AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD

Research and Applications

Cross-lingual Unified Medical Language System entity linking in online health communities

Yonatan Bitton,[1] Raphael Cohen,[1] Tamar Schifter,[2] Eitan Bachmat,[1] Michael Elhadad,[1]* and Noémie Elhadad[3]

[1]Department of Computer Science, Ben Gurion University, Beer Sheva, Israel, [2]Gertner Institute for Epidemiology and Health Policy Research, Tel HaShomer, Israel, and [3]Department of Biomedical Informatics, Columbia University, New York, New York, USA

*Corresponding Author: Michael Elhadad, Department of Computer Science, Ben Gurion University, Beer Sheva, Israel; elhadad@cs.bgu.ac.il

ABSTRACT

Objective: In Hebrew online health communities, participants commonly write medical terms that appear as transliterated forms of a source term in English. Such transliterations introduce high variability in text and challenge text-analytics methods. To reduce their variability, medical terms must be normalized, such as linking them to Unified Medical Language System (UMLS) concepts. We present a method to identify both transliterated and translated Hebrew medical terms and link them with UMLS entities.

Materials and Methods: We investigate the effect of linking terms in Camoni, a popular Israeli online health community in Hebrew. Our method, MDTEL (Medical Deep Transliteration Entity Linking), includes (1) an attention-based recurrent neural network encoder-decoder to transliterate words and mapping UMLS from English to Hebrew, (2) an unsupervised method for creating a transliteration dataset in any language without manually labeled data, and (3) an efficient way to identify and link medical entities in the Hebrew corpus to UMLS concepts, by producing a high-recall list of candidate medical terms in the corpus, and then filtering the candidates to relevant medical terms.

Results: We carry out experiments on 3 disease-specific communities: diabetes, multiple sclerosis, and depression. MDTEL tagging and normalizing on Camoni posts achieved 99% accuracy, 92% recall, and 87% precision. When tagging and normalizing terms in queries from the Camoni search logs, UMLS-normalized queries improved search results in 46% of the cases.

Conclusions: Cross-lingual UMLS entity linking from Hebrew is possible and improves search performance across communities. Annotated datasets, annotation guidelines, and code are made available online (https://github.com/yonatanbitton/mdtel).

Key words: UMLS, natural language processing, online health communities

INTRODUCTION

Online health communities (OHCs) have become increasingly relevant as a source of real-world evidence for researchers and informational support for patients and their caregivers.[1–3] The content of OHC posts is challenging to process automatically, but it is a critical step to enable downstream applications to advance biomedical knowledge, to support community moderators, and to support patients in finding and making sense of the rich information contained in OHC posts. Named-entity recognition has been investigated in English OHCs and social media.[4,5] In languages other than English, a set of additional challenges compound natural language

processing of health texts, including the fact that most languages do not have a version of the Unified Medical Language System (UMLS) available to them to normalize concepts.[6] In this article, we focus on the challenge of cross-lingual entity linking (XEL), which consists in mapping entity mentions in text of a source language to entities in a knowledge base, specifically UMLS,[7] in a target language.

We ground our work on a popular Hebrew-speaking OHC, Camoni.[8] A salient aspect of Hebrew health texts is the heavy use of transliteration of medical terms. Transliteration consists of writing a word from a source language (eg, English) in a different target script (eg, Hebrew alphabet) using the closest corresponding letters. Transliterations are thus phonetically similar to the source word.

We performed an empirical analysis of the way medical terms are mentioned in the Camoni communities. We find that most of the medical terms are transliterated, and furthermore, the transliterations include unreliable abbreviations and misspellings. Different people transliterate the same word in different forms, which creates variability. As discussed in Goldberg and Elhadad,[9] when foreign words are borrowed in languages with similar alphabets and sound systems, it is easy to map them back to the source terms, as the words are written in similar manners (usually, only small suffix variants are introduced). Words borrowed into languages with different writing and sound systems (eg, English words in Japanese, Hebrew, and Arabic texts) are more challenging. In addition, automatic translation does not cover most of the observed mentions of medical terms. In our test, automatic Hebrew-English translation in the medical domain fails to translate over 38% of the medical terms (Supplementary Appendix S1).

The high prevalence of transliteration of medical terms leads to high variability, which is a challenge for information retrieval and data mining.[10,11] Camoni, like many OHCs, relies on Google Custom Search application programming interface (API) as their search engine. Many user queries mention medical terms, which are very likely to include noisy transliterations. For example, a Hebrew query with a spelling variant such as "How do I know I have fibromyalegia?" does not return any results when "fibromyalgia" is transliterated. In this work, we empirically assess the prevalence of this issue, find it related in the case of Hebrew medical terms to the variability introduced by transliteration, and introduce a method based on UMLS normalization of medical terms in the patient-authored content.

Our key contributions are:

- An unsupervised method for creating a transliteration dataset. The dataset is used for training a state-of-the-art transliteration model;
- An efficient method to match the identified transliterated entities into their UMLS concept unique identifier (CUI);
- A contextual relevance model developed to detect medical term mentions in each disease-specific community.

