Automated Suggestion of Tests for Identifying Likelihood of Adverse Drug Events

Ritwik Banerjee*, Yejin Choi*, Gaurav Piyush*, Ameya Naik* and I. V. Ramakrishnan*
* Department of Computer Science
Stony Brook University
Stony Brook, NY 11794-4400

***Abstract*—Adverse drug events (ADE) caused by use, misuse or sudden discontinuation of medications trigger hospital emergency room visits. Information about a wide range of drugs and associated ADEs is provided in online drug databases in the form of narrative texts. Even though some ADEs can be detected by observable symptoms, several others can only be confirmed by laboratory tests. In this paper, we present a system that provides automated suggestion of tests to identify the likelihood of ADEs. Given a patient's medications and an optional list of signs and symptoms, our system automatically produces the laboratory tests needed to confirm possible ADEs associated with these drugs. The basis of our application is to map clinical symptoms to medical problems and laboratory tests. Towards that, we use template-based extraction and shallow parsing techniques from natural language processing to extract information from narrative texts. We employ relevance ranking measures to establish correspondence between the tests and ADEs. Our evaluation based on a sample set of 40 drugs shows that this system achieves relatively high sensitivity.**

I. Introduction

Adverse Drug Events (ADEs) are unintended and undesired reactions experienced by an individual due to the use, misuse, or discontinuation of medication. Several studies have reported that among the adult population, over 12% of emergency room (ER) visits are caused by ADEs [1]–[3]. Safety and quality of patient healthcare are strengthened when a medical problem caused by a drug is promptly and correctly identified. Evidence of an ADE based on a patients clinical symptoms thus provides an important data point for clinical decision making.

As there exist way too many drugs, physicians cannot be expected to have memorized all possible ADEs associated with them. Often this is not a problem due to the availability of electronic pharmaceutical databases (e.g. [4], [5]). To assess the likelihood of ADEs, physicians manually look up pharmaceutical databases that hold extensive information on drugs including narratives of their adverse effects. This manual lookup and review is a time-consuming process, prone to lapsed vigilance, and often brings about a failure to order appropriate diagnostic tests [6].

Some ADEs can be detected by observable symptoms, but several others can only be confirmed by laboratory tests. And even though laboratory testing is the single highest-volume medical activity driving clinical decision making, the process of ordering diagnostic tests and acting upon them remains vulnerable to errors [7], [8].

In this paper, we propose a clinical decision support (CDS) system which automatically *pushes* laboratory test suggestions

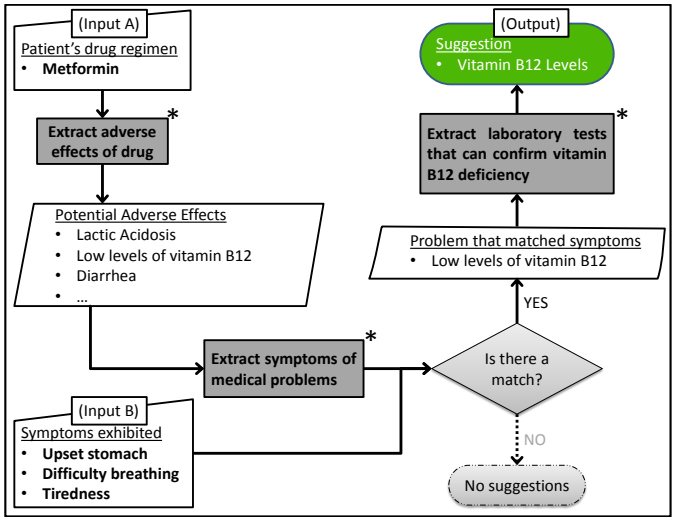

Fig. 1. A use-case scenario: patient taking metformin (Input A) and exhibiting certain symptoms (Input B). Symptoms are matched to a likely adverse effect, and a diagnostic test to confirm this possiblity is suggested (Output). The information extraction steps corresponding to $\mathcal{I}_1$, $\mathcal{I}_2$ and $\mathcal{I}_3$ are marked *.

to confirm (or invalidate) potential adverse effects of a patient's drug regimen. Our application exploits natural language information provided in online pharmaceutical databases. To this end, we use natural language processing (NLP) and template-based techniques to extract three types of information:

- ($\mathcal{I}_1$) Potential adverse effects of a drug.
- ($\mathcal{I}_2$) Observable symptoms associates with medical problems.
- ($\mathcal{I}_3$) Medical problems identified by abnormal values in laboratory test results.

