



Towards new information resources for public health—From WORDNET to MEDICALWORDNET

Christiane Fellbaum^{a,b,*}, Udo Hahn^c, Barry Smith^{d,e}

^a Department of Psychology, Princeton University Green Hall, Washington Rd., Princeton, NJ 08544, United States

^b Berlin-Brandenburg Academy of Science, Berlin, Germany

^c Jena University Language and Information Engineering (JULIE) Lab, Friedrich-Schiller-Universität Jena, Jena, Germany

^d IFOMIS, Universität des Saarlandes, Saarbrücken, Germany

^e Department of Philosophy, State University of New York at Buffalo, USA

Received 21 July 2005

Abstract

In the last two decades, WORDNET has evolved as the most comprehensive computational lexicon of general English. In this article, we discuss its potential for supporting the creation of an entirely new kind of information resource for public health, viz. MEDICALWORDNET. This resource is not to be conceived merely as a lexical extension of the original WORDNET to medical terminology; indeed, there is already a considerable degree of overlap between WORDNET and the vocabulary of medicine. Instead, we propose a new type of repository, consisting of three large collections of (1) medically relevant word forms, structured along the lines of the existing Princeton WORDNET; (2) medically validated propositions, referred to here as *medical facts*, which will constitute what we shall call MEDICALFACTNET; and (3) propositions reflecting laypersons' medical beliefs, which will constitute what we shall call the MEDICALBELIEFNET. We introduce a methodology for setting up the MEDICALWORDNET. We then turn to the discussion of research challenges that have to be met to build this new type of information resource. We build a database of sentences relevant to the medical domain. The sentences are generated from WordNet via its relations as well as from medical statements broken down into elementary propositions. Two subcorpora of sentences are distinguished, MedicalBeliefNet and MedicalFactNet. The former is rated for assent by laypersons; the latter for correctness by medical experts. The sentence corpora will be valuable for a variety of applications in information retrieval as well as in research in linguistics and psychology with respect to the study of expert and non-expert beliefs and their linguistic expressions. Our work has to meet several considerable challenges. These include accounting for the distinction between medical experts and laypersons, the social issues of expert-layperson communication in different media, the linguistic aspects of encoding medical knowledge, and the reliability, volume, and emergence of medical knowledge. The work described here has been tested in a small pilot experiment [39] and awaits large-scale implementation.

© 2005 Elsevier Inc. All rights reserved.

Keywords: WordNet; MedicalWordNet; MedicalFactNet; Medical terminology; Lexical database; Expert terminology; Layperson's terminology

1. Introduction

In the last two decades, WORDNET has evolved as the most comprehensive computational lexicon of general English. In this article, we discuss its potential for creating an entirely new kind of information resource for public health,

called MEDICALWORDNET. This resource is not meant to be merely a lexical extension of the original WORDNET to medical terminology (indeed, WORDNET already contains a large fraction of the medical terminology in everyday use). Rather, we propose the transformation of WORDNET into a large collection of medically validated propositions (which will constitute the MEDICALFACT-NET) and medical beliefs (which will constitute the MEDICALBELIEFNET). We introduce a methodology for setting up the entire MEDICALWORDNET structure. We then turn to an in-depth dis-

* Corresponding author. Fax: +609 2582682.

E-mail addresses: fellbaum@princeton.edu (C. Fellbaum), hahn@coling.uni-freiburg.de (U. Hahn), phismith@buffalo.edu (B. Smith).

cussion of research challenges that have to be met to build this new type of information resource.

2. WORDNET

WORDNET [29,17] is a large electronic lexical database of English. WORDNET was originally conceived as a full-scale model of human semantic organization, where words and their meanings are related to one another via semantic and lexical relations. One kind of evidence, on the basis of which this model was constructed, comes from word association norms [36]. Given a lexical stimulus (a noun, verb, or adjective), clear response patterns can be discerned. Stimuli and responses very frequently stand in specific semantic relations such as synonymy, antonymy, hypernymy, and meronymy. For example, “car” might elicit its synonym “vehicle” or its hyponyms (subordinate terms, expressing more specific meanings) like “truck” or “convertible.” The stimuli “truck” and “convertible,” in turn, might elicit their common hypernym (superordinate term with a more general meaning) “car.” “Car” might evoke its meronyms (terms expressing parts), such as “tire,” “brake,” and “windshield.” Conversely, the response to “steering wheel” might be its holonym (term expressing a corresponding whole) “car.” The relation of antonymy, or semantic contrast, is particularly salient from a cognitive point of view. People’s response to “cold” is almost invariably “hot,” and vice versa [13]. Such association patterns are believed to shed light on the way the mental lexicon is organized.

But besides the apparent psychological reality of the relations outlined above, there is also a textual correlate. Words that are synonyms, antonyms, hyponyms/hypernyms, and meronyms/holonyms of one another are more likely to co-occur in a single context than are unrelated words. This makes sense, given that a coherent text or discourse necessarily involves semantically related words. The co-occurrence patterns also are believed to reinforce the association among semantically related words. For the present purpose, the UML’s co-occurrence table could be inspected for data that would strengthen the development of MWN [34].

WORDNET shows that the bulk of the English lexicon can be organized by means of these relations. Further WORDNETS have since been successfully built in over 30 languages—often typologically and genetically unrelated—demonstrating the crosslinguistic validity of its design (for instance, for 13 European languages—Dutch, Italian, Spanish, Czech, Estonian, German, and French plus six Balkan languages—other than English, cf. [48,45]).

2.1. A brief survey of WORDNET

The building block of WORDNET is a synonym set, or *synset*, consisting of all the words that can be substituted for one another in given types of sentential contexts without change of truth value in the sentences involved. Not all the members of a given synset will be interchangeable

in all contexts. Examples are “bank, bench” or “fall, drop.” The current version of WORDNET contains over 117,000 synsets.

Synsets are connected to one another by means of bidirectional labeled arcs representing semantic relations in such a way as to constitute a dense semantic network. For example, the synsets “car, automobile” and “vehicle” are connected to each other through the hypernymy/hyponymy relation; “wheel” and “car” are linked via the meronymy/holonymy relation. In principle, WORDNET’s structure should ensure that all hyponyms (types) of “car” are also hyponyms of “vehicle” (and that all hyponyms of “car” refer to objects having referents of “wheel” as parts).

Adjectives encode properties, and their semantics seem to be best captured by a contrast, or antonymy, relation, illustrated by the synset pairs “ill, sick,” “well, healthy.” Concepts expressed by verbs have a particularly rich inventory of relations. Most verbs are “manner” elaborations of other verbs. Thus, “running” is a manner or way of “moving” and “babbling” is a manner or way of “talking.” Causation links pairs like “raise” and “rise,” because if somebody *raises* something, it will also *rise*. A range of entailment relations link pairs like “default” and “owe” and “arrive” and “leave.” If someone *defaults* on a loan, he necessarily thereby *owes*; if I *arrive* somewhere, it follows that I have *left from* somewhere.

While word meanings are represented largely in terms of a word’s relations to other words and synsets, every synset additionally contains a short definition (or *gloss*) and, in most cases, at least one sentence illustrating the usage of the synset members.