Our method is applicable to other languages with different scripts and we document steps to reproduce both intrinsic and extrinsic evaluation methods for XEL applied to UMLS.

Background and Significance

Entity linking is the task of linking spans in text to concepts in a knowledge base such as the UMLS. It is particularly useful in the clinical and health domains, when documents are written in free text and frequently refer to biomedical concepts.[12,13] XEL is needed when a document is in a source language, different from the language of the labels used in the target knowledge base. About 71% of the concept names in UMLS are labeled in English. Other languages occur much less: for Spanish, ~10.2%, and for French, ~2.85%; for Hebrew, only 485 terms occur in the UMLS.[14] When the source and target languages operate over different alphabets and sound systems, both translation and transliteration of terms must be handled.

Several CLEF eHealth challenges (2015-2019) have focused on named-entity recognition in English and French in biomedical articles, with applications to multilingual information extraction from health reports. The 2017-2018 tasks[10,11] explore automatic assignments of International Classification of Diseases–Tenth Revision codes to health-related documents in English, French, Hungarian, Italian, and German. The 2015-2016 tasks focus on information retrieval in biomedical domains. Xu et al[15] introduce a model to identify cross-lingual candidates for concept normalization using a character-based neural translation model trained on a multilingual biomedical terminology. They use UMLS data in Spanish, French, Dutch, and German. Our approach is similar and extends this work on the following dimensions: (1) we focus on OHCs with noisy text as opposed to scientific articles; (2) we study the case of transliteration in Hebrew, which uses a different alphabet and sound system than English; and (3) we introduce an unsupervised method to create a transliteration dataset, which can be used across languages with varying scripts and without existing resources.

A tagger that detects entities in Arabic medical documents[16] tags terms with semantic class such as medical problem, test, and treatment, a task introduced in the 2010 i2b2 challenge. Data from Wikipedia, DBpedia, and SNOMED-CT (Systematized Nomenclature of Medicine Clinical Terms) are leveraged, and a binary classifier for each semantic category is trained. Our work differs, as we perform a coarser form of detection for medical terms in general, without classifying their type, but aim for high recall in an unsupervised manner. In our intrinsic evaluation, we reproduce a similar task for Hebrew for the UMLS semantic groups of *Disorders* and *Chemical or Drugs*.

Previous work has addressed the task of detecting transliterations[9] and map them back to source terms.[17,18] The NEWS 2018 Shared Task on Transliteration of Named Entities focused on identifying transliterations in multiple languages in the news domain.[19] In our work, we build on a neural machine translation model for named-entity transliteration developed for this task.[20] The model uses a deep attentional recurrent neural network (RNN) encoder-decoder model. This architecture was ranked first in several tracks at the NEWS 2018 Shared Task when applied to Hebrew.

MATERIALS AND METHODS

Our approach to cross-lingual UMLS entity linking is called Medical Deep Transliteration Entity Linking (MDTEL). Given a community post document in Hebrew $(w_1, w_2 \ldots w_n)$, we want to tag the spans $(s_1 \ldots s_k)$, which contain medical terms and link them with a UMLS CUI (Figure 1). Figure 2 shows the overall structure of the processing pipeline.