Based on this extracted knowledge, on one hand we map symptoms to medical problems, and the problems to laboratory test results, while on the other hand we map drugs to their adverse effects. Fig. 1 shows a use-case scenario where a patient has been prescribed the drug Metformin, and is experiencing the following symptoms: (a) an upset stomach, (b) difficulty in breathing and (c) overall tiredness.

For information extraction, we use shallow parsing and pattern matching techniques to extract relevant text from available natural language databases. Further, we adapt MetaMap [9], a tool for recognizing medical concepts, to extract medical terms. The three types of information $\mathcal{I}_1$, $\mathcal{I}_2$ and $\mathcal{I}_3$ are obtained by employing these steps on three separate databases, viz. Drugs.com [10], the MedlinePlus encyclopedia [11] and Rush University Medical Center's health encyclopedia [12], respectively.

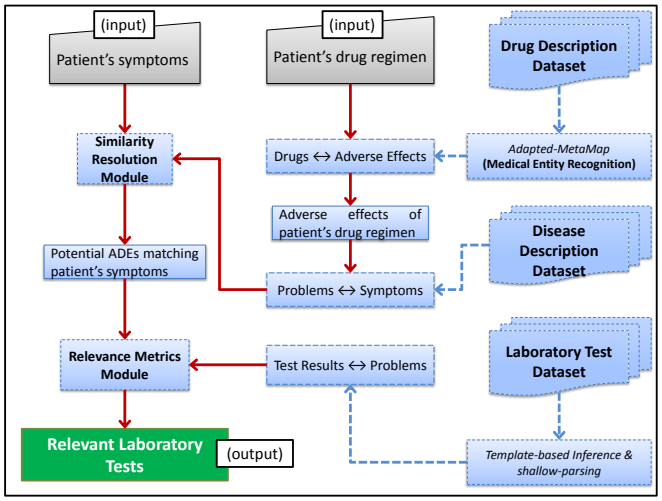

Fig. 2. Overview of the automated diagnostic test suggestion process. Offline datasets, modules and processes are shown in blue dashes. Information corresponding to $\mathcal{I}_1$, $\mathcal{I}_2$ and $\mathcal{I}_3$ are represented as bidirectional maps ($\leftrightarrow$). Input-dependent portion of the process flow is shown in red arrows.

In many cases, the output consists of more than one suggestion. This is due to multiple disorders partially or wholly matching the patient's symptoms, as well as multiple diagnostic tests having the ability to identify a medical problem. Providing a clinician with all such suggestions can lead to alert fatigue [13]. We thus employ a similarity resolution module (SRM) to filter out spurious suggestions, and then compute the relevance of each suggestion to provide a ranked list.

The rest of this paper consists of a discussion of related work in II, followed by an overview of our method in III. The datasets along with the information extraction steps are described in IV. Sections V and VI explain the SRM and the ranking measures before the experiment results are presented in VII.

II. Related Work

The topic of pharmacovigilance, defined as the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other drug-related problem [14], has received substantial research attention in the medical literature (*e.g.* [15], [16]). Computer-aided approaches in this area can be divided into two orthogonal categories: (a) extraction of new ADEs from population data, and (b) detection or identification of already known ADEs in patients.

Extraction of ADEs based on population studies.
This body of work aims to discover previously unknown ADEs from population-based historical clinical data – primarily a pharmacoepidemiological approach. Much of this work is based on mining biomedical research literature [17]–[20]. The main drawback is that learning requires large amounts of annotated training data [18], [21], [22], or else suffers low accuracy [19], [23], [24]. Getting large amounts of annotated text is expensive. We remark that our problem is not the discovery but the identification of known ADEs that are described in narrative form in existing pharmaceutical databases.

Identification of ADEs in patients.
Prior research tackling identification problem has focused on creating ADE alerts in patient data. In one line of work, alerts are formulated as rules that get triggered whenever signs and symptoms in patient data satisfy the conditions attached to these rules. The conditions are built with medical terms in the patient's symptoms and ADE descriptions. These terms are drawn from discharge summaries [25], [26], ADE reports [27], [28] and ambulatory care notes [29]. NLP tools (e.g. MedLEE [30]) have also been used to extract them.

TABLE I. Categories and their UMLS semantic types.

