

## Helping Healthcare Consumers Understand: An “Interpretive Layer” for Finding and Making Sense of Medical Information

Dagobert Soergel<sup>a</sup>, Tony Tse<sup>b</sup>, Laura Slaughter<sup>c</sup>

<sup>a</sup> College of Information Studies, University of Maryland at College Park, MD, USA

<sup>b</sup> Lister Hill National Center for Biomedical Communications, National Library of Medicine, Bethesda, MD, USA

<sup>c</sup> Department of Biomedical Informatics, Columbia University, New York, NY, USA

### Abstract

Healthcare consumers need to find, comprehend, and interpret health information before making informed decisions. Recent work by others and our own work suggest that mismatches in representations of health information used by consumers and professionals occur at different levels of knowledge representation, such as terminology (i.e., form or surface structure and concept or meaning) and semantic relationships. A challenge for consumer health informatics research is to devise a comprehensive strategy to bridge the gap between consumer understanding and biomedical knowledge at all levels. We propose a framework to inform the design of an “interpretive layer” to “mediate” between lay (illness model) and professional (disease model) perspectives, at all levels. In our view, the goal is to assist consumers in identifying terms to describe their needs, finding and **understanding** relevant information, and applying that knowledge for informed healthcare decision making.

### Keywords:

Medical Informatics Applications, Consumer Participation, Comprehension, Information Storage and Retrieval

### Introduction

We propose an “interpretive layer” to assist healthcare consumers (“consumers”) in constructing better mental models of their medical problems, formulating effective queries to express their needs, navigating health information systems, **understanding the documents found**, and acting on them appropriately. Put differently, we need to help consumers find, understand, and use medical information when and where it is needed. This can be achieved by bridging mismatches in knowledge representation between the biomedical professional perspective and the “narrative” lay perspective [1] and by filling in gaps in consumer knowledge. In this paper we present research results that support the construction of an interpretive layer, raise issues, and motivate further research to learn more about the perspectives and needs of consumers.

### Background

Access to medical information facilitates informed decision-making: Informed consumers require less time for physician explanations, are in a better position to select among healthcare op-

tions consistent with personal preferences and values [2], and may be more likely to comply with physicians’ instructions and to adopt a healthy lifestyle. According to [2], “Trying to explain, in a balanced way, the complex issues ... cannot be done quickly. ...educational materials ... may be helpful ... so that clinicians’ limited time can be spent not on basic education, but on tailoring the management strategy to patient’s preferences.” (p. 2357)

Other consumer health informatics applications being developed include decision support (e.g., Should I visit a physician?), patient views of electronic records, and quality control of online information [3]. While evaluations of such systems show general user satisfaction and acceptance, analyses of terms and relationships in consumer-authored questions and messages suggest that mismatches in representation may frequently hinder information retrieval and comprehension.

Recent research has reported mismatches at several levels of knowledge representation: terminology (form and concept, lexical and semantic) (e.g., [4-6]), semantic relationships [7], and explanatory/mental models [1,8]. In summary, the findings show that mismatches at each of these levels have implications for finding, understanding, and applying medical information.

Another major barrier is the different viewpoints (mental models; conceptual frameworks) about health and sickness held by laypersons and specialists. Medical cognition research on patients’ and physicians’ notions of pathology reveals two distinct mental models: an illness model and a disease model, respectively (Figure 1).

We will present results from our research into the medical terms and concepts consumers use and into the relationships between concepts that occur in medical questions posed by consumers, and show how they could be used to construct an interpretive layer.

### Methods

#### Forms and Concepts from Consumer Health Expressions

Fourteen lay volunteers extracted over 55,000 consumer health expressions (words or phrases representing unique medical concepts) from 2,000 anonymized postings at 12 Web-based health discussion forums on various disease-related topics, based on selective sampling. For nearly 25,000 normalized form types accounting for over 35,000 form tokens<sup>1</sup> we looked for mappings

to the Unified Medical Language System® (UMLS®) concepts using semi-automatic methods including MetaMap [9]. We reviewed the extracted surface forms that could not be mapped in the context of the original postings, noting patterns of deviation from professional forms in the expressions used by non-specialists. A recent study has demonstrated promising results using MetaMap for mapping free text, authored by lay persons, to UMLS concepts by limiting the source terminologies [10].