- In an offline process, we prepare a forward transliteration model to transliterate the UMLS into Hebrew. In addition to the transliterated UMLS terms, we include translations for key words that are related to the OHC main topic selected from a manually collected medical glossary. The result of this module is a list of pairs (<source UMLS term in English>$_i$, <target Hebrew term>$_i$). That is, given a term such as "diabetes," we generate multiple
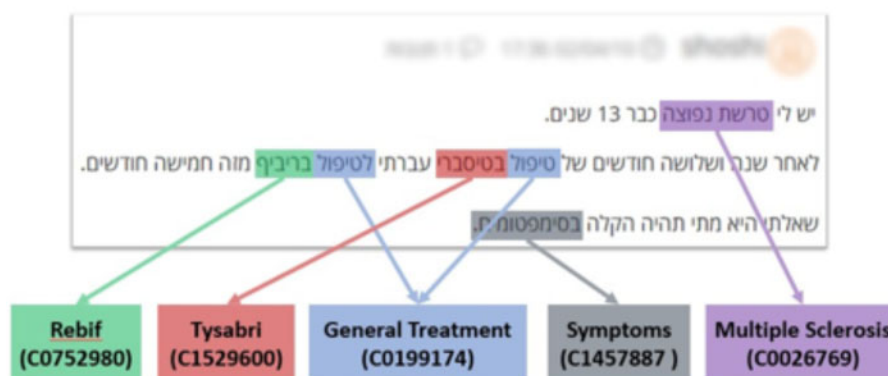
Figure 1. This forum post contains 26 words, and 6 spans that link to 5 different unique identifiers of Unified Medical Language System medical terms. Notice that a span can contain more than 1 word (like the term "multiple sclerosis"), a single Unified Medical Language System concept unique identifier can be referenced from several places in the same post.
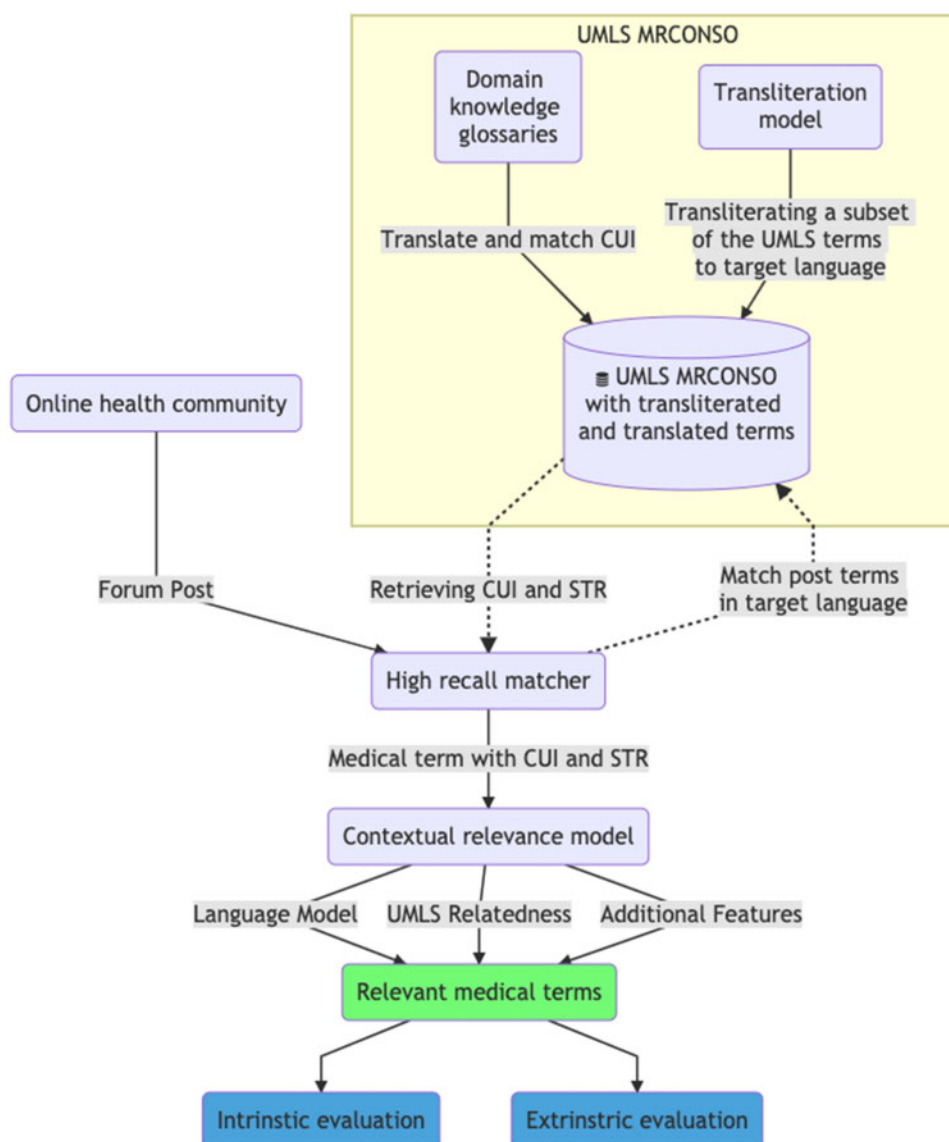

Figure 2. Cross-lingual entity linking processing pipeline: offline, a filtered subset of Unified Medical Language System (UMLS) is transliterated and translated into Hebrew, producing pairs <CUI English term; Hebrew term>; a post from the online health community is passed to the High recall matcher which searches for matches, by intersecting n-grams of the posts text with the Hebrew medical terms in the pairs, producing a list of candidate concept unique identifiers (CUIs). The contextual relevance model uses language models features, the UMLS Relatedness package, and additional features in order to filter relevant medical terms in the context of the post. STR: String.

candidate Hebrew terms such as a highly likely transliteration "*di-a-bi-tis*" and its translation "*sa-ke-re-t.*"

- A matcher to match spans from the post to the transliterated or translated UMLS Hebrew terms, producing a high-recall list of candidate matches.
- A contextual relevance model that filters the high-recall list by detecting matches that are not used in a medical sense.

We perform 2 evaluations to assess cross-lingual entity linking performance: (1) in an intrinsic test, we measure precision, recall, and F1 score of the XEL model on a manually annotated dataset of Hebrew documents with marked UMLS concepts for *Disorders* and *Chemicals or Drugs*; and (2) in an extrinsic evaluation, we quantify the impact of UMLS term normalization on Hebrew posts from the OHC on information retrieval quality.