The representation of lexical meaning in WORDNET turned out to be very useful for a number of natural language processing (NLP) applications. Members of a synset can be substituted for one another as key words in information retrieval; and the synonyms thereby identified used to filter out thematically coherent documents [47]. Alternatively, many NLP applications depend crucially on word sense disambiguation. Polysemy, the existence of one-to-many mappings from word forms to meanings, creates difficulties for automatic systems, though it does not pose a problem for human speakers. By exploiting the fact that each form-meaning pair in WORDNET occupies a unique position in the network, polysemous word forms like “bed” can be successfully disambiguated. One sense has “piece of furniture” as its hypernym; another is a hyponym of “plot, piece of land;” another is a “kind of stratum,” etc. Additionally, each sense has its own distinct set of hyponyms, meronyms, and so forth [26].

2.2. WORDNET and biomedicine

WORDNET was designed as an all-purpose rather than as a domain-specific lexical resource. Over the two decades of its construction, many technical terms from a wide range of domains have been included within it. Unfortunately, this was done haphazardly rather than in a planned fashion and it was not directed towards any particular type of

application. Moreover, WORDNET's builders are not domain experts, so that entries with technical meanings are not always reliable. One real problem is that many of WORDNET's definitions are unintelligible to laypersons. Some medical terms are even obsolete, such as "unction" and "ichor."

Another consequence of the fact that the medical entries in WORDNET were made by non-experts is that the associated hierarchies tend to be shallow, lacking intermediate nodes expressing meanings intelligible to, and salient for, medical experts. Still such missing nodes are often important for understanding nodes that are included.

A fairly systematic by-product of WORDNET's synset structure is the shared membership of expert and non-expert, or folk, terms in the same synset. Examples are "upper jaw, maxilla" and "hay fever, pollinosis." The co-existence of medical and folk terms within the same synsets presents a potentially powerful advantage for information retrieval, allowing for translation between expert and non-expert language. But this is true only in those cases where all synset members do indeed share a common meaning. In the medical domain, this frequently turns out not to be the case. Serious miscommunication may result in those settings where one party uses a term in its technical, medical sense and the other party uses the same term under the mistaken assumption that the same disorder or symptom is being referred to (cf. also our discussion in Section 5.6 below). Non-expert use of medical terminology can be characterized by fuzziness. In such cases, the meanings of given words as used by experts and non-experts diverge. Thus, people frequently confuse symptoms of a common cold with those of a flu. As a result, a patient with a cold may seek, receive, and follow on-line advice for treatment of the flu or vice versa.

WORDNET presents information generically and categorically and does not account for modality. This means that the relations among its entities are represented as being necessarily true and there is no room for probability, optionality, or conditionality. As a result, potentially important information is not provided. For example, "blister" is given as a kind of body part (skipping several intermediate levels), which to a naive user suggests that every body has blisters. The fact that blisters are a transient phenomenon associated with an injury is also not represented. Another example is "sprain," defined as "a painful injury to a joint caused by a sudden wrenching of its ligament." This gloss does not admit the possibility of torn ligaments, which is in fact a common by-product—or cause—of a sprain.

In a series of papers, Olivier Bodenreider and Anita Burgun have performed rigorous experiments on the connection of WORDNET to the medical and biological domain. To gain an understanding of how anatomical concepts are defined by a specialized medical dictionary (Dorland's [15]) and WORDNET as a general, i.e., layperson's terminological system, Bodenreider and Burgun [4] selected after some preprocessing 420 definitions using WORDNET's glosses as definitions; 134 anatomical terms were defined in

WORDNET, 213 in Dorland's, and 117 occurred in both sources. They found (by manual analysis) that genus-differentia definitions prevail in both general and specialized resources and that hierarchical relations are the principal type of relation found between the definiendum and the noun phrase head of the definiens. Although comparing domain-specific technical definitions with the medical definitions produced for use by laypersons, Bodenreider and Burgun do not raise the issue of the comprehensibility of either one to a non-expert audience.

Bodenreider and Burgun find that anatomical definitions are characteristically of the form: superordinate plus distinguishing feature (the latter expressed through some form of adjectival modification or relative clause). This way of defining words is in fact the canonical one (for nouns, and, to some degree, for verbs as well) and lexicographers follow it as much as possible when writing definitions. MEDICALWORDNET will observe this standard consistently in its augmentation and standardization of WORDNET's definitions, drawing on the results of the studies of best practice in the formulation of definitions in biomedical terminologies and ontologies [41,6,40].

Burgun and Bodenreider [9] also deal with the issue of terminological overlap between a layperson's terminological resource (WORDNET) and the specialized vocabularies of the Metathesaurus of the *Unified Medical Language System* (UMLS) [37]. By focusing on two semantic classes, viz. "Animal," a general class, and "Health Disorder," typical of the medical domain, they identify, for the latter, 2% of the domain-specific concepts from the UMLS in WORDNET, while 83% of the domain-specific concepts from WORDNET are found in the UMLS (for the "Animal" class the coverage data is 19 and 51%, respectively). Terms from WORDNET absent in UMLS are usually found to be lay terminology, an interesting finding as far as our approach is concerned (an example is WORDNET's "kissing disease" which appears in the UMLS, but only under the heading "infectious mononucleosis").

From a more focused perspective, Bodenreider et al. [5] assessed the coverage of WORDNET for terminology from molecular biology and genetic diseases. They extracted four major categories (phenotype, molecular function, biological process, and cellular component) from LOCUSLINK¹ and mapped it to synsets from the noun hierarchy of WORDNET. Furthermore, all terms from the *Gene Ontology* database² were also mapped to WORDNET to evaluate the latter's coverage of the domain of genes and gene products. Predictably, the coverage for highly specialized terms turned out to be low, ranging from 0% (for gene products) to 2.8% (for cellular components). Removing specialization markers (such as the use of hyphens, numbers, and capitals) from the terms and using synonyms significantly increased the rate of mappings of genetic disease names

¹ www.ncbi.nlm.gov/LocusLink, now superseded by ENTREZ GENE at www.ncbi.nlm.gov/entrez, last visited on June 14, 2005.

² www.geneontology.org/, last visited on June 22, 2005.

to WORDNET entries, boosting the overall mapping rate to a range from 27.4 to 31.4%. Still, WORDNET contains many of the most common terms for single gene disorders (e.g., “Huntington’s disease”), as well as many of the high-level terms from the *Gene Ontology*. Therefore, Burgun and Bodenreider conclude that WORDNET is likely to be a useful source of lay knowledge in the framework of a consumer health information system on genetic diseases.

To provide a preliminary estimate of the extent of WORDNET’s medical coverage, we conducted our own experiments and derived a test lexicon composed of 2838 single-word medical terms from an existing digitalized lexical resource for medical language processing (the LINKBASE system of the Belgian NLP company L&C),³ which was constructed independently of WORDNET by medical professionals. We transformed LINKBASE into an alphabetically ordered term list and eliminated automatically all acronyms, multi-word terms, proprietary terms, terms containing numbers, and (selecting a more or less arbitrary threshold) terms greater than ten characters in length. Of the 2838 terms then remaining, only eleven were not present in any form in WORDNET 2.0. Almost all missing terms were compounds such as “bedwetting” and “breastfed,” i.e., words composed from constituent parts which are present in WORDNET with the relevant meanings. Such compounds ought, however, to be included in a lexicon wherever their meanings are not the sums of the meanings of their parts.

Turning to usage scenarios for WORDNET in the medical environment, Xiao and Rösner [50] show how WORDNET can be used as a tool for simplifying information extraction from MEDLINE. Parsing tools are used to extract verbs from the corpus of MEDLINE abstracts, and it is then shown that very many (both low- and high-frequency) verbs are grouped together into WORDNET synsets in such a way that, within this specific discourse domain, there is only one semantic relation linking all the verbs in each of the relevant synsets. In this way it is possible to simplify the process of information abstraction by reducing the number of relations that need to be taken into account in the analysis of texts.