Category	UMLS Semantic Types
Medical Problem	Disease or Syndrome; Sign or Symptom; Body System; Laboratory or Test Result; Mental or Behavioral Dysfunction; Cellular or Molecular Dysfunction; Mental Process; Individual Behavior; Neoplastic Process; Acquired Abnormality; Anatomical Abnormality; Congenital Abnormality
Drug	Organic Chemical; Biologically Active Substance; Pharmacologic Substance; Amino Acid, Peptide or Protein; Steroid; Clinical Drug

Rules vary in complexity. Simple rules look for keywords or simple linguistic patterns [29], [31], [32]. As noted by Murff *et al.* [33], such triggers do not achieve a desirable level of accuracy. Additionally, the rules are learned from a set of documents obtained from a small number of hospitals, often one. In other words the data set from which the rules are derived is far from being comprehensive. Not surprisingly, such systems report low accuracy even when employing more sophisticated NLP tools to extract relevant medical terms [25], [26]. More complex decision rules have been used which combine keywords and laboratory findings [27], [28], [34], [35]. Such rules are, however, manually curated. Notable exceptions include the PSIP project [36], which performs more complex semantic mining. Their approach does not, however, extend to external natural language databases.

Identification using external knowledge sources.
Some recent work has explored the utility of available semi-structured knowledge sources (e.g. RxNorm [37]) to detect and prevent ADEs [38]–[41]. They continue to offer shallow coverage, though, because of a focus on very specific drugs (e.g. [38], [41]) or because they continue to build rules manually (e.g. [39]). More importantly, this body of work does not exploit laboratory test data.

In summary, despite substantial research in ADE detection methods, gaps remain. Our approach attempts to fill this gap with three salient characteristics:

(1) Extensive utilization of available pharmaceutical databases.
(2) Use of automated measures that provide coverage over a large set of drugs.
(3) Combines diagnostic test information with ADE information to aid CDS.

III. Method Overview

We use three separate medical knowledge repositories to automatically suggest diagnostic tests that identify potential ADEs. The first, a drug description dataset, is used to extract

TABLE II. MEDICAL ENTITY EXTRACTION

"Subnormal vitamin B12 levels have been reported, and may result in anemia or neuropathy."

Medical Entity	Semantic Category	Score (0–1000)
Subnormal vitamin B12 levels	Laboratory or Test Result	906
Anemia	Disease or Syndrome	746
Neuropathy	Disease or Syndrome	1000

the adverse effects of drugs. The second, the laboratory test dataset, is used to obtain the medical problems indicated by abnormal test results. The third dataset contains descriptions of various medical problems. We use it to map diseases and disorders to their observable symptoms. The overview of our approach is presented in Fig. 2.

Information extraction from these datasets involves using the medical entity extraction tool MetaMap. It extracts medical terms from natural language data and assigns them semantic types defined by the Unified Medical Language System (UMLS) [42]. For the purposes of this work, we combine several semantic types into two categories, *medical problems* and *drugs*, as shown in Table I. The distinction between our semantic categorization and that followed by MetaMap is further explained in Sec. IV-A.

Given a patient's list of medications and an optional list of exhibited symptoms, our application identifies a list of medical problems $\{p_1, \ldots, p_k\}$ fitting two criteria:

- Each p_i is a potential adverse effect of at least one of the drugs being taken by the patient.
- The symptoms of each p_i match at least some of these exhibited symptoms.

For the second criterion, i.e. matching symptoms, synonyms are resolved so that equivalent symptoms are identified. For example, if the disease description dataset provides "fatigue" as a symptom of an ADE while the input symptoms contain "tiredness", the *synonym resolution module* enables identifying them as equivalent. Sec. V provides further details.

Finally, we perform a relevance ranking of laboratory tests based on the similarity between (a) medical problems identified by test result, and (b) the list of problems $\{p_1, \ldots, p_k\}$ obtained above.

IV. INFORMATION EXTRACTION

In the scope of this work, we extract information to map drugs to their adverse effects, medical problems to their observable symptoms and laboratory tests to the medical problems they identify. These mappings are obtained by extracting information from the drug description dataset, the disease description dataset, and the laboratory test dataset, respectively.

A. The Drug Description Dataset

We extract information about drugs from Drugs.com, which consists of $5,856$ drugs. Like the previous dataset, this too is semi-structured, i.e. different types of information are pre-labeled (e.g. sections labeled "adverse effects", "dosage", etc). This corpus provides medical knowledge in complex descriptive paragraphs. We adapt MetaMap to extract and categorize relevant information.