Illness Model (Consumer)	Disease Model (Professional)
Based on everyday experience and social belief systems	Based on biomedical knowledge
A sociocultural construct, possibly with moral implications	Pathophysiological phenomena affecting the body and its structures
Narrative structure; story-like, based on memorable events or “episodes” that disrupt routines of “normal” life	Pragmatic representation of problem, based on underlying biomedical processes; allows for inferences
Little or no biomedical domain knowledge	Deep understanding of biomedical processes and disease characteristics from clinical education and experience
“Common-sense” explanations; superficial temporal or spatial associations (“heuristics of availability”)	Clinical data-driven diagnoses (“heuristics of representativeness”)

Figure 1 - Comparison of models held by laypersons (illness model) and healthcare professionals (disease model) (Adapted from [1])

In the mapping, we distinguished close matches and approximate matches (UMLS concepts broader, narrower, and associated). We consulted the original postings during the mapping process in an attempt to assign concepts that reflect the original intent. Furthermore, we consulted a variety of medical references and a physician (details available in [11]).

### Semantic Relationships from Consumer Questions and Physician-Provided Answers

We coded the semantic relationships in twelve question-answer pairs collected from eight “Ask-a Physician” Web sites (selective sampling), starting with the set of semantic relationships from the UMLS and modifying it as needed. Because several of the questions had multiple parts, 20 distinct questions were identified and characterized. We analyzed the patterns of relationship occurrence within consumer questions, within physician answers, and between questions and answers (details available in [7]).

## Results

At each level analyzed, mismatches between consumer and professionals concepts and forms were observed. Overall, the results were consistent with the illness model [1].

1. Represents 64% of all extracted consumer form tokens

## Forms

Thirteen percent of the forms used by health consumers were observed to be “non-regular” (Figure 2); often the meaning of these forms depends heavily on context (the form *type 1* is sufficient to represent *insulin-dependent diabetes mellitus* in a diabetes forum, but otherwise ambiguous). This is important in machine understanding of consumer questions.

Category	Example
Context Dependency	cut out (1. restriction; 2. excision)
Qualifier	slightly more
Misspelling	gaulbladder (gallbladder)
Shortened Form	alz (Alzheimer Disease)
Defining Phrase	breast removal (mastectomy)
Describing Phrase	ticking bomb (occurs suddenly)
Exemplar	aspirin (analgesic)
Colloquialism	pee (urination)
Neologism	ADD-lets (ADD patients)
Transformation	break (fracture)
Doctor-ese	Postop

Figure 2 - Preliminary categories for characterizing “non-regular” consumer health expressions

The five most frequently observed normalized forms and their UMLS concepts (in parentheses) include: *doctor* (C0031831, Physicians), *pain* (C0030193, Pain), *diagnose* (C0011900, Diagnosis), *test* (C0086143, Tests, Diagnostic), and *symptom* (C0683368, symptom <1>). Consensus forms, expressions commonly used by the observed discourse group to describe medical concepts, were also detected. Examples include *diagnosis*, *fatigue*, *side effect*, and *health*. The largest number of different forms (*highest* expressive variability) was found mapped to the UMLS concept “C0278140, Severe pain”, expressed in many different ways, such as *very painful*, *so much pain*, *terrible pain*. A dictionary of all these variations is essential for machine understanding of consumer texts.

## Concepts

Almost 9,000 form types (36% of ~25,000) accounting for nearly 30,000 form tokens (84% of ~35,000) mapped closely to a concept (~5,300 unique UMLS concepts); however, consumer forms often differed from professional forms in the UMLS. Analyzing closely mapped-to UMLS concept occurrences by semantic type, we found that 34% were related to the UMLS semantic type group, *pathology*. Within that group, 51% belong to constituent semantic types *sign or symptom* and *findings*. Other frequent semantic types include *age*, *family relations*, and entities or processes at an “everyday” level, such as in *anatomy* and *medical procedures*.

Expressions that were not “successfully” mapped to UMLS concepts fell into several categories, including:

- general ambiguity (*fuzzies*),
- lack of specificity (*private area*), and
- non-credible concepts (*mucoïd plaque*).

