer, Department of Psychology, University of Saskatchewan, for manually creating one of the sets of clusters for evaluation purposes. The

second set was created by Dr. Allen Finley, Departments of Anaesthesia and Psychology, Dalhousie University.

The paper is based on research presented previously at the Hawaii International Conference on System Sciences, 2005 and 2006. The 2006 paper was recognized as the best paper in the Information Technology and Health Care Track.

#### REFERENCES

EDITOR'S NOTE: The following reference list contains the address of World Wide Web pages. Readers, who have the ability to access the Web directly from their computer or are reading the paper on the Web, can gain direct access to these references. Readers are warned, however, that

1. these links existed as of the date of publication but are not guaranteed to be working thereafter.

2. the contents of Web pages may change over time. Where version information is provided in the References, different versions may not contain the information or the conclusions referenced.

3. the authors of the Web pages, not CAIS, are responsible for the accuracy of their content.

4. the author of this article, not CAIS, is responsible for the accuracy of the URL and version information.

- Abidi S.S.R. et al. (2004) "Knowledge Management in Pediatric Pain: Mapping On-Line Expert Discussions to Medical Literature," 11th World Congress on Medical Informatics (MEDINFO'2004), San Francisco, 7-11 Sept. 2004
- Abidi, S.S.R., M. Kershaw, and E. Milios. (2005) "Augmenting GEM-Encoded Clinical Practice Guidelines with Relevant Best-Evidence Autonomously Retrieved from MEDLINE", *Health Informatics Journal*, (11)1, pp. 57–72.
- Berkhin P. (2002) Survey of Clustering Data Mining Techniques, Technical Report, Accrue Software, Inc.
- Bodenreider, O. and A.T. McCray (2003) "Exploring semantic groups through visual approaches", *Journal of Biomedical Informatics*, 36(6), pp. 414-432
- Chen, Z. et al. (2006) "Accessing Tacit Knowledge in the Pediatric Pain e-Mail Archives", *Proc. of the 39<sup>th</sup> Hawaii International Conference on System Sciences*, Hawaii, January 4-7.
- Curran-Smith, J., S.S.R. Abidi, and P. Forgeron. (2005) "Towards a Collaborative Training Environment for Children's Pain Management: Leveraging an Online Discussion Forum, *Health Informatics Journal*, (11)1, pp. 31–43.
- Cutting, D.R. et al. (1992) "Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections", *Proc. of the 15<sup>th</sup> Annual International ACM SIGIR Conf. on Research and Development in Information Retrieval*, Copenhagen, Denmark, pp. 318-329.
- Dittenbach, M., A. Rauber, and D. Merkl (2002) "Uncovering Hierarchical Structure in Data Using the Growing Hierarchical Self-Organizing Map", *Neurocomputing*, (48)1-4, pp. 199-216.
- Han, J. and M. Kamber, (2001) *Data Mining: Concepts and Techniques*, New York, Morgan Kaufmann Publishers.

- Haynes B. et al. (1994) "Developing Optimal Search Strategies For Detecting Clinically Sound Studies In MEDLINE", *Journal of the American Medical Informatics Association*, pp. 447-458.
- Jain, A.K., M.N. Murty, and P.J. Flynn, (1999) "Data Clustering: A Review", ACM Computing Surveys, (31)3, pp. 264-323.
- Kohonen, T. 1982) "Self-Organized Formation Of Topologically Correct Feature Maps", *Biological Cybernetics*, (43), pp. 59-69.
- Koller, D. and M. Sahami, (1997) "Hierarchically Classifying Documents Using Very Few Words", Proc. of the 14<sup>th</sup> International Conf. on Machine Learning, Nashville, Tennessee, pp. 170-178.
- Korfhage, R.R. (1997) Information Storage and Retrieval, New York, Wiley Computer Publishing.
- Larsen, B. and C. Aone, (1999) "Fast and Effective Text Mining Using Linear-time Document Clustering", *Proc. of the 5<sup>th</sup> ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining*, San Diego, California, pp. 16-22.
- Lewis, D.D. and K.A. Knowles, (1997) "Threading Electronic Mail: A Preliminary Study", Information Processing and Management, (33)2, pp. 209-217.

MetaMap Transfer Home (2005) http://mmtx.nlm.nih.gov/ (current June 15, 2005).