TABLE III. SEMI-STRUCTURED INFORMATION

Laboratory test dataset (Vitamin B12 Level Test)	**Disease description dataset** (Vitamin B12 deficiency)
Vitamin B12 Level Definition ... Normal Values ... Values of less than 200 pg/mL are a sign of a vitamin B12 deficiency. Causes of vitamin B12 deficiency include diseases that cause malabsorption (*e.g.* Celiac diseases and Crohn's disease), not enough vitamin B12 in diet, ... Conditions that can increase vitamin B12 levels include liver disease, ...	Symptoms can include: • Diarrhea or constipation • Fatigue, lack of energy • Light-headedness when standing up • Loss of appetite • Pale skin • Problems concentrating • Shortness of breath, mostly during exercise • Swollen, red tongue or bleeding gums

∗Linguistic cues of template-based inference are underlined.

Medical Entity Recognition (MER) is the first step of our process. A "medical entity" is a given instance of a medical concept or category (e.g. *Metformin* is an instance of a *pharmaceutical substance*). As shown in Table II, recognizing such entities consists of (a) detecting the relevant terms and (b) identifying the semantic category. In this section, we first describe how we perform MER from complex narratives, and then present the shallow parsing and template-based extraction technique we use for simpler texts.

1) The Adapted-MetaMap method for MER: In the scope of this work, we use the UMLS semantic types[1] to categorize two types of medical entities: *drugs* and *medical problems*.

To extract medical entities from narratives, we use MetaMap, which splits a text into sentences and phrases and subsequently identifies medical terms. Moreover, it maps these terms to UMLS semantic categories, together with associated scores. These available categories, however, do not suffice for our proposed application. Table II presents a typical example: MetaMap identifies "subnormal vitamin B12 levels" as a test result instead of a potential adverse event. To enhance the precision of MetaMap for our application, we make the following revisions to it:

- Filter candidates based on empirically determined threshold on the score.
- Filter candidates based on additional restrictions on semantic types. E.g. the semantic type "Element, Ion, or Isotope" included by MetaMap in the category "Chemicals and Drugs" is excluded from our "Drug" category).
- Include some additional semantic types. E.g. the semantic type "Laboratory or Test Result" included by MetaMap in the category "Phenomena" is included by us in the "Medical Problem" category. The complete list of categories and their constituent UMLS semantic types are provided in Table I.

In the remainder of this paper, MetaMap-based MER revised as above is called *adapted-MetaMap*.

B. The Laboratory Test Dataset

For extracting information about laboratory tests and procedures, we use the Rush University Medical Center's health en-

[1]UMLS contains a semantic network which organizes a meta-thesaurus vocabulary of over two million medical concepts into 135 semantic types.

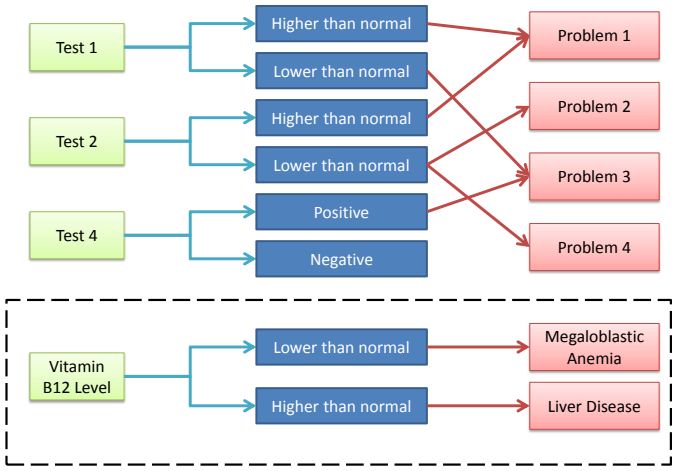

Fig. 3. Mapping abnormal laboratory test results to medical problems.

cyclopedia. It comprises of 603 tests, where information about each test is provided in semi-structured narratives. Table. III shows a typical semi-structured template.

1) Shallow Parsing and Template-based Inference: Some information is presented in relatively simple form. These include short sentences that serve as data labels, i.e. sentences from which the *type* of information can be inferred. Table IV illustrates this inference process with the vitamin B12 level test. Many test results, for example, have two types of abnormal results: (a) results above the reference range, and (b) results below it. Clauses of the form "causes of ... deficiency include ..." allows our application to relate the subsequently mentioned medical entities to the correct type of test result.

TABLE IV. TEMPLATE-BASED INFORMATION INFERENCE.

Text snippet	*Causes of vitamin B12 deficiency include diseases that cause malabsorption (for example, Celiac disease and Crohn's disease), ...*
Inference	Subsequent text lists medical problems resulting from vitamin B12 deficiency.