While some problems might have been resolved by greater access to context, it is not clear how to define others, such as concepts from non-allopathic belief systems (*homeopathy*) and

“pseudoscientific” or marketing concepts (*magic bullet*). Such concepts will be difficult to identify and represent—again a problem for machine understanding of questions.

**Semantic Relationships**

Figure 3 shows a sample individual question-answer analysis.

<p><i>Question text:</i> I have had migraines frequently for the last twenty years and during the last ten years I have had two TIA's (Transient Ischemic Attack). Could you please outline the different symptoms and causes for these disorders. Are these two conditions related? More importantly, could you suggest prevention and treatment options as well?</p>
<p><i>Answer text:</i> The headache that accompanies migraine is due to dilation of superficial vessels. ...if severe, can cause TIAs and this is known as hemiplegic migraine. The flashing lights that often precede migraine attacks are also caused by this vessel narrowing but are not TIAs and not hemiplegic migraine. .... Sometimes vascular headaches that resemble migraine are caused by intracranial arteriovenous malformations ...</p>
<p>Examples of Relationships:  <i>In the Question:</i>                  BRINGS_ABOUT:                      Cause [X (unknown)]                      Effect [migraines]  <i>In the Answer:</i>                  COMPARED_TO                      CompareObjectA [vascular headaches]                      CompareObjectB [migraine]                      ComparativeAB [resembles]                  Implied Between the Question and Answer:                  ISA                      Superordinate [migraine]                      Subordinate [hemiplegic migraine]</p>

Figure 3 - Sample question-answer pair (truncated) with semantic relationship examples [with permission from PCS Ventures]

Table 1 - Types of relationships by frequency

Relationship Type	Frequency (%)		
	Question (N = 97)	Answer (N = 334)	Implied (N = 78)
functionally_related	53	43	35
temporally_related	24	9	4
conceptually_related	10	22	10
additional_relations	7	8	2
topologically_related	5	12	10
isa	1	6	37

Table 1 summarizes the semantic relationship types observed in 12 pairs of question-answers. An important result of this study is a revised inventory of relationship types modified from the UMLS Semantic Network inventory with emphasis on relationship types occurring in health consumer questions and their answers. The relationship types are represented not simply as binary relationships, but as frames with the necessary slots [7].

**Discussion**

Consumer health expressions not in agreement with professional terminology hinder effective information retrieval of documents [5]. Gaps in medical knowledge, faulty mental models, and mismatched concepts and relationships may lead to: inadequate comprehension and misperception of needs; the selection of potentially ineffective query terms; and, even more importantly, to miscommunication and misinterpretation of the information found. Without direct access to professional reasoning and related biomedical knowledge or an easily understandable “interpretation” of these, consumers rely on their own often incomplete mental models.

Differences in perspectives and information introduce an additional complexity: patient needs for information access and learning are often different from those perceived by professionals. Our analysis of the “implicit” needs motivating the consumer-authored questions and mismatches in related professional responses are consistent with the distinction between the lay “illness model” and professional “disease model” [1]. Understanding consumer perspectives and framing information intended for consumers accordingly is crucial in facilitating accessibility. However, after providing an “entry point” for consumers, we believe that an additional “conceptual scaffold” may support learning about biomedical concepts relevant to individual situations.

**An Interpretive Layer for Health Consumer Information**

We need an interpretive layer to assist health consumers in

- understanding their health problems
- finding the proper concepts and terms to search
- making sense of the information they find
- integrating new knowledge into their mental models.

A major underlying function of such an interpretive layer is **educational**: to help consumers acquire medical knowledge and correct misconceptions. The goal is to mediate between lay information seekers’ needs and mental models and the professional perspective at various levels — form, concept, and semantic relationship (Figure 4).

Because changes in health status (e.g., illness events) are typically episodic in acute disease and vary from event to event, consumers do not need comprehensive biomedical knowledge about many topics. Rather, they need “just-in-time” learning, the intent of which is **not** to replace, “water down,” or mask technical information that could benefit patients [12]. Instead, the goal is to assist users by supplementing and providing cues based on their existing knowledge and experience and to lead them step by step to understanding domain-specific knowledge important to their health problem. A system built on understanding how users express, think about, and relate medical concepts will provide flexibility in meeting the varied needs of consumers in different situations.