- Montani S, and R. Bellazzi. (2002) "Supporting Decisions In Medical Applications: The Knowledge Management Perspective. *Intl. Journal of Medical Informatics*; (68), pp. 79-90.
- National Center for Biotechnology Information. (2006) *PubMed Clinical Queries*. <u>http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.shtml</u> (current March 28, 2006).
- National Library of Medicine.(2006) *Fact Sheet UMLS® Semantic Network* <u>http://www.nlm.nih.gov/pubs/factsheets/umlssemn.html</u> (current March 28, 2006).
- Nonaka, I. (1994). A Dynamic Theory Of Organizational Knowledge Creation. Organization Science, (5)1, pp.14-37.
- Porter, M. (2004) http://www.tartarus.org/~martin/PorterStemmer/ (current June 9, 2004).
- Qi, Q. et al. (2005) "Accessing Tacit Knowledge in the Pediatric Pain e-Mail Archives", Proc. of the 38<sup>th</sup> Hawaii International Conference on System, Sciences, Hawaii.
- Rossille, D., J. Laurent, and A. Burgun. (2005) "Modeling A Decision-Support System For Oncology Using Rule-Based And Case-Based Reasoning Methodologies", *Intl. Journal of Medical Informatics*, (74), pp. 299-306.
- Sebastiani, F., (2002) "Machine Learning in Automated Text Categorization", ACM Computing Surveys, (34)1, pp. 1-47.
- Stein, B., S.M. zu Eissen, and F. WiBbrock. (2003) "On Cluster Validity and the Information Need of Users", 3<sup>rd</sup> IASTED International Conferencer on Atificial Intelligence and Applications, Benamaldena, Spain, September.
- Steinbach, M., G. Karypis, and V. Kumar, (2000) "A Comparison of Document Clustering Techniques", Proceeding of Text Mining Workshop, KDD, 2000.
- SUMSearch, (2006) Society of General Internal Medicine, University of Texas, USA, http://sumsearch.uthscsa.edu/ (current March 28, 2006).

- Tague, J. and M. Shepherd, (1978) "Redundancy as a Measure of Classification Similarity", *Journal of Informatics*, (2)3, pp. 123-134.
- WAIS.(2004) (Version:iubio-wais-8-b5) <u>http://www.iiasa.ac.at/docs/R\_Library/ libsrchs.html</u> (current June 9, 2004).
- Willett, P., (1988) "Recent Trends in Hierarchic Document Clustering: A Critical Review", Information Processing & Management, (24)5, pp. 577-597.
- Wyatt J C. (2001) "Management Of Explicit And Tacit Knowledge", *Journal of the Royal Society of Medicine*, (94).

Young, J.F., (1971) Information Theory, New York, Wiley.

Zamir, O. et al. (1997) "Fast and Intuitive Clustering of Web Documents", KDD,, pp. 287-290.

#### APPENDIX: SAMPLE THREAD: OPIODS AND MENINGITIS

Cluster Keywords: Opioids, days, ivs, guillain, meningococcemia, ptosis, spontaneous, agonist, annequin, nubain, relieve, neurologist, nsaids, acetaminophens, reduce ...

Date: Wed, 04 Jan 1995 16:54:48 -0500 (EST)

From: poster

Subject: opioids and meningitis

#### X is a 13 month (9.8kg) old boy suffering from acute meningitis

(pneumocoque) treated with IV cefotaxime; at day three, I have been called as pediatric pain consultant to assess X; I have discovered an extreme painfull state: one could not handle or touch him without producing screaming. The child was unable to move spontaneously he looked paralysed by pain and hypertonia; he also presented a neurological complication : ptosis at the right side. The pain treatment was IV acetaminophen. The first day I have prescribed IV Nalbuphine (weak opioid u antagonist and agonist) 11mg/24h after a loading dose of 1.4 mg; Pain at rest has been succesfully relieved but not the mobilisation pain; the dose has been increased at 14 mg/day wihout relieving the pain associated with moving; he has moved spontaneously limbs 2 days later; nalbuphine has been stopped 4 days later. Neurological examination and CT scan have been still normal (except ptosis) during this period. No opioid's side effects have been observed.

What do you think of this case ? Have you any experience with opioids and acute meningitis ?

Dr Poster, Pediatric pain unit, Poster Hospital

\_\_\_\_\_

Date: Wed, 04 Jan 1995 17:27:25 -0500 (EST) From: first reply Subject: re: opioids and meningitis

Is there any periosteal involvement? If so an NSAID (ibuprofen or naproxen) may be much more effective than even opioid.

-----

Date: Wed, 04 Jan 1995 19:06:32 -0400 From: second reply Subject: Re: opioids and meningitis

#### Poster writes:

>X is a 13 month (9.8kg) old boy suffering from acute meningitis...

>extreme painfull state: one could not handle or touch him without

>producing screaming....

>The first day I have prescribed IV Nalbuphine ...

>succesfully relieved but not the mobilisation pain;...

>has moved spontaneously limbs 2 days later; nalbuphine has been stopped 4

>days later. Neurological examination and CT scan have been still normal...

I have used IV morphine for similar severe meningitis pain, with success. I wouldn't hesitate to use a pure opioid agonist (in conjunction with acetaminophen, NSAID, and/or tricyclics). However, it sounds like you have the situation under control.

Second Reply, Associate Professor, Dept and University

-----

Date: Thu, 05 Jan 1995 18:58:32 -0800 (PST) From: Third Reply Subject: Re: opioids and meningitis

I wonder if the problem is not due to severe arachnoiditis that is secondary to the inflammation. I would suggest a trial of steroids in this patient, perhaps in combination with a benzodiazepine to reduce the spasm. Narcotics may reduce the pain but I would not like to keep X on them for too long. Third Reply

-----

#### **ABOUT THE AUTHORS**

**Syed Sibte Raza Abidi** is an Associate Professor in the Faculty of Computer Science and is Director of Health Informatics for the Faculty and is the graduate coordinator for the Master of Health Informatics program. He has a wide range of research interests and specializes in knowledge management and health informatics.