Data
Camoni corpus
The Camoni communities have about 20 000 registered members and 100 000 unique visitors per month.[21] Camoni is organized around 39 disease-specific communities (Supplementary Table S1.2). We extract text from 3 communities (diabetes, sclerosis, and depression), for a total of 55 000 posts and 2.5 million tokens (Supplementary Table S1.3).

Gold-standard annotation of UMLS terms
To test XEL performance, we constructed an annotated dataset in which all mentions of specific UMLS terms are annotated. Annotation guidelines (https://github.com/yonatanbitton/mdtel) accounted for linguistic properties of Hebrew text to handle compound nominal expressions, aggregated prepositions, and conjunctions. We focused this annotation on the most common semantic groups: *Disorders* and *Chemical and Drugs*.[22] The guidelines were refined through 3 rounds of test with 3 annotators on 50 posts and analyzing disagreement.

Annotation was carried out in the Doccano online annotation tool (https://doccano.herokuapp.com/) (Figure 3). Two fifth-year medical students annotated documents and prepared a set of posts from the 3 OHCs. A total of 802 forum posts were annotated with overall 4106 term mentions and 1700 unique terms (Supplementary Table S2.1). The Cohen's kappa interannotator agreement was 0.76, 0.76, and 0.71 for the diabetes, sclerosis, and depression communities, respectively, indicating high agreement. For the intrinsic evaluation, given the annotated mentions, we executed a pretrained model of our method and aligned the Hebrew mentions with UMLS CUIs. One of the annotators then validated the linking as accurate or not.

General medical term annotation
In addition to the fine-grained UMLS term mentions gold dataset, we prepared a coarse annotation data to tag any mention of a "medical term." We collected 100 posts of each community and manually tagged relevant medical terms. Statistics of this dataset are shown in Supplementary Table S5.1. In the processing pipeline, we use this small dataset to train a binary relevance classifier for text mentions conditioned on the community.

Building a transliteration model
We first build a transliteration model capable of mapping English medical terms to multiple likely Hebrew transliterations (transliteration is a noisy process even when people do it). We learn a character-based transducer that maps English strings to corresponding Hebrew strings. The model is based on a neural model based on an RNN encoder-decoder architecture.[20]

To train such a model, a large dataset of pair (En, He) term pairs is required. We introduce an unsupervised method to synthesize such a dataset, instead of creating it through manual annotation. We act under the hypothesis that medical terms are usually transliterated. We initialize the data from a dictionary of English medical terms and add some words from medical glossaries we collected from trusted Web sources.[23] We then use Google Translate to translate those words into Hebrew. In some cases, the terms are mapped to transliterations, and in others to Hebrew word translations. This step yielded 106 573 pairs. Because we assume most of the mappings are transliterations, we build on the generalization capability of the RNN architecture to filter out the cases of translations in this collected data.

Mapping UMLS terms to Hebrew terms
To produce the list of <En, He> UMLS terms, we apply the following steps:

1. **Filter UMLS terms by type:** For each evaluation experiment, we filter the UMLS in different ways: in the intrinsic evaluation, we focus on the disorders and drugs semantic groups. In the extrinsic evaluation, we include concepts expected in typical consumer health corpus and keep concepts with TTY in adjectives, drug names, supplementary concepts, common names, entry, hierarchical terms, brand names, finding names, clinical synopsis, ingredients, scientific names, active substance, main heading, and chemicals. The filtered terms are mostly abbreviations or acronyms. We collected 2.8 million medical terms; 640 000 of these are single words, and 2.160 million are multiwords.
2. **Transliterate:** We construct a transliteration model and apply it on the selected UMLS terms. We collect the top 3 transliterations predicted for each UMLS term. This process results in a list of 8.4 million Hebrew terms mapped to their source UMLS CUI for general medical terms.
3. **Translate:** For some medical terms, a Hebrew translation is more likely to be used than a transliteration (eg, the Hebrew translation of "diabetes" is more common than its transliteration). To identify those terms, we collect a list of pairs (x_i, y_i), where y_i is the Hebrew translation of the English word x_i in case a transliteration is infrequent. The identification of these terms is described in Supplementary Appendix S2. This process produced 14 766 Hebrew translated (and nontransliterated) terms mapped to their source UMLS CUI.

UMLS XEL
We now describe how to exploit the dataset of medical term pairs (transliterated and translated) and perform UMLS entity linking on the OHC documents. We proceed in 2 steps: (1) produce candidate matches with high recall and (2) filter candidates using a machine learning model that was trained to detect relevant terms in the context of a post.