WORDNET’s design allows users with specific technical applications to augment the database, primarily by adding new terms as leaves to the existing branches of its taxonomic and part-whole hierarchies. Such enriched WORDNETS retain all of the original information, and the added words are semantically specified in terms of WORDNET’s original relations. Turcato et al. [46] and Buitelaar and Sacaleanu [8] describe an attempt to extend the German WORDNET with synsets pertaining to the medical domain using automatic methods, in particular the detection of semantic similarity from co-occurrence patterns in a domain-specific corpus. The results, while good, are hampered by problems of lexical polysemy and by the notorious German tendency

for compound formation, which leads to potentially open-ended lexicon growth and thus poses problems for automatic word sense recognition and discrimination. One clear conclusion from these studies is that fully automated lexical acquisition provides inadequate results, and that much of the acquisition and curation work must be performed manually. Our proposal below reflects this conclusion.

3. Methodological considerations: from words to sentences

The previous section provided some evidence that WORDNET carries great potential for linguistic applications in the biomedical domain. However, communication exclusively on the lexical level—the level of single words—is necessarily limited to the mere identification of entities, viz. objects, properties, and events. For more effective language processing words have to be embedded in phrases and sentences. In fact, linguists are well aware that the meanings of words are to a great extent dependent on the contexts in which they are used. It is for this reason that definitional sentences were added to WORDNET’s synsets in the course of its development. However, much of the misleading medical information in WORDNET is carried precisely by these sentences, which were, of course, not intended to convey medical information to health care consumers.

We therefore propose to build a database of sentences containing medical terminology drawn from WORDNET and other lexical resources. The database will consist of two validated subcorpora of sentences, MEDICALBELIEFNET and MEDICALFACTNET, both providing meaningful contexts for relevant medical terms. MEDICALBELIEFNET will consist of sentences that receive high marks for assent by laypersons and is thus designed to constitute a representative fraction of the beliefs about medical phenomena (both true and false) distributed through the general population of English speakers. MEDICALFACTNET will consist of sentences that receive high marks for correctness on being assessed by medical experts; it is thus designed to constitute a representative fraction of the beliefs about medical phenomena which are intelligible to non-expert English-speakers. For each of the major selection steps—deriving fact statement via coders, raters’ decisions on their status as medical facts or beliefs by raters—we will quantify the degree of intra- and inter-coder as well as intra- and inter-rater consistency.

Our corpora will be restricted to grammatically complete, syntactically simple sentences of the English language which have been rated as understandable by non-expert human subjects in controlled questionnaire-based experiments. They will be restricted in addition to sentences that are self-contained, in the sense that they contain no anaphora or indexical expressions or any linguistic elements that need to be interpreted with respect to other sentences. The sentences will be generated from WORDNET’s synsets and the relations among them, as well as from publicly available medical information sources (which will also supply new words and word senses that are missing in WORDNET).

³ <http://www.landcglobal.com/index.php>.

Compiling MEDICALFACTNET and MEDICALBELIEFNET in tandem will allow us not only to build new sorts of applications for information retrieval in the domain of consumer health, but also to pursue new avenues of research in linguistics and psychology, for example in exploring individual and group differences in medical knowledge and vocabulary, in understanding non-expert medical reasoning and decision-making, and in making a systematic assessment of the disparity between lay and expert beliefs and vocabulary as concerns medical phenomena. A further, long-term goal of our work is encyclopedic in nature: to document as far as possible in its entirety the medical knowledge that can be understood by average adult consumers of healthcare services in countries like the United States today.

3.1. Sources of sentences

The sentences that will constitute the candidate input to MEDICALBELIEFNET and MEDICALFACTNET in the pilot phase of our work will derive from two sources. One is the existing Princeton WORDNET itself; the other are medical information sources validated by experts. A preliminary experiment showed that, as concerns the latter, the most promising starting-points for both term and sentence generation are certain online information sources focusing on the coverage of common health problems and targeted specifically to non-specialist users.

3.1.1. Deriving sentences from on-line resources

In a preliminary experiment, sentences were derived by fact statement coders, i.e., researchers in medical informatics, from fact sheets on airborne allergens in the National Institute of Allergy and Infectious Diseases's (NIAID) Health Information Publications and on hay fever and perennial allergic rhinitis in the UK NetDoctor's Diseases Encyclopedia.⁴

The initial documents were divided into paragraph-length sections, and participants were instructed to associate with each section complete, simple and self-contained sentences expressing the generic medical knowledge contained therein, as far as possible drawing on terms used in the original sources. Participants were instructed to eliminate sentences containing anaphora, indexical expressions, formulations of instructions, warnings and the like, and to replace them wherever possible by complete statements (assertions) constructed via simple syntactic modifications. Participants were instructed to include only such terms and information that they themselves judged to be intelligible to non-experts. They were further instructed *not* to avoid redundancy. Below is a text from the NIAID's Health Information Publications factsheet:

“There is no good way to tell the difference between allergy symptoms of runny nose, coughing, and sneezing

and cold symptoms. Allergy symptoms, however, may last longer than cold symptoms.”

These two sentences yielded the following simple statements (or ‘fact statements’ in our terminology):

- (1) Allergies have symptoms.
- (2) Colds have symptoms.
- (3) A runny nose is a symptom of an allergy.
- (4) Coughing is a symptom of an allergy.
- (5) Sneezing is a symptom of an allergy.
- (6) Cold symptoms are similar to allergy symptoms.
- (7) A cold is not an allergy.
- (8) Allergy symptoms may last longer than cold symptoms.

3.1.2. Deriving sentences from WordNet

To derive sentences from WORDNET itself we treat the database as a set of edges between terms in a graph, each edge or link being of the form **tLu** (where **L** ranges over “is-a,” “part-of,” etc.) and **t**, **u** stand for WORDNET terms. Some members of the resulting class of **tLu** tuples can be transformed automatically into English sentences with a minimal amount of post-processing. For example where **t** and **u** are nouns, each “**t is-a u**” formula can be transformed into sentences of the forms “**a t is a u**” and “**a t is a type of u**” (with corrections for articles and plurals), as in: “a cut is a type of wound;” “an abrasion is a wound;” “patients are people;” etc. In other cases, sentences must be derived by less straightforward rules.

3.2. Ratings

The sentences generated by these methods will serve as inputs to validations carried out by human participants. The judgements will be made with respect to two criteria, *understandability/comprehensibility* and *agreement*. Each sentence is rated by three participants. Both laypersons and physicians will rate the sentences for agreement; the judgements of the laypersons will yield MEDICALBELIEFNET, while those of the physicians will constitute MEDICALFACTNET, reflecting the differences in medical knowledge of the two respective populations.

Subjecting WORDNET-derived sentences to the judgements of medical experts will also, as a by-product of this methodology, constitute a test of WORDNET's medical coverage and can thus be used for quality-assurance for MEDICALWORDNET. For example, if a term is incorrectly located in a hierarchy, the statement where this term is defined as a type of its hypernym will be rejected by the medical raters.

3.2.1. Rating for comprehensibility

As a first step, all sentences pass through a *comprehensibility filter*. Laypersons are recruited as participants and asked to rate the sentences for their understandability on a five-point Likert scale ([1] = “I don't understand this sentence at all”; [5] = “I completely understand this sentence.”).

⁴ www.netdoctor.co.uk.