**Zhixin Chen** holds degrees in Computer Science from China and a Master of Health Informatics degree from Dalhousie University. He is currently working as a health informatician for the Capital District Health Authority in Halifax, Nova Scotia, Canada.

**G.** Allen Finley holds appointments in the Departments of Anaesthesia & Psychology at Dalhousie University and practices at the IWK Children's Hospital in Halifax, Nova Scotia, Canada. He is one of the founders of the pediatric pain list server.

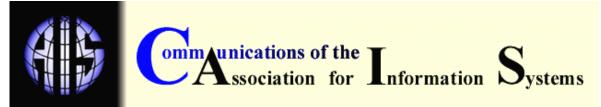
**Qigang Gao** is an Associate Professor in the Faculty of Computer Science and has research interests in Machine intelligence and knowledge-based systems, computer vision and image analysis, data warehousing and data mining, data visualization, and GUI.

**Qiufen Qi** is a medical doctor trained in China and holds a Master of Computer Science degree from Dalhousie University. She is currently in the Interdisciplinary Ph.D. program at Dalhousie University with a research focus in Health Informatics.

**Michael Shepherd** is a Professor in the Faculty of Computer Science and is co-director of the Web Information Filtering Lab (<u>http://www.cs.dal.ca/wifl</u>). He has served on the Board of Governors of Dalhousie University and as the Director of Health Informatics for the Faculty of Computer Science. His research areas are information retrieval, web information systems and health informatics.

Copyright © 2006 by the Association for Information Systems. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints or via email from <u>ais@aisnet.org</u>

.



ISSN: 1529-3181

### EDITOR-IN-CHIEF

Joey F. George Florida State University

<b>AIS SENIOR EDITORI</b>	AL B	OARD	Junionaly		
Jane Webster		Joey F. George		Kalle Lyytinen	
Vice President Publications		Editor, CAIS		Editor, JAIS	
Queen's University		Florida State Univers	ity		ern Reserve University
Edward A. Stohr		Blake lves		Paul Gray	
Editor-at-Large		Editor, Electronic Pul	olications		Editor, CAIS
Stevens Inst. of Technology		University of Houstor	า	Claremont	Graduate University
CAIS ADVISORY BOA	RD				
Gordon Davis	Ken	Kraemer	M. Lynne Mar	kus	Richard Mason
University of Minnesota	Univ.	of Calif. at Irvine Bentley College			Southern Methodist Univ.
Jay Nunamaker	Henk	Sol Ralph Sprague		le	Hugh J. Watson
University of Arizona	Delft	University	University of I	Iawaii	University of Georgia
CAIS SENIOR EDITOR	RS				
Steve Alter	Chris	Holland	Jerry Luftman	l	
U. of San Francisco	Mand	hester Bus. School	Stevens Inst.	of Technolog	у
CAIS EDITORIAL BO					
Erran Carmel	Fred	d Davis Gurpreet Dhill		lon	Evan Duggan
American University	Uof A	rkansas, Fayetteville	Virginia Commonwealth U		U of Alabama
Ali Farhoomand	Jane	Fedorowicz	Robert L. Glass		Sy Goodman
University of Hong Kong	Bentl	ey College	Computing Trends		Ga. Inst. of Technology
Ake Gronlund		Guthrie	Alan Hevner		Juhani livari
University of Umea	Califo	ornia State Univ.	Univ. of South Florida		Univ. of Oulu
	K.D. Joshi Michel Kalika		Jae-Nam Lee		Claudia Loebbecke
Washington St Univ.		Paris Dauphine	Korea University		University of Cologne
Sal March		McCubbrey	Michael Myers		Dan Power
Vanderbilt University		ersity of Denver	University of Auckland		University of No. Iowa
Kelley Rainer		Tallon	Thompson Teo		Craig Tyran
Auburn University		on College	Natl. U. of Singapore		W Washington Univ.
	Upkar Varshney Chelley Vician		Doug Vogel		Rolf Wigand
Georgia State Univ.			City Univ. of Hong Kong		U. Arkansas, Little Rock
Vance Wilson		Wolcott	Ping Zhang		
U. Wisconsin, Milwaukee	U. of	Nebraska-Omaha	Syracuse Uni	versity	
DEPARTMENTS					
Global Diffusion of the Internet.			Information Technology and Systems.		
Editors: Peter Wolcott and Sy Goodman			Editors: Alan Hevner and Sal March		
Papers in French		Information S		Healthcare	
Editor: Michel Kalika			Editor: Vance	Wilson	
ADMINISTRATIVE PE					
		gan Ramsower	Chris Furner		Cheri Paradice
AIS, Executive Director		lisher, CAIS			CAIS Copyeditor
Georgia State University	ty Baylor University		Florida State Univ.		Tallahassee, FL