

# PatientsLikeMe: Consumer Health Vocabulary as a Folksonomy

Catherine Arnott Smith, PhD<sup>1</sup> and Paul J. Wicks, PhD<sup>2</sup>

<sup>1</sup>University of Wisconsin-Madison; <sup>2</sup>PatientsLikeMe Inc., Cambridge, MA

## Abstract

*PatientsLikeMe is an online social networking community for patients. Subcommunities center on three distinct diagnoses: Amyotrophic Lateral Sclerosis, Multiple Sclerosis and Parkinson's Disease. Community members can describe their symptoms to others in natural language terms, resulting in folksonomic tags available for clinical analysis and for browsing by other users to find "patients like me". Forty-three percent of PatientsLikeMe symptom terms are present as exact (24%) or synonymous (19%) terms in the Unified Medical Language System Metathesaurus (National Library of Medicine; 2007AC). Slightly more than half of the symptom terms either do not match the UMLS, or are unclassifiable. A clinical vocabulary, SNOMED CT, accounts for 93% of the matching terms. Analysis of the failed matches reveals challenges for online patient communication, not only with healthcare professionals, but with other patients. In a Web 2.0 environment with lowered barriers between consumers and professionals, a deficiency in knowledge representation affects not only the professionals, but the consumers as well.*

## Introduction

As consumers have gained increased access online to the literature of healthcare professionals, they have also formed their own powerful communities of expertise, and so the very notion of "expertise" has undergone expansion. Internet-based technologies have great potential not only to empower consumers in general, but allow patients to make a meaningful contribution to the ongoing conversation of healthcare provision.

## Background

The recognition that healthcare terminology is not consumer English is as old as healthcare itself; the consumer health vocabulary problem was not invented, but considerably exacerbated, by the Internet. Medical informatics researchers have explored consumer health vocabulary in various dimensions, documenting the severity of the consumer-professional gap<sup>1,2</sup> communications dysfunctions, and implications for health literacy initiatives<sup>3</sup>.

However, medical terminology, like any other community-specific sublanguage, can be learned by outsiders. Redlich found this in 1945, in the first published study specifically assessing patients' understanding of medical terms<sup>4</sup>. Individual levels of education had little to do with understanding of the correct definition; instead, the time a patient had spent on the ward was the most important factor. Expertise simply at being a patient is an expertise in its own right. This speaks to the importance of community to the construction of experts.

*Community* is a distinguishing characteristic of Web 2.0. Sites can promote collaboration and distribution of information by peer members. When boundaries separating individuals and communities disappear through the use of social networking, traditional authority roles disappear also.<sup>5</sup> Community-building technologies empower not only because they traverse boundaries create offline, but because they erase them completely.

Web 2.0 privileges augmented content over semantic architecture; for example, a user-generated taxonomy called a *folksonomy* can be established through the construction and collaboration of user-generated index terms, or tags (such as those at Amazon.com, Flickr, Technorati, and Craigslist). The word *folksonomy*, coined in ironic opposition to *taxonomy*, was first used in 2006. Folksonomy facilitates networking of related concepts and related interests, thus creating related people; in Web 2.0, "seeing what other users are thinking about is as much a part of the site as finding what you need."<sup>6</sup>

How can users directly contribute to vocabularies? The literature of biomedical informatics is largely silent on the question. Users and consumers exist primarily as sources of feedback. Pain language, one particular aspect of patient communication, has been researched primarily to develop assessment instruments. Typically, patients assess, rate and indicate agreement with term lists, but do not themselves generate new terms (for examples, see <sup>7,8</sup>). Consumer health vocabulary researchers do rely on analysis of consumer-generated terms, but this represents anonymous consumers in absentia (see <sup>9</sup> for a review).

When patients, are consulted about their term preferences in an information systems context, their

input are intended to improve existing term lists and interfaces constructed by experts.<sup>10</sup> One notable exception is documented in<sup>11</sup>, where researchers asked other researchers to describe themselves for representation in a database; these self-generated descriptions were made available for keyword searching.

We propose that consumer and patient folksonomies are a rich source of vocabulary for enhancing communication in healthcare. This work has particular implications for the personal health record, which can be expected to require a mix of personal and professional vocabulary.

PatientsLikeMe ([www.patientslikeme.com](http://www.patientslikeme.com)) is an online social networking community allowing members to track their progress with clinical scales, share information, and learn more about their condition. In December 2007 PatientsLikeMe consisted of three subcommunities: Amyotrophic Lateral Sclerosis (ALS; founded November 2005), Multiple Sclerosis (MS), and Parkinson's Disease (PD, both founded April 2007).