Our entity-linking method has 2 contributions:

- UMLS entity linking is conditioned on the topic of the document. We use a different relevance classifier for each disease-specific community.
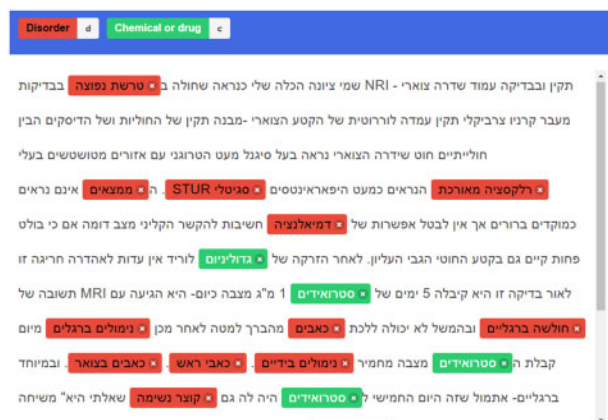
Figure 3. Doccano online annotation tool with the Hebrew Unified Medical Language System schema.

```
for message in corpus do
    matches-for-message ⇐ empty-list
    n-grams ⇐ generate message words n-grams(N = 3)
    for gram in n-grams do
        for i in range(1, N+1) do
            heb-candidate = string(gram[:i])
            best-match = searcher.ranked-search(heb-candidate, sim-threshold)[0]
            if filter-model.predict(heb-candidate, best-match, message-context) == 0 then
                matches-for-message.append(best-match)
            end if
        end for
    end for
end for
```

Figure 4. Overall algorithm for entity linking—combining high-recall n-grams matching and contextual filtering.

- Relevance classification is contextual—the same term may be tagged as a medical mention in some contexts, but not in others.

High-recall matching

To generate a high-recall list of candidates, we scan all posts in each community, and search all matches for any term y_i from the pairs dataset (either transliterations or translations). We match both the observed tokenized form of the Hebrew text in the post and its lemmatized text. Lemmatization is particularly important in Hebrew because most function words (articles, prepositions, and conjunctions) appear in Hebrew in a form agglutinated to the next word. We use the YAP model[24] for Hebrew lemmatization. We use a fuzzy string matching algorithm to match between the UMLS transliterated terms and the post tokens in a parallel manner.[25] Because many UMLS terms are multiword expressions, we match n-grams (1,2,3) in the text. In the full algorithm, *filter_model* refers to the relevance model described in Figure 4.

Contextual relevance ranking

The relevance classifier takes the whole post as a context to determine whether a term mention is used in a medical sense, because some words may be considered medical in some contexts and not medical in other contexts. We expect that in a nonmedical corpus like Wikipedia, medical terms will have low probability, unless they are common phrases. We train a Hebrew language model on a Hebrew Wikipedia corpus[26] with 3.8 million sentences, which were split 80% train and 20% validation. We trained multiple neural language models (Supplementary Appendix S3). Using this model, we calculate $p_{lm}(match \mid context)$ and combine a UMLS similarity feature between the candidate term and the name of the community

Table 1. Intrinsic evaluation: Entity-level recognition (exact span) performance on gold-standard dataset

	Accuracy	f1_score	Precision	Recall	Support
Diabetes	0.97	0.73	0.71	0.75	314
Sclerosis	0.98	0.76	0.82	0.71	306
Depression	0.99	0.75	0.77	0.73	262
Weighted average	0.98	0.75	0.77	0.73	—

(eg, "diabetes") and count-based features. We train binary classifiers on the gold-standard and general datasets we prepared.

This classifier is eventually applied on the list of high-recall candidates produced earlier and provides the final identified terms in the original Hebrew post, each linked to the corresponding UMLS concept.

RESULTS

Intrinsic evaluation

In the intrinsic evaluation experiment, we train a model to recognize drugs and disorders mentions in a post. The settings of this experiment are the following:

- When preparing the list of pairs <En, He>, we filter the UMLS to only include drugs and disorders.
- We train the contextual filter model for each class—disorders and drugs—using a random forest decision tree classifier on a split of the gold-standard dataset to filter the spans returned by the high-recall matcher. We use a 75%-25% train-test split of the dataset.
- Similar to previous work, we measure entity level metrics (exact match of the span) and token-level metrics (using BIO encoding of spans) and report Precision, Recall and F1 score.

Results are presented in Table 1 (see also Supplementary Appendix Tables S4.2-S4.5). F1 score for exact entity recognition is 0.75 overall, with the less stringent token-level measure reaching 0.78 (accounting for partial match and overlaps). Precision is generally better than recall for all metrics. The source of the low recall is that the list of pairs we have generated does not include the target terms that occur in the posts, and not the contextual filter.

Supplementary Appendix Table S4.6 indicates the proportion of the high-recall candidates filtered out by the contextual filters for each community and each type of term. On average, about half of the candidates are filtered out based on context. In ablation tests, we confirm that F1 score drops significantly (from 0.75 to 0.64) when using the full UMLS instead of selected semantic groups (diseases and drugs). Without contextual filter, F1 score drops from 0.75 to 0.57. The link to UMLS CUI was accurate in 96% of the cases.