The interpretive layer must serve the following functions:

- Understanding natural language consumer health questions, including resolution of ambiguities and detection of misconceptions with the goal of extracting query

terms for a document search or producing an entity-relationship representation for an answer search.

- Mapping user search terms to appropriate query terms, expanding query terms following various relationships and suggesting different avenues for search.
- Providing visualizations of concept hierarchies and of real-world relationships (such as hormonal cascades) that the user can assimilate and navigate, both in specifying a search topic and in clarifying concepts found in the search results.
- Providing assistance in understanding documents by explaining concepts and terms.
- Providing direct answers to user questions.

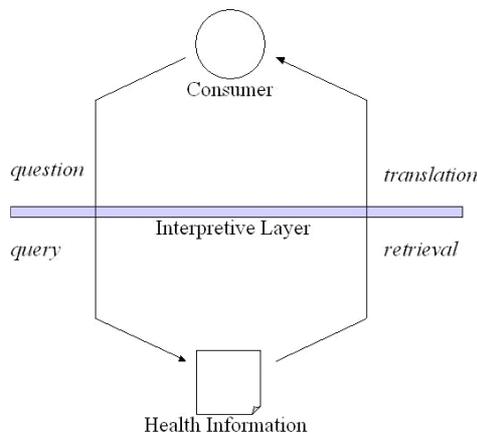


Figure 4 - Schematic showing "interpretive layer" and points of intervention

Such a system needs a knowledge base (KB) consisting of two interrelated and not clearly separable components:

- An educational consumer health ontology and thesaurus that leads from lay health expressions to a consumer-accessible hierarchical structure. Each concept record would have an easy-to-understand definition and relevant context-based usage information to help disambiguate expressions.
- Knowledge bases of (1) highly structured, quality biomedical knowledge and (2) representations of lay explanatory models (assertions and relationships).

Table 2 illustrates functions of the interpretive layer. While consumer health information systems have implemented some of these functions [13,14], we have not found any that implement the comprehensive approach integrating information retrieval and education proposed here.

### Research base and future research

Our research contributes to building such a KB: The large database of mapped consumer health expressions can serve as the starting point for a consumer health ontology and thesaurus, and the patterns and matching rules facilitate computer-assisted processes for expanding the thesaurus; this will require ongoing ef-

fort. The relationship inventory we developed with a focus on types of information important for health consumers provides a structure for the ontology and for a formal encoding of medical knowledge.

Table 2: Hypothetical interpretive layer examples

**Information need.** A consumer wants to know **what treatments are available for shingles**, which she suspects she has.

A **consumer health information system with an interpretive layer** contains an **educational ontology** with a lay definition for *shingles*: "a viral infection by the same virus that causes chickenpox, characterized by symptoms ranging from an itchy rash to severe pain, usually for several days." The **professional KB** contains biomedical facts such as "Herpes zoster (shingles) is caused by human herpesvirus 3, the virus responsible for varicella (chickenpox). Treatments include antiviral drugs, steroids, antidepressants, anticonvulsants, and topical agents." The **KB of lay explanatory models** contains common explanatory beliefs, such as "shingles are caused by bacteria" or "antibiotics are used to treat shingles", enabling the system to explain the error of such beliefs.

**Search: Query formulation.** The consumer may type the query *treatment for shingles* into a search box. If she has an explanatory model that includes "shingles is caused by bacteria," the query might be *antibiotics for shingles*.

**Browse: Finding the category.** If the consumer does not know the term, she could start in a visual display of the body and select a body part and region. Or she could start in a two-window display showing the facets of anatomical classification: *type of tissue* and *body region*; from the *tissue type* hierarchy she would select *skin (surface)* and from the *body location* hierarchy she would select *abdominal region*. The system would then present descriptions and pictures of frequent skin conditions in the selected body region.

**Search: Query expansion.** Based on the domain KB, the search interface would show the technical term (*herpes zoster*), a definition, and a link to additional information. If *antibiotics* were also searched, the causative relationship would be triggered: Pharmaceutical substance <treats> Disease or syndrome. However, the professional KB would identify the error in the assertion "Antibiotics <treats> herpes zoster." The system may then initiate a dialog with the consumer, explaining the nature of the conflict. Additional information based on the "illness model," such as common symptoms in lay and professional terms (e.g., "burning pain and tingling sensations on the skin [neuralgic pain]" and treatments (e.g., "drugs that destroy viruses [antiviral medications]"), might be presented. In addition, alternative queries may also be suggested, with an opportunity to reformulate the query.