The 4,914 unique users are patients (68%), caregivers (17%), guests (10%), researchers (4%) and doctors (1%). This community is international; members are nationals of more than 13 countries, with a slight majority in the United States (51%). Members who report time since diagnosis show a mean of 32 months (ALS), 66 months (MS), and 53 months (PD). Of the 1810 total members reporting their gender, 44% are male and 56% female; the mean age across all 3 communities is 50 years old.

Each community member is asked to track ten "core symptoms" of their condition. The core list was first generated for the ALS community with input from healthcare practitioners and the literature, for example<sup>12</sup>, and then modified for use in the MS and PD communities. Members can also report, in natural language, any additional symptoms they are experiencing. The result is a semi-structured alphabetical list which patients can use as an assist for future symptom reporting. It also permits comparison with symptoms reported by other people. The terms become "live" immediately, but are periodically reviewed for normalization as necessary. The contribution of patients to the naming of symptoms has already had a clinical effect. PatientsLikeMe researchers found a statistically significant association between excessive yawning, reported as a symptom within the ALS community, and bulbar onset of ALS disease; excessive yawning was twice as common in bulbar-onset ALS patients as those with limb-onset ALS. After this association was

confirmed, the term "excessive yawning" was relocated to the core symptom list for ALS.<sup>13</sup>

An old research question with new implications for Web developers is this: What language do patients use to describe their conditions? And what are the implications of patient- and consumer-contributed terms for patient- and consumer-oriented information systems?

## Methods

As of September 2007, 376 symptom terms had been contributed by PatientsLikeMe community members. Two coders working independently analyzed these raw, un-normalized terms for consonance with the Unified Medical Language System (2007 AC) in December, 2007 and achieved 100% inter-rater agreement.

## Results

Forty-three percent of the patient-submitted terms from PatientsLikeMe communities are present either as exact matches to the Unified Medical Language System Metathesaurus (2007AC) [24%], or as synonym matches [19%]. Most exact matches were contributed by SNOMED CT (93%), followed by Read Codes (88%) and MedDRA (86%).

Exact Match	Synonym	Not Found	Other*	TOTAL
90	71	209	6	376
24%	19%	56%	3%	

\*=Duplicates (1; <1%); Impossible to determine patient's meaning (2; 1%); False match (3; 1%)

**Table 1.** PatientsLikeMe symptom terms (3 communities): Agreement with the UMLS Metathesaurus

Six UMLS semantic types represented 92% of the terms: Sign or Symptom (38%); Disease or Syndrome (25%); Finding (24%); Pathologic Function (3%); Mental/Behavioral Dysfunction (2%) and Body Part, Organ or Organ Component (2%). Eight other types accounted for 1% of the terms each.

Two hundred nine terms submitted by PatientsLikeMe community members did not match the UMLS Metathesaurus (2007AC). The reason these terms failed are presented in Table 2 below.

Category	Number	Example
Fragment or phrase	90	<i>Positive test for borrelia burgdorferi at Bowen Research</i>
Other	70	<i>Banding; L'Hermitte Sign</i>
Two more concepts in one expression	55	<i>Choking/coughing while eating/drinking</i>
Misspelling	11	<i>Ischemic colitus</i>
Temporal characteristics	9	<i>Sneezing during breakfast</i>
Vague	6	<i>Cramps; Other dropped foot severe</i>
Slang	3	<i>Word fishing</i>
False Matches	3	<i>Ear burning</i>

**Table 2.** Failure analysis of nonmatching patient-submitted symptom terms

### Discussion

The PatientsLikeMe data is interesting regarding on the nature of a “consumer” health vocabulary. It is noteworthy that no nursing vocabularies are found in the top ten contributing matches to patient language in this study. This contrasts with the findings of those of Brennan and Aronson, who studied coverage of consumer vocabulary in email messages by 6 nursing terminologies, concluding that “these vocabularies address a particular part of the patient experience not addressed in other health care vocabularies.” They found that nursing vocabularies were able to provide “an accurate, if incomplete, representation of the terms patients use in their electronic mail messages.”<sup>14</sup>

It is in the ontology of semantic types that we see the considerable challenge, not only of communication between healthcare professionals and patients, but of the ability of consumer-generated content to completely represent a clinical situation. Only 38% of the patient-submitted symptom terms are actually considered “Signs or Symptoms” by the UMLS. What other things are considered “symptoms” by community members? Any things that they see as affecting their health and well-being in any way – diseases (25%), physical and mental processes, functions and dysfunctions such as acquired abnormalities (6%) and injuries (1%), and even a bacterium (1%).