Often, the sentences will be presented in groups, where each sentence involves the same phenomenon or entity, as the example in Section 3.1.1 shows. Raters will be encouraged not to reflect on the statements with which they are presented but to pass immediately from one statement to the next, ignoring any connections between them. To counteract sequencing effects—resting on the fact that people will spontaneously seek to create some coherence even when reading incoherent discourse—we will permute sentences randomly as between different raters, and also introduce sentences about extraneous subject matter. We will also present to all raters groups of independently rated sentences which are designed for benchmarking in evaluations of our own raters and to ensure that results derived from non-reliable raters can be excluded from the final results of the experiment. Only those statements which receive a score of [4] (“I understand this sentence fairly well”) or higher from each of the two raters will serve as input to the next rating step. Hence, it is a major goal of the construction of the MEDICALBELIEFNET, if not its whole *raison d'être*, to discard those medical facts lay users do not understand.

3.2.2. Rating for agreement by laypersons

For the construction of MEDICALBELIEFNET, those sentences that have received high marks for understandability are presented to a second group of layperson raters who rate them for agreement. The raters indicate whether or not they agree with the statements by selecting a score from a five-point Likert scale, ranging from [1] = “I do not agree at all” to [5] = “I agree completely.” In this task, raters will be encouraged to reflect upon their answers wherever they deem it necessary. Statements receiving a score of at least [4] (“I pretty much agree”) from each of the two raters will be included in the MEDICALBELIEFNET sentential corpus.

3.2.3. Rating for agreement by medical experts

To construct MEDICALFACTNET, we will recruit physicians as participants in a parallel rating experiment. They will be asked to judge only those sentences that were highly rated for agreement by the lay participants in the earlier experiment. Here, raters will be encouraged to take their time and even to use reference works whenever they are uncertain. Sentences will be rated on the same five-point Likert scale that was presented to the laypersons. Although the judgements refer to *agreement*, given the expertise of the raters, we interpret the results of this experiment to reflect on the statements' *correctness* as well.

Those sentences that receive a score of [5] from each of the two raters will be included in the MEDICALFACTNET sentential corpus. The entire workflow for the creation of MEDICALWORDNET is depicted in Fig. 1.

4. Some preliminary results

Smith and Fellbaum [39] report on a preliminary experiment to derive basic sentences from medical factsheets. Researchers in medical informatics generated 1644 sentenc-

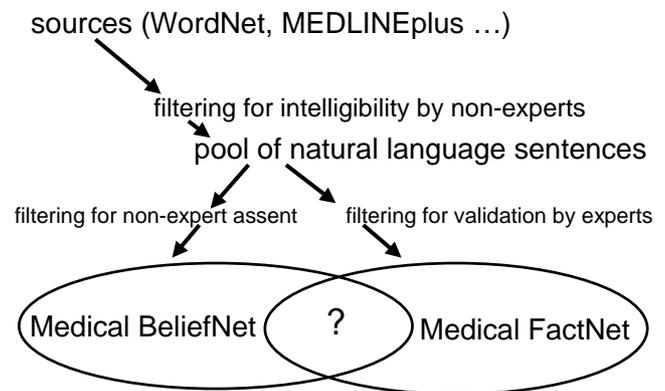


Fig. 1. Workflow for the creation of MEDICALWORDNET.

es in approximately 20 person hours. Five hundred of the sentences were evaluated by pairs of beginning medical students. Fifty-eight percent of the sentences were given the highest score for correctness (5) by both members of each pair. However, a closer analysis showed that the weighted kappa measure for inter-rater agreement was too low to make the results statistically significant. Future experiments will require much larger samples.

5. Challenges for MEDICALWORDNET

Our vision for a new type of information resource for public health outlined in Section 3 is, quite obviously, affected by a series of methodological challenges that need to be addressed before a truly useful infrastructure in the public health domain can be conceived along the lines suggested. These challenges range from social issues of expert-layperson communication (cf. Sections 5.1 and 5.2) and linguistic aspects of encoding medical knowledge (cf. Sections 5.3 and 5.6), to issues pertaining to the very existence of an expert-layperson distinction (cf. Section 5.4), to the status, volume and emergence of medical knowledge (cf. Sections 5.5, 5.8, and 5.9) and to the role MEDICALWORDNET is designed to play as an information system (cf. Section 5.7).

5.1. Doctor–patient communication

The skills of a physician in general practice comprise the ability to acquire relevant and reliable information through communication with patients in non-expert language and to convey diagnostic and therapeutic information in ways tailored to the individual patient. Since the physician, too, is a member of the wider community of non-experts and continues to use non-expert language for everyday purposes, one might assume that there are no difficulties in principle keeping him from being able to formulate medical knowledge in a vocabulary that the patient can understand. As Slaughter [38] and Smith et al. [42] have shown, however, there are limits to this competence. The former examines dialogue between physicians and patients in the form of question–answer pairs, focusing especially on the relations documented in the UMLS Semantic Network.

(<http://www.nlm.nih.gov/pubs/factsheets/umlsemn.html>).

Only in some 30% do the answers given by professionals match the consumers' queries in terms of their semantic relations. An example of one such question–answer pair taken from Slaughter [38, p. 224] illustrates the kinds or linguistic problems that can arise:

Question: “My 7-year-old son developed a rash today that I believe to be chickenpox. My concern is that a friend of mine had her 10-day-old baby at my home last evening before we were aware of the illness. My son had no contact with the infant, as he was in bed during the visit, but I have read that chickenpox is contagious up to 2 days prior to the actual rash. Is there cause for concern at this point?”

Answer: “(a) Chickenpox is the common name for varicella infection. [...] (b) You are correct in that a person with chickenpox can be contagious for 48 h before the first vesicle is seen. [...] (c) The fact that your son did not come in close contact with the infant means he most likely did not transmit the virus. (d) Of concern, though, is the fact that newborns are at higher risk of complications of varicella, including pneumonia. [...] (e) There is a very effective means to prevent infection after exposure. A form of antibody to varicella called varicella-zoster immune globulin (VZIG) can be given up to 48 h after exposure and still prevent disease.”

First, there are lexically rooted mismatches in communication (which may in part reflect legal and ethical considerations) between experts and non-experts. The professional substitutes the medical term, *varicella*, for the folk term *chickenpox* throughout the reply. Second, the physician refers to *the virus*, without spelling out the relation between *chickenpox* and *virus*. Expanding the range of concepts and terms without clarifying their relations is a source of miscommunication and confusion. Third, the questioner requests a yes/no-judgement on the possibility of contagion in a 10-day-old baby. But only section (c) of the answer responds to this question, and this in a way which involves multiple departures from the type of non-expert language that the questioner can be presumed to understand. Fourth, much of the information given is generic and independent of the particular case or context. Patel et al. [32] noted that this is the case even where requests relate to specific episodic phenomena (occurrences of pain, fever, reactions to drugs, etc.). In our example, all sections except for (c) are of this generic kind. The reply is in the form of context-independent statements about causality, about types of persons or diseases, about typical or possible courses of a disease. Accordingly, one major motivation for MEDICALFACTNET is to make such generic medical information accessible to non-experts.

5.2. Online (medical) communication

Understanding patients requires from the practitioner both explicit medical knowledge and tacit linguistic competence dispersed across large numbers of more or less isolated practitioners. This is not a problem so long as this

knowledge is to be applied in a controlled setting, as in face-to-face communication with patients. However, as a result of recent developments in technology, including telemedicine and Internet-based medical query systems [12], we now face a situation where dispersed knowledge requires extensive management. Ely et al. [16] and Jacquemart and Zweigenbaum [22] have shown that clinical questions are expressed via a small number of different syntactic–semantic patterns—about 60 patterns account for 90% of the questions. Yes/no questions like the following are frequent: “Do hair dyes cause cancer?,” “Can I use aspirin to treat a hangover?” With the right sort of information resource, questions of this form can automatically be transformed into answer statements: “Hair dyes can cause bladder cancer,” “Aspirin doesn't help in case of a hangover,” etc. These answers can be linked further to relevant and authoritative sources.