Reported F1 performance of English UMLS linkers is in the range of 0.72-0.82 (cTAKES [clinical Text Analysis and Knowledge Extraction System]).[13,27] Our method reaches 0.75 F1 with a small training dataset (about 1000 mentions) of social media text, suffering mainly from low recall. Supplementary Appendix S4 provides detailed error analysis.

Extrinsic evaluation

In this evaluation, we identify all medical terms mentions, without fine-grained classification into semantic groups, according to the following setup:

Table 2. MDTEL UMLS entity linking performance on test data

Community	F1 score	Precision	Recall	ROC AUC	Accuracy: Filter model	Accuracy: Full algorithm	High-recall candidates filtered out
Diabetes	87.6	82.0	94.1	82.1	84.2	98.7	31.7%
Sclerosis	93.8	92.6	95.0	94.1	91.2	98.8	50.9%
Depression	87.5	87.5	87.5	97.9	85.9	99.1	53.8%

AUC: area under the receiver-operating characteristic curve; MDTEL: Medical Deep Transliteration Entity Linking; ROC: receiver-operating characteristic; UMLS: Unified Medical Language System.

- When preparing the list of pairs <En, He>, we filter the UMLS to include types expected in OHCs.
- We train the contextual filter model, using the same random forest decision tree classifier with a 75%-25% train-test split.

Table 2 summarizes our results. Performance is measured on the high-recall list of candidate terms. When considering all tokens in the posts, performance is reported as *accuracy full algorithm*. With accuracy ranging from 98.7% to 99.1% and F1 score between 87.5 and 93.8, the performance of MDTEL on noisy Hebrew text is similar to that obtained by specialized UMLS linkers in English such as cTAKES.[13] About half of the high-recall candidates are filtered out by the contextual relevance classifier, demonstrating the importance of context in identifying term mentions. Increased performance compared with the intrinsic evaluation is explained by (1) a larger number of instances for training and (2) coarser decision making—we do not attempt to identify the semantic type of the mention, only the fact that a medical term is mentioned.

Information retrieval improvement in OHCs with UMLS term normalization

Once term mentions are identified, we normalize them and assess the impact of this normalization on an information retrieval task. For example, given the search query submitted to Camoni ("How do you know I have fibromyalegia?"), we link the transliterated "fibromyalegia" with high similarity to the transliterated term "fibromyalgia" to generate the new query, "How do you know I have fibromyalgia?"

Experimental steps

1. We collect search queries submitted to Camoni from the previous 3 years. We focused on queries that occurred <20 times and are more likely to include infrequent transliteration forms. About 6500 queries are collected for each community.
2. Google Search Baseline: We collect the current search results for each query without any modification on the full Camoni site (using Google Custom Search API).
3. Google Spelling Suggestion Baseline: We use Google Spelling Suggestions API to propose a fixed query and collect search results for the fixed query.
4. UMLS Normalization: We linked the text of the query to UMLS and performed an alteration for each case where similarity between a linked term is high (>0.8) but not perfect (< 1.0). A query may include several medical terms, in which case we generate 1 fixed query for each normalized term, with only 1 medical term altered.
5. Get results of the normalized queries: For each search query, we collect search results using Google Custom Search API. For

Table 3. Quantitative information retrieval improvement using MDTEL

Community	Queries	Queries improved Google Spelling Suggestion	Queries improved MDTEL UMLS Linking
Diabetes	6581	22.7%	45.2%
Sclerosis	6325	22.6%	46.5%
Depression	7302	22.2%	47.3%

MDTEL: Medical Deep Transliteration Entity Linking; UMLS: Unified Medical Language System.

each original query, we select the normalized query returning the most search results.

Analysis of experimental results

We measure the percentage of queries where the normalized queries got more results than the original query. Table 3 compares the recall improvements introduced by UMLS normalization and Google Spelling Suggestion.

MDTEL improves recall for about twice as many queries as the data-driven Google correction method. This difference is explained by the following factors: we focus on medical terms, we consider several alterations for each query, and our process exploits pre-trained data that are not observed frequently in search queries, and hence, for which Google's approach is not likely to find corrections.

We perform manual analysis of the search results with MDTEL UMLS normalization to verify that in the cases with increased search results, MDTEL also improves the relevance of the search results. Given the original and the fixed query Q and Q_{fixed}, and the corresponding search results $Answers(Q)$ and $Answers(Q_{fixed})$, we asked medical students to annotate the following questions:

- Does Q_{fixed} have the same intended meaning as Q?
- Are $Answers(Q_{fixed})$ documents that match the meaning of Q?

In this experiment, we presented as $Answers(Q)$ the title of the top 10 search results received in the website as ranked by Google API: articles or posts titles, without their content.

To account for the cases where the original query had very few answers, we sample 100 cases in which UMLS normalization increased the search count from <5 (there are 1314, 1272, and 1491 such queries for the diabetes, sclerosis, and depression communities, respectively). We ask the question: "Do you find a relevant answer to Q in $Answers(Q_{fixed})$?" We do not show Q_{fixed} to the annotator to avoid introducing bias. Full results are shown in Supplementary Appendix S5. A total of 94% of the answers are positive, indicating that UMLS normalization did not alter the intended meaning of the query (Supplementary Table S5.2).