As most natural language data sources (e.g. an online encyclopedia) use similar snippets across most documents, we were able to infer the type of information by using a combination of pattern matching and phrase chunking. The location of the relevant snippets were identified by studying HTML templates of our data sources.

2) Mapping Test Results to Medical Problems: Using the shallow-parsing and template-based inference, we identify the types of abnormal test results (e.g. lower/higher than normal, positive/negative). Further, we use adapted-MetaMap to extract the medical problems associated with such abnormal results. This way, we obtain a directed tripartite graph which maps each abnormal test result to a medical problem (Cf. Fig. 3).

C. The Disease Description Dataset

In order to determine the observable symptoms associated with medical problems, we make use of the National Library of Medicine's MedlinePlus encyclopedia [11]. Among other things, MedlinePlus contains semi-structured data about the symptoms of various diseases and disorders. To extract these

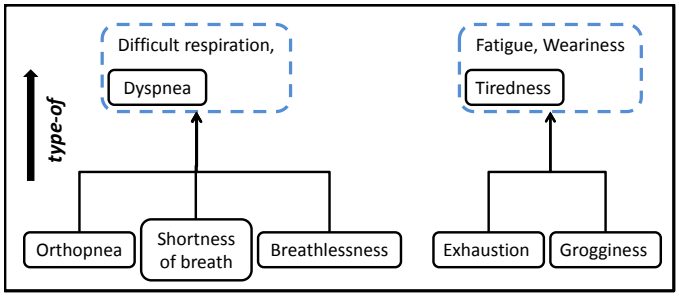

Fig. 4. Terms identified as *similar* by the similarity resolution module when the path length $l = 1$. The outer dashed boxes denote synonym sets (i.e. $l = 0$). The member of synonym sets found in our data and its hyponyms (i.e. the *type-of* relation) are shown in solid boxes.

symptoms, the template-based inference methods described earlier suffice. A typical example is illustrated in Table. III.

V. SIMILARITY RESOLUTION

Collating information from multiple sources has its advantages, but natural language data often exhibits great variety. During experimentation, we often found instances where identical or very similar symptoms were expressed in different words or phrases. Furthermore, it was observed that ignoring such variations had a negative effect on identification of relevant ADEs.

To tackle this issue at the simplest level, we resort to semantic similarity measures based on WordNet [43], which is able to directly identify several synonymous terms (e.g. "fatigue" and "tiredness"). Identifying the semantic similarity in a more general sense (*i.e.* going beyond synonyms), however, requires a more work. Many symptoms are expressed in words or phrases that are used interchangeably. Further, many observable symptoms are expressed in broader semantics than the actual symptom. For example, a patient suffering from a "histamine headache" may have her symptom listed only as "headache". Such closely related expressions share a *type-of* relationship. If expression e_1 is a type of e_2, it is said that e_1 is a *hyponym* of e_2, and e_2 is a *hypernym* of e_1. WordNet stores hypernymy and hyponymy relations in a directed acyclic graph structure. The further away two words are from each other in this graph, the less similar they are. In other words, the path length l between a term t and its hypernym/hyponym t' is inversely related to their semantic similarity $s(t, t')$, *i.e.* $l \propto 1/s(t, t')$.

For our purposes, we identify two words as synonymous if they are semantically similar *enough*. To determine this similarity, we collect symptom terms together with their synonyms as well as their hyponyms and hypernyms. We run several experiments with the path length varying from 0 (i.e. considering only synonyms) to 3. Fig. 4 illustrates the kind of similar terms encountered for path lengths $l = 0, 1$.

VI. RELEVANCE RANKING

The final stage of our application involves ranking several potentially useful laboratory tests according to their relevance to a given set of medical problems. In order to do this, we run adapted-MetaMap on the test descriptions, which are sections

TABLE V. SUGGESTIONS WITHOUT ADDITIONAL SYMPTOMS

Evaluation Set		Sensitivity	Precision
All Drugs	All Symptoms	0.87	0.82
Nonpsychotropic Drugs	All Symptoms	0.86	0.85
Nonpsychotropic Drugs	*minus symptoms associated with the nervous system*	**0.89**	**0.86**

explaining what problems the test might detect. This process yields a set of medical problems P_t for each test, viz. the diseases or disorders confirmed by abnormal values. The same process on the adverse effects of a patient's list of medications yields another set of medical problems P_m. The tests are ranked in descending order of the Jaccard similarity coefficient, defined as the size of the intersection divided by the size of the union of the sample sets

$$J(P_t, P_m) = \frac{|P_t \cap P_m|}{|P_t \cup P_m|} \tag{1}$$

VII. EVALUATION

Due to the lack of available gold standard labels, we manually evaluate our application on a set of 40 drugs obtained from the list of top 100 (by sale) drugs on Drugs.com. Of these, 10 were psychotropic medications. In this section, we present the results of two experiments: (a) without any symptoms being provided, and (b) with a list of symptoms.