As an example, MEDLINEplus is described in its online documentation as a source of authoritative and up-to-date medical information for both experts and non-experts. Enquirers can use MEDLINEplus⁵ like a dictionary, choosing health topics by keywords. Alternatively, they can use the system's search feature to gain access to a database of relevant online documents selected for reliability and accessibility on the basis of pre-established criteria. Table 1 (taken from [42]) shows the problems that can arise when a system fails to take account of the special features of the knowledge and vocabulary of typical non-expert users. Here success in finding the needed information depends too narrowly on the precise formulation of the query text. Thus, “tremble” and *trembling* call forth different responses (one lists caffeine, the other phobias), even though the terms in question differ only in a morphological affix that does not involve any distinction of meaning. Such problems are characteristic of information services of this kind. Experienced Internet users are of course familiar with the limitations of search engines, and so they are able to modify the formulation of their queries in order to get more informative and, hence, better results. Even experienced users, however, will not be able to overcome the arbitrary sensitivities of an information system, and the latter cannot have the goal of bringing non-experts' ways of using language into line with that of the system.

Patel et al. [32] make clear that if a medical information system is to mediate between experts and non-experts, then it must rest on an understanding of both expert and non-expert medical vocabulary. But terms, or word forms, are not always associated with word meanings in a clear-cut and unambiguous fashion; and the problem of lexical polysemy is compounded when different speaker populations are involved. A lexical database must represent all and only the meanings of each given term in such a way that these meanings can be clearly discriminated and mapped onto word occurrences in natural text and speech. Achieving

⁵ medlineplus.gov.

Table 1
Mismatch of MEDLINEplus and non-expert discourse

Query text	MEDLINEplus response (with links to documents sorted by the following keys)
<i>tremor</i>	Tremor , Multiple Sclerosis, Parkinson's Disease, Degenerative Nerve Diseases, Movement Disorders
<i>intentional tremor</i>	Tremor , Multiple Sclerosis, Parkinson's Disease, Spinal Muscular Atrophy, Degenerative Nerve Diseases
<i>tremble</i>	Anxiety, Parkinson's Disease, Panic Disorder, Caffeine, Tremor
<i>trembling</i>	Anxiety, Parkinson's Disease, Panic Disorder, Phobias, Tremor

these ends is one of the hardest challenges facing both theoretical and applied linguistic science today.

It is generally agreed that the meanings of highly polysemous terms cannot be discriminated without consideration of their sentential contexts [33]. People manage polysemy without apparent difficulties; but modeling human speakers' capacity for lexical disambiguation in automatic language processing systems is hard. The idea underlying the present proposal draws on currently emerging NLP methodologies that harness the ability of powerful and fast computers to store and manipulate both lexical databases and large text collections (or 'corpora'). One strategy is to train automatic systems on large numbers of semantically annotated sentences that are naturally used and understood by human beings, and to exploit standard pattern recognition and statistical techniques for purposes of disambiguation [26] (for a survey, cf. [21]). Words and the representation of their senses, stored in lexical databases, can be linked for this purpose to specific occurrences in corpora.

5.3. Medical facts in natural language

MEDICALWORDNET, linguistically speaking, introduces a new type of text genre, viz., the fact statement. Fact statements are grammatically correct, syntactically very simple ('noise-free') sentences. They bear some resemblance to Chomsky-style "base sentences," i.e., simple declarative sentences that can serve as input to form more complex structures like questions, embedded clauses, etc. While the approach of creating a database of almost 'atomic' fact sentences (which has its historical roots in the philosophy of science as well) is intriguing at first sight, it involves the imposition of severe restrictions on expressibility. In the same way that discourse is not just a sequence of (even non-basic) assertions but rather a rhetorically fully linked set of simple propositions [28], medical knowledge can hardly be expressed as just a *set* of intentionally unrelated fact statements. One may anticipate a significant loss of informativeness when complex diagnostic and therapeutic 'micro-theories' are boiled down to sets of very simple fact statements that lack any further level of structural organization.

Even if we assume that the lack of such further structuring can be compensated for via sheer size in the fact corpus in such a way as to give rise to only little loss of content, we have to bear in mind that size itself will bring problems of its own (cf. Section 5.7). For it is clear that in constructing the MEDICALWORDNET, we will face an exceedingly large and growing number of fact statements, all contained in

more or less complex source assertions (cf. the two-sentence assertion in Section 3.1.1 which yields already eight fact statements). The need to express interrelations between a myriad of plain fact statements will then immediately arise both for the sake of maintaining the huge collection of single statements and for the sake of updating the statement base.

To say that fact statements are self-contained means that they are short of all reference to any sort of linguistic context (as via anaphora). The elimination of contextual information from the original complex natural language assertions is not a trivial task for fact statement encoders. While decontextualization constitutes a less pressing problem for the resolution of pronominal anaphora like "it," "they," and "its" (this process is largely grammar-guided), its consequences are more serious in the case of nominal and bridging anaphora, where hyponymic or meronymic knowledge, respectively, is required to resolve the reference relations. For example, a speaker or writer may refer to "(this) fruit" after having earlier mentioned the discourse referent "apple." "Apple" is a hyponym (subordinate, more specific concept) of "fruit," and this knowledge establishes the (probable) co-reference of the referents targeted by these words. Similarly, a speaker or writer may refer to "pits" after having referred earlier to "peaches"; "pits" here must be interpreted as a meronym (part) of "peaches." Unfortunately, nominal and bridging anaphora of the type illustrated above constitute by far the largest proportion of anaphora in biomedical texts [19].

5.4. Medical experts and laypersons

Fact statements must be understandable by non-experts, defined as "average adult consumers of healthcare services." While it is fairly clear who the "medical experts" are (people with an educational background in medicine), it is much harder to determine the complementary group of medical non-experts. It has already been shown that laypersons faced with the need to acquire information about diseases or other medical issues of immediate interest to themselves or to their families are able to substantially increase their proficiency (e.g., by using the Internet [20]), locally with regard to some given medical topic. For other medical topics, however, the same layperson will possess the low level of competence of other laypersons. Hence, prior knowledge specific to individuals must be an experimental variable that can be controlled for, though it is hard to manipulate this variable given the large number of specialities in the medical domain. Furthermore, medical

understanding grounded in commonsense knowledge is clearly dependent on the general, domain-independent intellectual abilities and educational background of the laypersons involved.

Furthermore, it is difficult to determine the degree and level of understanding of the raters or of the potential end users of a system like MEDICALWORDNET. Assessment of one's own comprehension and ability to learn has been shown to be unreliable [25]. Therefore, a non-expert reading a statement or text might not only miss minor subtleties of content but seriously misinterpret the text without realizing that this is the case.

Another problem is posed by medical knowledge that simply cannot be phrased in a way that it is intelligible to a non-expert. For instance, biochemical explanations of the (mal)functioning of the brain or of genetically motivated diseases might require considerable background knowledge of chemistry or genetics, and consequently attempts to communicate biochemical information might fail for a large subgroup of non-experts lacking this background. This problem is of minor importance, however, given the overall goal of MEDICALWORDNET, which is to capture the entirety of medical information that can be understood by the (average) non-expert.