In a second evaluation, we account for cases where the original query yielded many results (>5). We sampled 100 examples of Q, Answers(Q), Answers(Q_{fixed}), showed them to the annotator as Q, Answers(A), Answers(B), and asked: "Which answers are better?" To avoid bias, the annotator is not told which column corresponds to MDTEL. Full results are shown in Supplementary Table S5.3. In 96% of cases, UMLS normalization either improved or preserved the quality of the search results, with about 36% of the cases improved.

Finally, we ask the annotator to directly assess whether the replacement performed by MDTEL respects the intended meaning of the query. The replacement was found problematic in only 4 cases of 300. Supplementary Appendix S5 provides detailed error analysis.

DISCUSSION

User text normalization at indexing time

In all experiments reported previously, we limit ourselves to normalization of the query at search time. We expect that patient-authored posts also contain misspellings of medical terms in Hebrew. Normalizing the posts through UMLS linking at the time they are posted would increase search efficacy and ease text analytics. Practically, this process is more complex to implement than query alteration because it must be verified by the author of the post or through postediting curation. The potential improvement in search efficacy is assessed in the Supplementary Appendix.

Generalizing to other languages

The MDTEL method is generalizable to other languages with different writing script than English. It relies on several components, we list language-specific requirements and annotation efforts for each step:

1. **Transliteration model:** To train the transliteration model, the only resource needed is an English medical terms list, which can be taken from our reference.[23] It is useful to add domain-specific terms, and to include translated terms in cases in which the translation is more common than the transliteration. The generation of the synthetic dataset to train the transliteration model requires access to an automatic translation API (eg, Google Translate).
2. **High recall matcher**: This component uses the expanded UMLS MRCONSO table to build the list of pairs. The matcher itself requires a morphological analyzer (we use YAP for Hebrew) to lemmatize the text of the post.
3. **Contextual relevance model:** This is a machine learning model, and it needs labeled dataset in order to learn. We experimented with several dataset sizes, and for all experiments, using 500 posts as labeled dataset (25% of those posts as a test set) provided usable performance. The guidelines for annotation must be adapted to each language (eg, addressing morphological and syntactic constructs such as compound words).

Our method relies on the fact that transliterations for medical terms are frequent. This must be validated for each language. A simple test can be done in order to verify this: taking the English medical terms, and measure how often translations are not words in the target language. In Hebrew, we found 37% (28 308 of 77 698) of the resulting translations are not Hebrew words.

CONCLUSION

We present MDTEL, a method for UMLS XEL in OHCs, and assess its value in a named-entity recognition and information retrieval tasks in a popular Israeli community. UMLS XEL is a challenging task in non-English languages, especially for languages with different writing and sound systems such as Chinese, Japanese, Hebrew, and Arabic. We observe that most medical term mentions in patient-authored text are formed by transliteration. Our method exploits advances in neural methods to generate transliterations from source text. We enhance a deep attentional RNN encoder-decoder transliteration model by synthesizing a set of term pairs starting from the UMLS vocabulary in an unsupervised manner.

We also note that the classification of an n-gram as a UMLS medical concept mention depends on its context, both the general topic of the document and the surrounding tokens. We take this context dependence into account and build a relevance classifier for each community in the OHC that accounts for contextual neural language model features when detecting medical term mentions.

On a domain of noisy Hebrew text characterized by very high variability (1.43 average forms for each medical term), MDTEL achieves performance similar to that achieved on English text by leading UMLS linkers on a vocabulary of over 8 million distinct forms. We demonstrate the efficacy of UMLS linking in patient-authored text by analyzing the improvements introduced in the OHC search engine by applying MDTEL on search queries. Both quantitative and manual performance analysis demonstrate the high value of UMLS terms in improving search. Cross-lingual UMLS linking also enables text analytics—we measured the prevalence of medical terms in patient-authored text (about 5%), the distribution of UMLS contexts in forums, which are key features enabling sensemaking and trend detection for community moderators.

AUTHOR CONTRIBUTIONS

YB, RC, EB, ME, and NE worked on study design and algorithms. YB implemented and carried out experiments. TS participated in designing evaluation study design. All coauthors participated in the writing of the manuscript.

SUPPLEMENTARY APPENDIX

Supplementary Appendix is available at *Journal of the American Medical Informatics Association* online.

CONFLICT OF INTEREST

The authors have no competing interests to declare.

REFERENCES

1. Civan A, Pratt W. Threading together patient expertise. *AMIA Annu Symp Proc* 2007; 2007: 140–4.
2. Mamykina L, Nakikj D, Elhadad N. Collective sensemaking in online health forums. In: proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems; 2015: 3217–26.
3. Zhang S, O'Carroll Bantum E, Owen J, Bakken S, Elhadad N. Online cancer communities as informatics intervention for social support:

conceptualization, characterization, and impact. *J Am Med Inform Assoc* 2017; 24 (2): 451–9.