**Comprehension of documents found.** On displaying a document, the interpretive layer would parse the forms, identify relevant medical concepts, and show them within a hierarchical context. It would also enhance the document text on-the-fly, providing lay expressions next to technical terms.

More work is needed

- to explore and understand common consumer mental models of medical problems;
- to better understand consumer information needs and information seeking behaviors;
- to explore ways to frame biomedical knowledge optimally for comprehension by laypersons including ways to facilitate comprehension through visualization; and
- to devise ways to evaluate the impact of such a system.

## Conclusion

While considerable progress has been made in making health information available to laypersons, numerous conceptual and technical challenges remain before the promise of widespread accessibility to consumer health information can be fully realized. One problem facing consumers is finding and making sense of basic biomedical knowledge underlying medical information where and when it is needed. The proposed interpretive layer framework aims to serve as a resource for linking consumers with health information systems by addressing ways to bridge mismatches between consumer and professional representations of health-related knowledge. In particular, we propose the development of components such as a thesaurus, including an educational consumer health vocabulary, as proposed by Zielstorff [15], and a knowledgebase of medical knowledge in a form accessible to health consumers. It is hoped that this framework will inspire further work in understanding and developing tools to assist consumers in accessing healthcare information.

## References

- [1] Patel VL, Arocha JF, Kushniruk AW. Patients' and physicians' understanding of health and biomedical concepts: relationship to the design of EMR systems. *J Biomed Inform.* 2002;35(1):8-16.
- [2] Barry MJ. Involving patients in medical decisions. How can physicians do better? *JAMA.* 1999 Dec;282(24):2356-7.
- [3] Eysenbach G. Consumer health informatics. *BMJ.* 2000 Jun 24;320(7251):1713-6.
- [4] Tse T, Soergel D. Exploring medical expressions used by consumers and the media: an emerging view of consumer health vocabularies. *Proc AMIA Symp.* 2003: 564-8.
- [5] Zeng Q, Kogan S, Ash N, Greenes RA, Boxwala AA. Characteristics of consumer terminology for health information retrieval. *Meth Inf Med.* 2002;41(4):289-98.
- [6] Smith CA, Stavri PA, Chapman WW. In their own words? A terminological analysis of e-mail to a cancer information service. *Proc AMIA Symp.* 2002:697-701.
- [7] Slaughter L, Soergel D. How physicians' answers relate to health consumers' questions. *Proc ASIST.* 2003:28-39.
- [8] Forsythe DE. New bottles, old wine: hidden cultural assumptions in a computerized explanation system for migraine sufferers. *Med Anthropol Q.* 1996;10(4):551-74.
- [9] A.R. Aronson, Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc. AMIA Symp.* 2001:17-21.
- [10] Brennan PF, Aronson AR. Towards linking patients and clinical information: detecting UMLS concepts in e-mail. *J Biomed Inform.* 2003 Aug-Oct;36(4-5):334-41.
- [11] Tse T, Soergel D. Procedures for mapping vocabularies from non-professional discourse and a case study from the medical domain. *Proc ASIST.* 2003. p. 174-83.
- [12] Ogden J, Branson R, Bryett A et al. What's in a name? An experimental study of patients' views of the impact and function of a diagnosis. *Fam Pract.* 2003 Jun;20(3):248-53.
- [13] Bouhaddou O, Lambert JG, Miller S. Consumer health informatics: knowledge engineering and evaluation studies of medical HouseCall. *Proc AMIA Symp.* 1998:612-6.
- [14] McRoy SW, Liu-Perez A, Ali SS. Interactive computerized health care education. *JAMIA.* 1998 Jul-Aug;5(4):347-56.
- [15] Zielstorff RD. Controlled vocabularies for consumer health. *Biomed Inform.* 2003 Aug-Oct;36(4-5):326-33.

### Address for correspondence

Dagobert Soergel  
 College of Information Studies, University of Maryland  
 Hornbake Building, South Wing  
 College Park, MD 20742-4345 USA  
[dsoergel@umd.edu](mailto:dsoergel@umd.edu)