5.5. Truth status of medical knowledge

Perhaps the strongest claim implied by fact statements is their status as medical truths. This is obviously a problematic notion, and the claim to truth will be difficult to sustain under all circumstances. The empirically grounded approximation of medical truth is the explicit goal of activities under the heading of *evidence-based medicine* [1,18]. Researchers involved are concerned with setting up guidelines for medical treatment given state-of-the-art experimental evidence and thus reflecting a “best practices” approach to the consensual beliefs concerning recommended actions (an ‘implementation’ of medical truths) in given areas of medicine [14]. This approach does not, however, cover all of medicine. Moreover, there are areas in medicine where one is very likely to encounter competing schools of thought as to what is to be counted as medical truth, and consensus about statements in such areas will clearly be very difficult to arrive at [3]. And while consensus beliefs about disorders evolve very slowly, beliefs concerning domains like drugs and therapies are subject to change over time. The assumption of one single established truth targeted by fact statements holding for all areas of medicine is thus in practice untenable. However, it is precisely through experiments such as the construction of MEDICALFACTNET that we shall be in a position to establish whether the idea of consensus about truth will indeed prove to be tenable in the domain at issue.

This problem does not, however, affect the categorial distinction between truth and belief that we propose to determine by means of a non-expert-to-expert filtering and our 5-point assessment scales. That is to say, there will

always be, among the totality of statements accepted as true (‘believed’) by non-experts some that are rejected (as non-facts) by experts. Our methodology for capturing this opposition (including the problematic expert–layperson distinction on which it rests, see Section 5.4) awaits sound experimental assessment in terms of its validity, reliability, and reproduceability. We cannot rule out that, in seeking to determine what ought properly to be contained in MEDICALFACTNET, we will discover that truths and mere beliefs, fuzzy, uncertain, incomplete, or default knowledge will be intertwined; these questions have been long debated in the field of artificial intelligence research [43, Chapter 6] and [7, Chapters 11 and 12].

5.6. Medical knowledge in natural language

In the long term, the engineering of medical lexical knowledge viewed from different backgrounds of expertise (or more precisely, of the true and false beliefs associated with the use of medical terms) will require the interfacing of WORDNET's lay medical terminology with broad-coverage expert-level terminological resources such as the UMLS Metathesaurus [37]. It is unlikely that this can be achieved by simple term matching. For instance, the layperson's understanding of “SARS,” which has no prior non-expert meaning, is likely to be different from the expert's understanding, in spite of the fact that they both use the same term. Furthermore, a term like “chickenpox” as used by laypersons might have different connotations from those of the corresponding expert's term “varicella infection” (cf. Bodenreider and Burgun [4, Section 5.2], and [9, Section 3.3]).

This leads us to certain serious technical implications of the distinction between folk and expert conceptualizations. First of all, the granularity (specificity and depth) of the representations of experts and laypersons differ considerably. While educated laypersons may have a satisfactory understanding of a diverse spectrum of base-level categories [31] covering all the types of phenomena they are likely to encounter, this kind of conceptual structure will constitute merely the starting point for the rich conceptual stratification that characterizes the understanding at an expert level. The observation by Burgun and Bodenreider [9, p. 81] concerning WORDNET's treatment of “grand mal epilepsy” and “generalized epilepsy” as synonymous, two terms whose meanings are treated differently in the UMLS nicely reflects this. As Burgun and Bodenreider explain: “Medically ‘grand mal epilepsy,’ also called ‘tonico-clonic epilepsy,’ is a kind of ‘generalized epilepsy,’ along with ‘tonic epilepsy,’ among others. Therefore, technically, ‘generalized epilepsy’ and ‘grand mal epilepsy’ are better represented in hierarchical relations as in the UMLS than as synonyms as in WORDNET.” It seems fair to assume that “epilepsy” constitutes, as a potential base level category, one of the hyponyms of “disease,” with no further stratification in the layperson's conceptualization.

From a formal, graph theoretical, perspective, there are serious implications for the topologies of the underlying concept graphs, composed of the semantic relations linking term nodes [7]. So any attempt at relating, or even merging, WORDNET and the constituent vocabularies of the UMLS Metathesaurus will be a real challenge because the concept graph of a layperson (represented by the WORDNET topology) and that of an expert (represented by the topologies of UMLS source terminologies) will differ greatly. The question then arises how structurally divergent graphs taken from laypersons and experts can be matched and merged to make use of the medical knowledge already available from the UMLS (cf. the work of Knight and Luke in the non-biomedical domain [24] and in the biomedical domain by Reed and Lenat [35]).

5.7. MEDICALWORDNET as an information system

If WORDNET is moving from a lexical database to MEDICALWORDNET, a propositional knowledge base encoded in natural language statements, entirely new requirements from an information system perspective will arise. One concerns the update policy. Since medical knowledge, especially in the domain of drugs and therapies, is changing rapidly—new knowledge is being discovered, established knowledge claims reassessed and possibly adjusted, old or outdated knowledge claims are eliminated—the proposition base must reflect the volatile nature of the domain. Now given the huge number of single, *logically* fully independent and, hence, unrelated, fact statements, provisions have to be made to locate the ones affected by changes and to modify them accordingly, to identify and remove potential contradictions, etc.

The problem may be crucial anticipating that lots of paraphrases of the same underlying proposition may occur in different verbal surface realizations (consider, e.g., “Sneezing is a symptom of flu.” vs. “Flue has sneezing as a symptom.” vs. “Influenza has sneezing as a symptom.”—we cannot avoid getting results like this, given the way we have set up the instructions for generation of the fact statements thus far). While for the acquisition and collection of medical facts and beliefs this variety does not create any harm, from an information system perspective there is a great potential for not being able to track the same propositions (encoded as different sentences) to properly reflect all effects of changes in medical knowledge in the database of belief and fact sentences that occur over time. To the best of our knowledge, we do not know of any natural language language system capable of recognizing and neutralizing paraphrases of propositions on such a large scale (covering the whole of medicine).

The ensuing tracking task is enormous and requires a substantial amount of *understanding* of the propositions. Technically, mechanisms for partitioning the proposition base might be useful to counteract the otherwise flat structure of sets of propositions [27,49]. Similarly, it may also be possible to do justice to alternative medical approaches, as

well as the expert/layperson distinction, by incorporating different view mechanisms [11,44].

MEDICALWORDNET as a proposition base also poses fundamentally new constraints on its user interface. The query language should be sophisticated enough to allow more than partial string matches, e.g., via regular expressions. Linguistically sensitive access via phrases or verbs, if desired, will require an appreciable amount of syntactic analysis (at a minimum, chunking, and shallow parsing) [23]. Further, some way of integrating already existing terminologies (e.g. as collects within the UMLS Metathesaurus) [37]) should be provided to link up to prior medical knowledge and support content-focused access. But it is an open issue as to what kind of support experts and laypersons might really want in a new scenario like the one outlined here. These are questions that could and should be investigated.

5.8. The volume of medical knowledge

The UMLS [37] in its 2004 edition supplies more than two million biomedical terms, and more than eleven million relations between these terms. Given the procedure for deriving base sentences outlined in Section 3.1.1, we would arrive at well over eleven million fact statements, simply by extracting the corresponding sentences from these relations. Upon recombination with already available fact statements from the original WORDNET, plus additional material from the Web, this will result in a very large overall number of statements. That huge number will certainly decrease due to the fact that most of these statements will go (far) beyond what a layperson can understand. Although we do not consider entering all of these propositions into MEDICALWORDNET—many of them will not pass the intelligibility filter—we still face a tremendous knowledge management and assessment problem. The remaining number of statements coming from the complete UMLS resource will still constitute a huge data set which will require a lot of efforts to judge in terms of its belief and fact status.