Some of these patients named *Type I diabetes*, *kidney stones*, and *carpal tunnel* as “symptoms”. They may be honestly expressing their belief that these conditions are effects and not causes, and reflect an underlying disease state for which they do not have, or do not know, a name. Other terms reflect clear

confusion over just what a symptom is. *Borrelia burgdorferi* is a good example. This is the bacterium that causes Lyme Disease. The patient who calls it a symptom may have meant to convey the process of differential diagnosis: Symptoms of Lyme Disease can mimic those of ALS, and thus this is one of the diagnoses that must be excluded in investigation of possible ALS.<sup>15</sup>

Some communication problems are caused by sparse “tagging” by community members. These “symptoms” include the human body and its aspects in its normal state (body parts, 2%; bodily substances, 1%, and behaviors, 1%). The gap here is not in understanding, but in the context. Patients listing body parts as symptoms must be expressing the location of the symptoms they feel—the *bladder*, the *left hand*, the *right hand*. But what are these symptoms? What do we make of *eye floaters* and *mucus* (More? Less? Different kind?) The clinician does not know; more importantly for a social networking site, expressly geared for pairing of patients experiencing similar life courses, other patients do not know.

The failure analysis of terms displayed in Table 2 completes the picture of communication challenges online. PatientsLikeMe terms not found in the UMLS Metathesaurus – either as a concept name, or a synonym – fell into seven categories.

*Fragments or phrases* (43%) represent in most cases the community member’s attempt to use the symptom list, not as the terminological assist intended by the developers, but as a means of dialogue with other people. Some phrases are not symptoms, but instead a brief medical history: *Had heart attack 16 mayos put stent in*. Some fragments are parts of sentences

describing a specific clinical event: *Positive test for borrelia burgdorferi at Bowen Research*. This kind of consumer phrasing has been called “definitions and descriptions”<sup>16</sup>.

Some terms have nested *two or more clinical concepts* (26%), whether symptoms or not, in one expression. These terms and phrases will present problems for community members just as they do for the postcoordinated UMLS Metathesaurus. Weakness in left arm and shoulder may describe 50% of a patient’s problem, if she has a weak left arm, but not completely, if her shoulder is unaffected. In order for this community member to find a patient “like me”, she must either overstate her own symptoms, by including her left shoulder, or she must create a new symptom—*weakness in left arm*—where none existed before. A postcoordinated list of symptoms depends, for its ultimate utility, on users who know where to look on that list. This condition holds whether the list is highly controlled by the centralized maintainers of the information system – for example, by the National Library of Medicine via MeSH—or by the distributed peers contributing idiosyncratic tags, as in folksonomy builders and PatientsLikeMe.

*Mis-spellings* (5%) can be ameliorated by the human editor who edits the terms. Editing, however, does not solve the problem of excessively vague expressions (3%), such as *cramps* (muscular? Menstrual?) Again, symptoms too vague for the UMLS to match can be assumed to be difficult for community members to match as well. How much of a symptom’s conceptualization and expression must be shared for a patient to find a patient like her?

Those in the *temporal* category (4%) encode times of day. The patient who contributed *Sneezing during breakfast* as a symptom clearly believes the context of the sneezing as important as the sneezing itself. Attention to temporal characteristics may be as important for patients as it can be for healthcare providers.

*Slang* (1.4%) occurred very rarely in this data, which may reflect community members’ understanding that these symptoms have clinical meaning and are also being mined for clinical purposes.

Finally, there are 70 symptom terms which can be not be classified anywhere else (*Other*, 33%). Most terms in this category are not found in the UMLS because they express either a problem or a body part in more granular terms than the UMLS “knows”: *atrophy of thigh muscles*, instead of *Muscle Atrophy* (UMLS CUI C0026846). Nine of these patient-contributed terms are synonyms for a formal clinical concept not

currently recognized by the >120 source vocabularies in the UMLS. *Bowel urgency, Cognitive confusion, Excess mucus, Foot drag, Gag response, Heart racing, Poor temperature regulation, Ringing in ears, Temperature dysregulation, Vitreal haemorrhage, L’Hermitte sign* and *Uhtoff’s Phenomenon*—are terms used in the medical literature and documented in Medline, but terms that the UMLS presently does not know about. These are all good examples of how patients can meaningfully contribute to the extended conversation that is healthcare.

## Conclusion

A website sponsored and tailored for a specific population is, among other things, a community information system; it is dependent on “a tight interplay between the organization of knowledge and communicative processes within communities of practice.”<sup>17</sup> Vocabulary is key to the communicative process, yet website developers, designers and maintainers of healthcare information systems do not generally consult with users about the vocabulary in which that information is provided. Designers tend to assume that their own preferences and skills are representative of the user.

Dutch researchers Oudshoorn and Somers<sup>18</sup> looked at three Dutch patient organizations and their websites. A dichotomy was found between implicit and explicit techniques used for knowledge representation by these three organizations. Implicit form of modeling rests on “think[ing] from the perspective of the target group”. Only a website devoted to young people with cancer relied on more explicit methods, basing representation in personal experience but also on “extensive interactions with young people”.