A. Suggestions based on patient's drug regimen only

When suggesting tests without being provided any of the patient's symptoms, our application is unaware of the prior likelihood of any ADE. All potential adverse effects are thus treated as equally probable. Therefore, instead of ranking by relevance, we check whether a test suggested by our application has the ability to identify *any* potential adverse effect associated with the given drug. A total of 226 tests were suggested for the 40 drugs, of which 186 were deemed relevant by a manual verification process. Additionally, 28 laboratory tests deemed relevant by human judges were not suggested by our system.

An analysis of the results showed that in general, test suggestions to identify potential ADEs associated with psychotropic medications performed poorly. The worst results obtained for the drug Adderall, where only 2 of the 11 suggested tests were judged as relevant. For other drugs, our analysis revealed that suggestions for adverse effects involving the nervous system were comparatively less accurate. Table V presents the results of this evaluation.

B. Suggestions based on patient's drug regimen and symptoms

Our second experiment takes into account the scenario where in addition to the list of medications, we also know a few symptoms exhibited by the patient. We test with each drug twice by providing two sets of symptoms s_1 and s_2 corresponding to different medical problems. For instance, *Metformin* medication was tested with the two sets {nausea, vomiting, low blood pressure, abdominal pain} and {fatigue, dizziness, dyspnea, diarrhea}, corresponding to the potential ADEs *lactic acidosis* and *vitamin B12 deficiency*, respectively. We thus obtain 80 test inputs for the 40 drugs. For each input, the output is a list of suggested laboratory tests ranked by the Jaccard similarity coefficient as described in Eq. 1.

TABLE VI. MEAN RECIPROCAL RANK OF SUGGESTIONS.

Evaluation Set	Mean Reciprocal Rank				
	no SRM	SRM			
		$l=0$	$l=1$	$l=2$	$l=3$
All Drugs	0.78	0.83	**0.88**	0.76	0.73
Nonpsychotropic Drugs	0.80	0.86	**0.89**	0.79	0.74
Nonpsychotropic Drugs (*minus symptoms associated with the nervous system*)	0.80	0.85	**0.93**	0.78	0.74

We also experiment with and without the similarity resolution module (SRM). Results show that resolving similarity leads to a significant improvement in performance when hyponymy/hypernymy is restricted to short paths. Longer paths simply result in more tests being suggested, even if they are for ADEs with quite different symptoms.

Our application provides a list of laboratory tests that can confirm (or invalidate) a medical condition associated with the patient's symptoms provided that the condition is also a known potential ADE associated with the patient's drug regimen. The output list is ranked by the relevance of the test in identifying a potential ADE associated with the given set of drugs and symptoms. To evaluate such a list, we use the *mean reciprocal rank* (MRR) [44], a widely used metric in information retrieval (*e.g.* [45], [46]) to evaluate systems that provide ranked results instead of a single answer. It is the average of the reciprocal ranks of the output list

$$MRR = \frac{1}{|S_i|} \sum_{k=1}^{|S_i|} \frac{1}{\text{rank}_k} \tag{2}$$

where S_i is the input set of drugs and symptoms, and rank_k is the rank of the correct suggestion for the k^{th} test input. The preformance measurements are presented in Table VI.

VIII. CONCLUSION

We presented an approach for automatically suggesting diagnostic tests to identify potential ADEs. We do this by extracting information from available datasets about adverse drug effects, laboratory tests and symptoms of various medical problems. We present experiments that suggest diagnostic tests with and without relevance ranking. Further, we incorporate methods involving hyponymy/hypernymy to resolve semantically similar terms and demonstrate that resolving natural language similarities boosts performance.

Our future work in this direction intends to incorporate multiple corpora for each type of information extraction to improve coverage and to extend the current application to identify adverse effects arising from drug-drug interactions. We also intend to extend the similarity resolution work to account for medical paraphrases.

ACKNOWLEDGMENT

The authors would like to thank Dr. Mark C. Henry, Professor and Chair of the Department of Emergency Medicine at the Stony Brook School of Medicine for his insights and helpful comments.