The problem of size becomes even more daunting when one considers principled ways of encoding medical knowledge. Consider the possibility of expressing (interesting subsets of) the transitive closure of medical facts starting from a fairly general statement such as “*drugs treat diseases*” down to the level of concrete assertions, say, “*Aspirin treats headache*.” Once we set up MEDICALWORDNET, do we want to explicitly and instantaneously enumerate all the derivable fact statements on the basis of the semantic relations available from the UMLS (in AI jargon, the approach of *read-time inferencing*), or do we want to generate such fact statements on demand only (a kind of *question-time inferencing*)? Read-time inferencing, in this example means that we take all instances of drugs and diseases we know of and instantiate all (drug, disease) pairs, incorporating appropriate (integrity or sortal) restrictions, if available, on which diseases are cured which drugs. Question-time inferences, on the other hand, try to establish a

treatment link between, say, “Aspirin” and “headache,” only if required by some query to the proposition base. If we opt for the first (read-time) proposal, the knowledge base will further explode (including the load involved in filtering out appropriate truth/belief statements). If we consider the second (question-time) alternative, we have to supply appropriate reasoning machinery (for example, via some form of Description Logic classifier or realizer [30]), which is currently not available in the context of WORDNET. It would also require that all fact statements be mapped to a formal representation level (e.g., Description Logic or full predicate logic) to compute inferences such as subsumption relations [2].

5.9. On discovery procedures

Finally, a serious empirical question concerns the *representativeness* of the sample from which fact statements are derived. Representativeness, or the lack thereof, directly affects the completeness (or selection bias) of the set of propositions derived. Once representativeness is achieved, the consistency of the statement coders and statement raters (inter- and intra-reliability) must be ensured. Although we envisage quantifying the degree of coder and rater consistency, it is also not yet clear what kind of consistency will really be measured—does it relate to the coders’ or raters’ competency with respect to the decisions as to truth/belief status, or to the layperson/expert distinction, or to the linguistic form of fact statements (are paraphrases acceptable?, are more specific statements as acceptable as more general ones and where do we stop generalizing?).

There are different criteria for assessing the fact statements. Considering work done on the evaluation of the results of information extraction systems [10], we may want to measure partial degrees of overlap (e.g., if “curing leprosy with penicillin” is the full statement, how do we deal with “curing leprosy” only?), or the completeness of facts (e.g., assume we knew from an oracle that a complex natural language utterances contained nine fact statements, how do we deal with the extraction of only seven from these nine? Who would provide the corresponding ground truth?), or in terms of a general-to-specific interval within a well-defined fact subspace (how would we scale the granularity steps, e.g., for “aspirin treats headache” up to “a drug treats a disease?”), or in terms of (in)consistency. The latter aspect is interesting insofar as we might also be challenged to deal in the rating step with ‘semi-true’ statements (depending on more and more complex context conditions) as well as undecidables, i.e., questions that even medical people have no certain answers to.

6. Conclusions and outlook

The MEDICALWORDNET project, as described in the first part of this paper, is to some degree a visionary enterprise. As described in the present paper and in [39], we have so far carried out pilot studies only to test the proposed meth-

odology. Lack of funding has prevented larger-scale implementations. A next step might apply the methodology to a limited domain and evaluate the analytical benefits.

Some of the problems that need to be solved have been discussed in this paper. Despite the concerns expressed, we conclude that the conception of a combined lexical and propositional database and the empirically established distinction between expert and lay medical knowledge is a useful one and deserves the considerable effort necessary to realize it. Whether our approach is feasible or not, bringing medical expertise to public health information platforms such as Internet portals is an issue of great societal importance and potential value for health consumers seeking medical advice without the need for (expert) intermediaries.

References

- [1] Abalos E, Carroli G, Eugenia Mackay M. The tools and techniques of evidence-based medicine. *Best Pract Res Clin Obstet Gynaecol* 2005;19(1):15–26.
- [2] Baader F, Nutt W. Basic description logics. In: Baader F, Calvanese D, McGuinness DL, Nardi D, Patel-Schneider PF, editors. *The description logic handbook, theory, implementation, and applications*. Cambridge, UK: Cambridge University Press; 2003. p. 43–95.
- [3] Barrett B, Marchand L, Scheder J, Beth Plane M, Maberry R, Appelbaum D, et al. Themes of holism, empowerment, access, and legitimacy define complementary, alternative, and integrative medicine in relation to conventional biomedicine. *J. Altern. Complement. Med.* 2003;9(6):937–47.
- [4] Bodenreider O, Burgun A. Characterizing the definitions of anatomical concepts in WORDNET and specialized sources. In *Proceedings of the 1st international conference of the global wordnet association*. Mysore, India, January 21–25, 2002, 2002. p. 223–239.
- [5] Bodenreider O, Burgun A, Mitchell JA. Evaluation of WORDNET as a source of lay knowledge for molecular biology and genetic diseases: A feasibility study. In Baud R, Fieschi M, Le Beux P, Ruch P, editors. *Medical Informatics Europe 2003—Proceedings of the 18th international congress of the european federation for medical informatics. The New Navigators: From Professionals to Patients, number 95 in Studies in Health Technology and Informatics*, St. Malo, France, May 4–7, 2003. Amsterdam: IOS Press; 2003. p. 379–384.
- [6] Bodenreider O, Smith B, Kumar A, Burgun A. Investigating subsumption in DL-based terminologies: a case study in SNOMED CT. In Hahn U, Schulz S, Cornet R, editors. *KR-MED 2004—Proceedings of the 1st international workshop on formal biomedical knowledge representation, collocated with the 9th international conference on the principles of knowledge representation and reasoning (KR 2004)*, Whistler, B.C., Canada, June 1, 2004. Bethesda, MD: American Medical Informatics Association (AMIA); 2004. p. 12–20. Published via <http://CEUR-WS.org/Vol-102/>.
- [7] Brachman RJ, Levesque HJ. *Knowledge representation and reasoning*. Amsterdam: Elsevier/Morgan Kaufmann; 2004.
- [8] Buitelaar P, Sacaleanu B. Extending synsets with medical terms WORDNET and specialized sources. In *Proceedings of the 1st international conference of the global wordnet association*. Mysore, India, January 21–25, 2002; 2002.
- [9] Burgun A, Bodenreider O. Comparing terms, concepts and semantic classes in WORDNET and the Unified Medical Language System. In *Proceedings of the NAACL 2001 Workshop ‘WORDNET and Other Lexical Resources: Applications, Extensions and Customizations’*, Pittsburgh, PA, June 3–4, 2001, New Brunswick, NJ: Association for Computational Linguistics, 2001; p. 77–82.
- [10] Chinchor N, Hirschman L, Lewis DD. Evaluating message understanding systems: an analysis of the third Message Understanding Conference (MUC-3). *Comput Linguist* 1993;19(3):409–47.