4. Demner Fushman D, Elhadad N. Aspiring to unintended consequences of natural language processing: a review of recent developments in clinical and consumer-generated text processing. *Yearb Med Inform* 2016; 25 (1): 224–33.

5. Gonzalez-Hernandez G, Sarker A, O'Connor K, Savova G. Capturing the patient's perspective: a review of advances in natural language processing of health-related text. *Yearb Med Inform* 2017; 26 (1): 214–27.

6. Neveol A, Dalianis H, Savova G, Zweigenbaum P. Clinical natural language processing in languages other than English. *AMIA Annu Sympos Proc* 2014: 1–3.

7. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res* 2004; 32 (database issue): D267–70.

8. Gertner Institute. "Camoni": Social network for people with dealing with chronic diseases. http://www.gertnerinst.org.il/e/health_society/social_networks/social_networks_Camoni/ Accessed January 2019.

9. Goldberg Y, Elhadad M. Identification of transliterated foreign words in Hebrew script. In Gelbukh A, ed. *Computational Linguistics and Intelligent Text Processing*. New York, NY: Springer; 2008: 466–77.

10. Névéol A, Anderson RN, Cohen KB, *et al*. CLEF eHealth 2017 multilingual information extraction task overview: ICD10 coding of death certificates in English and French. In: Cappellato L, Ferro N, Goeuriot L, Mandl T, eds. CLEF 2017 Working Notes. 2017: 1–3. http://ceur-ws.org/Vol-1866/invited_paper_6.pdf Accessed January 2019.

11. Névéol A, Robert A, Grippo F, *et al*. CLEF eHealth 2018 multilingual information extraction task overview: ICD10 coding of death certificates in French, Hungarian and Italian. In: Cappellato L, Ferro N, Nie J-Y, Soulier L, eds. CLEF 2018 Working Notes; 2018: 1–3. https://pdfs.semanticscholar.org/13db/484cd8cee4d99ab3f7ee9c00e06f3c74e0c1.pdf Accessed January 2019.

12. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp* 2001; 2001: 17–21.

13. Savova GK, Masanz JJ, Ogren PV, *et al*. Mayo clinical text analysis and knowledge extraction system (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc* 2010; 17 (5): 507–13.

14. Roller R, Kittner M, Weissenborn D, Leser U. Cross-lingual candidate search for biomedical concept normalization. *arXiv*: 180501646 Cs; 2018.

15. Xu J, Gan L, Cheng M, Wu Q. Unsupervised medical entity recognition and linking in Chinese online medical text. *J Healthc Eng* 2018; 2018: 2548537.

16. Cotik V, Rodríguez H, Vivaldi J. Arabic medical entity tagging using distant learning in a Multilingual Framework. *J King Saud Univ-Comput Inf Sci* 2017; 29 (2): 204–11.

17. Cohen R, Goldberg Y, Elhadad M. Transliterated pairs acquisition in medical Hebrew. In: proceedings of the Machine Translation and Morphologically-rich Languages Workshop of the Israel Science Foundation; 2011.

18. Cohen R, Elhadad M. Sideways transliteration: how to transliterate multicultural person names? *arXiv: 191112022*; 2019.

19. Chen N, Banchs R, Zhang M, Duan X, Li H. Report of news 2018 named entity transliteration shared task. In: proceedings of the Seventh Named Entities Workshop; 2018: 55–73.

20. Grundkiewicz R, Heafield K. Neural machine translation techniques for named entity transliteration. In: proceedings of the Seventh Named Entities Workshop; 2018: 89–94.

21. Cohen R, Elhadad M, Birk O. Analysis of free online physician advice services. *PLoS One* 2013; 8 (3): e59963.

22. Bodenreider O, McCray AT. Exploring semantic groups through visual approaches. *J Biomed Inform* 2003; 36 (6): 414–32.

23. Dictionary of English medical terms for LibreOffice/OpenOffice/Android/Word. https://github.com/glutanimate/wordlist-medicalterms-en Accessed January 7, 2019.

24. Seker A, More A, Tsarfaty R. Universal Morpho-syntactic Parsing and the Contribution of Lexica: Analyzing the ONLP Lab Submission to the CoNLL 2018 Shared Task. In: proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies; 2018: 208–15.

25. Okazaki N, Tsujii J. Simple and efficient algorithm for approximate dictionary matching. in *proceedings of the 23rd International Conference on Computational Linguistics*. Association for Computational Linguistics; 2010: 851–9.

26. Wikipedia Hebrew Corpus. http://u.cs.biu.ac.il/~yogo/hebwiki/ Accessed January 9, 2019.

27. Denecke K. Extracting medical concepts from medical social media with clinical NLP tools: a qualitative study. In: proceedings of the Fourth Workshop on Building and Evaluation Resources for Health and Biomedical Text Processing; 2014.