REFERENCES

[1] A.Capuano *et al.*, "Adverse drug events in two emergency departments in Naples, Italy: an observational study," *Pharmacol. Res.*, vol. 50, no. 6, pp. 631–636, Dec 2004.

[2] G.Trifiro *et al.*, "Adverse drug events in emergency department population: a prospective Italian study," *Pharmacoepidemiol Drug Saf*, vol. 14, no. 5, pp. 333–340, May 2005.

[3] P. J.Zed *et al.*, "Incidence, severity and preventability of medication-related visits to the emergency department: a prospective study," *CMAJ*, vol. 178, no. 12, pp. 1563–1569, Jun 2008.

[4] "Lexicomp." [Online]. Available: http://www.lexi.com/

[5] "Micromedex." [Online]. Available: http://www.micromedex.com/

[6] T. K.Gandhi *et al.*, "Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims," *Ann. Intern. Med.*, vol. 145, no. 7, pp. 488–496, Oct 2006.

[7] H.Singh, "Diagnostic errors: moving beyond 'no respect' and getting ready for prime time," *BMJ Qual Saf*, vol. 22, no. 10, pp. 789–792, Oct 2013.

[8] M.Zhi, E. L.Ding, J.Theisen-Toupal, J.Whelan, and R.Arnaout, "The landscape of inappropriate laboratory testing: a 15-year meta-analysis," *PLoS ONE*, vol. 8, no. 11, p. e78962, 2013.

[9] A. R.Aronson, "Effective mapping of biomedical text to the umls metathesaurus: the metamap program." in *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 2001, p. 17.

[10] "Drugs.com." [Online]. Available: http://health.rush.edu/HealthInformation

[11] Bethesda, "Medlineplus, National Library of Medicine (US)." [Online]. Available: http://www.nlm.nih.gov/medlineplus/

[12] "Rush university medical center's health encyclopaedia." [Online]. Available: http://health.rush.edu/HealthInformation

[13] R.Koppel *et al.*, "Identifying and quantifying medication errors: evaluation of rapidly discontinued medication orders submitted to a computerized physician order entry system," *J Am Med Inform Assoc*, vol. 15, no. 4, pp. 461–465, 2008.

[14] World Health Organization, "The safety of medicines in public health programmes: pharmacovigilance, an essential tool," 2006.

[15] R.Beuscart, W.Hackl, and C.Nøhr, *Detection and prevention of adverse drug events: information technologies and human factors*. IOS Press, 2009, vol. 148.

[16] C.Friedman, "Discovering novel adverse drug events using natural language processing and mining of the electronic health record," in *Artificial Intelligence in Medicine*. Springer, 2009, pp. 1–5.

[17] T.Agbabiaka, J.Savovic, and E.Ernst, "Methods for causality assessment of adverse drug reactions: a systematic review," *Drug Saf*, vol. 31, no. 1, pp. 21–37, 2008.

[18] H.Gurulingappa, A.Mateen-Rajput, L.Toldo *et al.*, "Extraction of potential adverse drug events from medical case reports," *Journal of biomedical semantics*, vol. 3, no. 1, p. 15, 2012.

[19] H.Théophile *et al.*, "An updated method improved the assessment of adverse drug reaction in routine pharmacovigilance," *Journal of clinical epidemiology*, vol. 65, no. 10, pp. 1069–1077, 2012.

[20] W.Wang *et al.*, "A drug-adverse event extraction algorithm to support pharmacovigilance knowledge mining from PubMed citations," *AMIA Annu Symp Proc*, vol. 2011, pp. 1464–1470, 2011.

[21] I.Segura-Bedmar, P.Martnez, and D.Sánchez-Cisneros, "The 1st DDI Extraction-2011 challenge task: Extraction of Drug-Drug Interactions from biomedical texts," in *Proc. DDI Extraction-2011 challenge task*, 2011, pp. 1–9.

[22] I.Segura-Bedmar, P.Martinez, and C.dePablo-Sanchez, "Using a shallow linguistic kernel for drug-drug interaction extraction," *J Biomed Inform*, vol. 44, no. 5, pp. 789–804, Oct 2011.

[23] K.Shetty and S. R.Dalal, "Using information mining of the medical literature to improve drug safety," *J. Am. Med. Inform. Assoc.*, vol. 18, pp. 668–674, May 2011.

[24] B.Percha, Y.Garten, and R. B.Altman, "Discovery and explanation of drug-drug interactions via text mining," *Pac Symp Biocomput*, pp. 410–421, 2012.