- [11] Claybrook BG, Claybrook A-M, Williams J. Denning database views as data abstractions. *IEEE Trans Softw Eng* 1985;SE-11(1):3–14.
- [12] Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. *Health Educ Res* 2001;16(6):671–92.
- [13] Deese J. The associative structure of some English adjectives. *J Verbal Learn Verbal Behav* 1964;3:347–57.
- [14] Denny K. Evidence-based medicine and medical authority. *J Med Humanit* 1999;20(4):247–63.
- [15] W.A. Newman Dorland, editor, *Dorland's Illustrated Medical Dictionary*. Philadelphia: W.B. Saunders Company, 27th edition, 1988.
- [16] Ely JW, Osheroff JA, Gorman PN, Ebell MH, Lee Chambliss M, Pifer EA, et al. A taxonomy of generic clinical questions: classification study. *Br Med J* 2000;321:429–32.
- [17] Fellbaum C, editor. *WORDNET: an electronic lexical database*. Cambridge, MA: MIT Press; 1998.
- [18] Gerber A, Lauterbach KW. Evidence-based medicine: why do opponents and proponents use the same arguments? *Health Care Anal* 2005;13(1):59–71.
- [19] Hahn U, Romacker M, Schulz S. Discourse structures in medical reports—watch out! The generation of referentially coherent and valid text knowledge bases in the MEDSYNDIKATE system. *Int J Med Inform* 1999;53(1):1–28.
- [20] Hart A, Henwood F, Wyatt S. The role of the Internet in patient–practitioner relationships: findings from a qualitative research study. *J Med Internet Res* 2004;6(3):e36.
- [21] Ide N, Véronis J. Introduction to the Special Issue on Word Sense Disambiguation: the state of the art. *Comput Linguist* 1998;24(1):1–40.
- [22] Jacquemart P, Zweigenbaum P. Towards a medical question-answering system: a feasibility study. In Baud R, Fieschi M, Le Beux P, Ruch P, editors. *Medical Informatics Europe 2003—Proceedings of the 18th international congress of the european federation for medical informatics. The New Navigators: From Professionals to Patients, number 95 in Studies in Health Technology and Informatics, St. Malo, France, May 4–7, 2003*. Amsterdam: IOS Press, 2003; p. 463–468.
- [23] Jurafsky D, Martin JA. *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice Hall; 2000.
- [24] Knight K, Luk SK. Building a large-scale knowledge base for machine translation. In AAAI'94—Proceedings of the 12th national conference on artificial intelligence, vol. 1. Seattle, WA, USA, July 31– August 4, 1994. AAAI Press & MIT Press, Menlo Park, CA, 1994; p. 773–778.
- [25] Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Social Psychol* 1999;77(6):1121–34.
- [26] Leacock C, Chodorow M, Miller GA. Using corpus statistics and WORDNET relations for sense identification. *Comput Linguist* 1998;24(1):147–65.
- [27] Lenat DB, Guha RV. Building large knowledge-based systems. Representation and Inference in the Cyc project. Reading, MA: Addison-Wesley; 1990.
- [28] Mann WC, Thompson SA. Rhetorical structure theory: toward a functional theory of text organization. *Text* 1988;8(3):243–81.
- [29] Miller GA. WORDNET: a lexical database for English. *Commun ACM* 1995;38(11):39–41.
- [30] Möller R, Haarslev V. Description logic systems. In: Baader F, Calvanese D, McGuinness DL, Nardi D, Patel-Schneider PF, editors. *The description logic handbook. Theory, implementation, and applications*. Cambridge, UK: Cambridge University Press; 2003. p. 282–305.
- [31] Murphy GL, Lassaline ME. Hierarchical structure in concepts and the basic level of categorization. In: Lamberts K, Shanks D, editors. *Knowledge, concepts, and categories*. Cambridge, MA: MIT Press; 1997. p. 93–113.
- [32] 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.
- [33] Pustejovsky J. *The generative lexicon*. Cambridge, MA: MIT Press; 1995.
- [34] Qing Z, Cimino J. Mapping medical vocabularies to the unified medical language system. In *Proceedings of the AMIA annual fall symposium, 1996*; p. 105–109.
- [35] Reed SL, Lenat DB. Mapping ontologies to Cyc. In *Proceedings of the AAAI 2002 Conference Workshop on 'Ontologies for the Semantic Web'*. Edmonton, Canada, July 2002. Menlo Park, CA: AAAI Press, 2002.
- [36] Rosenzweig MR. International Kent–Rosanoff word association norms, emphasising those of French male and female students and French workmen. In: Postman L, Keppel G, editors. *Norms of word association*. New York: Academic Press; 1970. p. 95–176.
- [37] UMLS. *Unified Medical Language System*. National Library of Medicine, Bethesda, MD, 2004.
- [38] Slaughter L. *Semantic Relationships in Health Consumer Questions and Physicians' Answers: A Basis for Representing Medical Knowledge and for Concept Exploration Interfaces*. PhD thesis, University of Maryland at College Park, 2002.
- [39] Smith B, Fellbaum C. Medical wordnet: a new methodology for construction and validation of information resources for consumer health. In *Proceedings of COLING, 2004*.
- [40] Smith B, Köhler J, Kumar A. On the application of formal principles to life science data: a case study in the GENE ONTOLOGY. In Rahm E, editor. *DILS 2004—Proceedings of the 1st international workshop on data integration in the life sciences, Lecture Notes in Artificial Intelligence, vol. 2994, Leipzig, Germany, March 25–26, 2004*. Berlin: Springer, 2004; p. 79–94.
- [41] Smith B, Rosse C. The role of foundational relations in the alignment of biomedical ontologies. In Fieschi M, Coiera E, Jack Li Y-C, editors. *MEDINFO 2004—Proceedings of the 11th world congress on medical informatics. Vol. 1, number 107 in Studies in Health Technology and Informatics, San Francisco, CA, USA, September 7–11, 2004*. Amsterdam: IOS Press, 2004; p. 444–448.
- [42] Smith CA, Stavri PZ, Chapman WW. In their own words? A terminological analysis of e-mail to a cancer information service. In Kohane IS, editors. *AMIA 2002—Proceedings of the annual symposium of the american medical informatics association. biomedical informatics: one discipline, San Antonio, TX, November 9–13, 2002*. Philadelphia, PA: Hanley & Belfus, 2002; p. 697–701.
- [43] Stefik M. *Introduction to knowledge systems*. San Francisco, CA: Morgan Kaufmann; 1995.
- [44] Storey VC, Goldstein RC. A methodology for creating user views in database design. *ACM Trans Database Syst* 1998;13(3):305–38.
- [45] Tuns D. Special issue on the BALKANET project. *Rom J Inf Sci Technol* 2004;7(12):191–238.
- [46] Turcato D, Fass D, Tisher G, Popowich F. Fully automatic bilingual lexical acquisition from EUROWORDNET. In *Proceedings of the NAACL 2001 workshop 'WORDNET and other lexical resources: applications, extensions and customizations'*. Pittsburgh, PA, June 3–4, 2001. New Brunswick, NJ: Association for Computational Linguistics, 2001.
- [47] Voorhees EM. Using WORDNET for text retrieval. In: Fellbaum C, editor. *WORDNET: an electronic lexical database*. Cambridge, MA: MIT Press; 1998. p. 285–303.
- [48] Vossen P, editor. *EUROWORDNET: a multilingual database with lexical semantic networks*. Dordrecht: Kluwer Academic Publishers; 1998.
- [49] Wiederhold G, Rathmann P, Barsalou T, Suk Lee B, Quass D. Partitioning and composing knowledge. *Inf Syst* 1990;15(1):61–72.
- [50] Xiao C, Rösner D. Finding high-frequent synonyms of a domain-specific verb in English sub-language of MEDLINE abstracts using WORDNET. In Sojka P, Pala K, Fellbaum C, Vossen P, editors. *GWC 2004—Proceedings of the 2nd international conference of the global wordnet association, Brno, Czech Republic, January 20–23, 2004*, 2004; p. 242–247.