The implicit-explicit distinction is a knowledge representation challenge, because making internalized understanding externally visible for the use of others is a difficult task. As any reader who has worked with healthcare data standards understands, symptoms and other expressions of the lived patient experience are both “unconscious and procedural ... hard to formalize and communicate to others.”<sup>19</sup>

Patients are the target audiences of patient-oriented websites. In a Web 2.0 environment, patients also contribute to the building and maintenance of these websites. Participation is reinforced by the strong value of empathy and identity politics in online community. Thus it follows logically that implicit knowledge representations generated by outsiders, what Oudshoorn & Somers call the ‘I-Methodology’, “cannot do justice to the patient’s experience because

it excludes the perspectives and needs of people with differing demographic characteristics from the design.”<sup>18</sup> Knowledge construction through social networking can elicit new healthcare concepts for healthcare vocabularies, coding sets, and classifications. The challenge for PatientsLikeMe and other online patient communities is to avoid recreating an I-Methodology through a perpetuation of selfish tagging. The results of this study reveal a range of challenges for online patient communication, not only for healthcare professionals, but for other patients. Vocabulary developers in the Web 2.0 era must understand the tension between unfettered, free expression and rigidly controlled terminologies in order to harness the real power of the folksonomy for enhanced communication and information retrieval. In a Web 2.0 environment with lowered barriers between consumers and professionals, a deficiency in knowledge representation affects not only the professionals, but the consumers as well.

**Acknowledgments and Disclosures.** The authors express their appreciation to the staff at PatientsLikeMe for data extraction and to the users of the site for sharing their data. The second author is an employee of PatientsLikeMe.com and holds stock options in the company.

### References

- Zeng QT, Tse T. Exploring and developing consumer health vocabularies. *J Am Med Inform Assoc* 2006; 13(1): 24-9.
- Smith CA, Stavri PZ, Chapman WW. In their own words? A terminological analysis of e-mail to a cancer information service. *Proc AMIA Symp* 2001; 697-701.
- McCray AT. Promoting health literacy. *J Am Med Inform Assoc* 2005; 12(2):152-63.
- Redlich FC. Patient's language: Investigation into use of medical terms. *Yale J Biol & Med* 1945; 17:427-453
- Matusiak KK. Towards user-centered indexing in digital image collections. *OCLC Syst & Serv Rsch* 2006; 22(4): 283-298
- Dye J. Folksonomy: A game of high-tech (and high-stakes) tag. *EContent* 2006; 29(3): 38-43
- De Conno F, Ripamonti C, Caraceni A, Saita L. Palliative care at the National Cancer Institute of Milan. *Support Care Cancer*. 2001 May; 9(3):161.
- Ohnhaus EE, Adler R. Methodological problems in the measurement of pain: a comparison between the verbal rating scale and the visual analogue scale. *Pain* 1975; 1(4): 379.
- Zeng QT, Tse T. Exploring and developing consumer health vocabularies. *J Am Med Inform Assoc* 2006; 13(1): 24-9
- Slaughter L, Ruland C, Rotegård AK. Mapping cancer patients' symptoms to UMLS concepts. *Proc AMIA Symp* 2005;699-703
- Friedman PW, Winnick BL, Friedman CP, Mickelson PC. Development of a MeSH-based index of faculty research interests. *Proc AMIA Symp*. 2000;265-9
- Forshew DA, Bromberg MB. A survey of clinicians' practice in the symptomatic treatment of ALS. *Amyotroph Lateral Scler Other Motor Neuron Disord*. 2003; 4(4):258-63
- Wicks P. Excessive yawning is common in the bulbar-onset form of ALS [letter]. *Acta Psc Scandinavica* 2007; 116 (1), 76
- Brennan PF, Aronson AR. Towards linking patients and clinical information: detecting UMLS concepts in e-mail. *J Biomed Inf* 2003; 36:334-341
- National Institute of Neurological Disorders and Stroke. 2007, Dec. 11. Amyotrophic Lateral Sclerosis Fact Sheet. Available online: [http://www.ninds.nih.gov/disorders/amyotrophiclateralsclerosis/detail\\_amyotrophiclateralsclerosis.htm](http://www.ninds.nih.gov/disorders/amyotrophiclateralsclerosis/detail_amyotrophiclateralsclerosis.htm). Date accessed: December 30, 2007.
- Tse T, Soergel D. Exploring medical expressions used by consumers and the media: An emerging view of consumer health vocabularies. *Proc AMIA Symp* 2003: 674-678.
- Wenger E. *Communities of practice: Learning, meaning and identity*. Cambridge: Cambridge University Press, 1998.
- Oudshoorn N, Somers A. Constructing the digital patient: Patient organizations and the development of health websites. *Inf, Comm & Soc* 2006; 9(5): 659.
- Spaniol M, Klamma R, Springer L, Jarke M. Aphasic communities of learning on the Web. *Int J Dist Ed Tech* 2006; 4(1):31-45.