[25] G. B.Melton and G.Hripcsak, "Automated detection of adverse events using natural language processing of discharge summaries," *J Am Med Inform Assoc*, vol. 12, no. 4, pp. 448–457, 2005.

[26] X.Wang, G.Hripcsak, M.Markatou, and C.Friedman, "Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study," *J Am Med Inform Assoc*, vol. 16, no. 3, pp. 328–337, 2009.

[27] T.Botsis, M. D.Nguyen, E. J.Woo, M.Markatou, and R.Ball, "Text mining for the vaccine adverse event reporting system: medical text classification using informative feature selection," *J. Am. Med. Inform. Assoc.*, vol. 18, pp. 631–638, Jun 2011.

[28] N. P.Tatonetti, G. H.Fernald, and R. B.Altman, "A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports," *J. Am. Med. Inform. Assoc.*, vol. 19, no. 1, pp. 79–85, June 2011.

[29] M. N.Cantor, H. J.Feldman, and M. M.Triola, "Using trigger phrases to detect adverse drug reactions in ambulatory care notes," *Qual. Saf. Health Care*, vol. 16, no. 2, pp. 132–134, Apr 2007.

[30] C.Friedman, "A broad-coverage natural language processing system." in *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 2000, p. 270.

[31] P. J.Haug, R. M.Gardner, R. S.Evans, B. H.Rocha, and R. A.Rocha, "Clinical decision support at Intermountain Healthcare," in *Clinical Decision Support Systems, Theory and Practice*, E. S.Berner, Ed. New York: Springer, 2007, vol. 2, pp. 159–189.

[32] B.Honigman *et al.*, "Using computerized data to identify adverse drug events in outpatients," *J Am Med Inform Assoc*, vol. 8, no. 3, pp. 254–266, 2001.

[33] H. J.Murff, V. L.Patel, G.Hripcsak, and D. W.Bates, "Detecting adverse events for patient safety research: a review of current methodologies," *J Biomed Inform*, vol. 36, no. 1-2, pp. 131–143, 2003.

[34] A. C.Seger *et al.*, "Development of a computerized adverse drug event (ade) monitor in the outpatient setting," DTIC Document, Tech. Rep., 2005.

[35] T. K.Gandhi *et al.*, "Outpatient adverse drug events identified by screening electronic health records," *J Patient Saf*, vol. 6, no. 2, pp. 91–96, Jun 2010.

[36] R.Beuscart, P.McNair, J.Brender *et al.*, "Patient safety through intelligent procedures in medication: the psip project," *Stud Health Technol Inform*, vol. 148, pp. 6–13, 2009.

[37] "Rxnorm." [Online]. Available: http://www.nlm.nih.gov/research/umls/rxnorm/

[38] E.Izquierdo-Garcia and I.Escobar-Rodriguez, "[Systematic review of new protease inhibitors interactions: telaprevir and boceprevir]," *Farm Hosp*, vol. 36, no. 6, pp. 469–482, Nov 2012.

[39] P. L.Smithburger, S. L.Kane-Gill, and A. L.Seybert, "Drug-drug interactions in the medical intensive care unit: an assessment of frequency, severity and the medications involved," *Int J Pharm Pract*, vol. 20, no. 6, pp. 402–408, Dec 2012.

[40] A.Wright *et al.*, "Preventability of adverse drug events involving multiple drugs using publicly available clinical decision support tools," *Am J Health Syst Pharm*, vol. 69, no. 3, pp. 221–227, Feb 2012.

[41] H. H.Tsai, H. W.Lin, Y. H.Lu, Y. L.Chen, and G. B.Mahady, "A review of potential harmful interactions between anticoagulant/antiplatelet agents and Chinese herbal medicines," *PLoS ONE*, vol. 8, no. 5, p. e64255, 2013.

[42] D.Lindberg, B.Humphreys, and A.McCray, "The unified medical language system." *Methods of information in medicine*, vol. 32, no. 4, pp. 281–291, 1993.

[43] C.Fellbaum, Ed., *WordNet: An Electronic Lexical Database*. MIT Press, 1998.

[44] E. M.Voorhees *et al.*, "The trec-8 question answering track report." in *TREC*, vol. 99, 1999, pp. 77–82.

[45] R.Baeza-Yates, B.Ribeiro-Neto *et al.*, *Modern information retrieval*. ACM press New York, 1999, vol. 463.

[46] L.Hirschman, J. C.Park, J.Tsujii, L.Wong, and C. H.Wu, "Accomplishments and challenges in literature data mining for biology," *Bioinformatics*, vol. 18, no. 12, pp. 1553–1561, 2002.