EXPRESSION OF LABORATORY EXAMINATION RESULTS IN MEDICAL LITERATURE

Takashi Okumura, Eiji Aramaki, Yuka Tateisi

National Institute of Public Health, Kyoto University, National Institute of Informatics
Japan
taka@niph.go.jp, eiji.aramaki@design.kyoto-u.ac.jp, yucca@nii.ac.jp

Abstract

Medical literature contains expressions of laboratory examination results, which are invaluable knowledge sources for building a disease knowledge base that covers even rare diseases. In this study, we analyzed such expressions of disease descriptions in open databases with manually built dictionaries and obtained the following results. First, we identified two major types of expressions for laboratory examination results, that with and without their test names in the expressions. Second, the study identified evaluative expressions that frequently appear in the description of the results. Third, presence of test names and evaluative expressions could classify the expressions into four major classes that demand independent strategies to interpret. The study illustrated that this is beyond the scope of the existing corpora in this domain mostly designed for medical records. Although the analysis is based on rudimentary statistics, it clarified the factors necessary for future corpus designs to promote further research.

Keywords: Laboratory examination, Medical literature, Corpus design

1. Introduction

Clinical Decision Support Systems necessitate disease knowledge bases that describe the relationship between disease and their associated signs and symptoms, as well as typical laboratory results. Because manual compilation of such a knowledge base requires considerable efforts, using Natural Language Processing (NLP) is a reasonable approach. In this regard, medical NLP tools, such as MetaMap (Aronson, 2001), cTakes (Savova et al., 2010), and MedLEE (Friedman et al., 1995), should contribute to automating the compilation process to some extent by extracting the expression of **clinical findings** in the descriptions of target diseases.

On the other hand, the interpretations of **laboratory examination results**, particularly those of rare diseases, are available only in review articles and case reports, and a disease knowledge base demands the processing of expression of laboratory examination results in such medical literature. However, studies have mainly targeted the extraction of clinical findings from a large number of clinical documents, and only a few studies (Zhang and Patrick, 2006) have attempted to interpret expression of laboratory findings in medical literature, which are readily available in a machine readable format on hospital information systems nowadays (McDonald et al., 2003).

Accordingly, this paper analyzes expressions of laboratory examination results to clarify the factors for future corpus design to facilitate further research activities. To this end, Section 2. outlines laboratory examinations and their result expressions, and presents a hypothetical classification model with four major categories. For verification of the model, Section 3. develops simple dictionaries and reports on the statistics of the public disease databases OMIM (John Hopkins University, 1987) and Orphanet (INSERM SC11, 1997). Section 4. discusses the results and the considerations required for future corpus design. Section 5. reviews previous studies that are highly relevant to laboratory examination results and Section 6. concludes the paper.

2. Expression of laboratory results

2.1. Overview of laboratory examinations

Laboratory examinations are performed at medical institutions and involve analyzing samples collected from patients, with the aim of measuring some property of the sample. For example, the white blood cell (WBC) count is an examination performed using a blood sample to determine the number of WBCs in a unit volume. Laboratory examinations are categorized according to the samples used, such as blood, urine, and stool. They are also categorized according to the method of analysis, such as hematological, immunological, physiological, and microscopical means, each of which has certain characteristic style for their result descriptions. To illustrate the characteristics of their results, the examinations are classified by the type of results (Table 1), analyzing a catalog of laboratory examinations available in Japanese clinical settings (Fumimaro Takaku (ed.), 2009). As shown, examinations with numerical results were dominant in the catalog (40.8%), and expression of their results is represented simply using numbers or adjectives such as *high*, *normal*, and *low*. The second most abundant in the ranking is examinations yielding only high/low results, without numbers (24.7%), followed by those with positive/negative results (19.9%). The remaining examinations, such as microscopical examinations and blood typing yield qualitative results.

Result Category	# of records	
Numerical	334	(40.8%)
High/Low	202	(24.7%)
Positive/Negative	163	(19.9%)
Load/Function test	29	(3.5%)
Typing/Genetic test	21	(2.6%)
Fraction/Isozymes	18	(2.2%)
Microscopical findings	14	(1.7%)
Others	37	(4.5%)
Total	818	(100.0%)

Table 1: Classification of examination type

		Evaluative expression	
		+	−
Examination name	+	Quantitative examinations (E1,2,3,4,5,6,7)	Qualitative examinations
	−	Descriptive expression (E8,9,10,11,12)	Nominal expression (E13)

Table 2: Classification of the expressions for examination results

2.2. Expression of laboratory examination results

Laboratory examination results in medical literature are mostly appear in free format. To analyze the expressions, we randomly selected 20 entries from OMIM (John Hopkins University, 1987) that included "syndrome" in the title and had 450 as the last three digits of the record (Okumura et al., 2013), and found 13 expressions of laboratory examination results in the annotated descriptions, which were classified as follows.

Results with examination name The simplest form of laboratory result expression found was that with an examination name, which can be formalized as either "evaluative adjective + examination" or "examination + copula + adjective", as listed below. Note that characteristic adjectives, such as *elevated*, *decreased*, and *normal*, were found in the expressions.

1. *elevated mean pulmonary artery pressure*
2. *elevated liver function enzymes*
3. *elevated transaminases*
4. *moderately elevated alkaline phosphatase*
5. *normal esophageal manometry findings*
6. *SGPT was chronically elevated*
7. *hCG stimulation tests showed normal testosterone responses*

Results without explicit name The second category of result expressions are those without explicit examination names. For example, the results of loading tests can be expressed in a descriptive manner, without their examination names. The names are omitted in such expressions, probably because the findings and the responses are more essential than their formal names in the context and domain experts can easily infer the actual name if needed. Note that, in the examples below, evaluative adjectives are found, such as *low*, *effective*, and *insensitive*.

8. *deletion of part of the short arm of chromosome 5*
9. *lymphocyte infiltration of the CNS*
10. *low to low-normal gonadotropin (152760) responses*
11. *Administration of testosterone enanthate was effective*
12. *insensitive to growth hormone*

Nominal expression of examination results Lastly, we found unique expression for a laboratory result that does not contain an examination name, evaluative adjective, nor qualitative results.

13. *hypercholesterolemia*

Hypercholesterolemia implies that there was a sampling of blood from the patient and the cholesterol level was high in the biochemical examination result. In this expression, the name of the laboratory examination and its result are naturally encoded into the nominal form. Although this is an established expression of a certain clinical finding, it is a possible form of examination result that must be detected for disease knowledge bases that include laboratory examination results, because clinicians may perceive it also as increased cholesterol. Such expressions include *hyperkalemia*, *hyperglycemia*, *viremia*, *bacteremia*, *acidemia*, *acidosis*, *alkalosis*, and *alkalemia*, all of which correspond to certain examination results.

2.3. Classification of laboratory examination results and technical challenges

As illustrated, there are various ways of expressing laboratory examination results, and notably, there are expressions of examination results without the name of the examination. This observation suggests that expressions of laboratory examination results can be classified into two major categories, that with and without examination names. Further, some of the expressions contained evaluative adjectives and others did not, a distinction that can also be utilized for classification. The identified axes can categorize the expression into four types, as illustrated in Table 2. "En" in the cells refers to the 13 expressions.

Unfortunately, the examples lack an expression with an examination name that depicts results in a qualitative manner. However, there are various cases in clinical medicine that fall in this category; for example, result expressions for urinary sediment analysis contain various *casts* with independent names. More importantly, the table suggests the existence of expressions without examination names. An illustration is a descriptive expression, *insensitive to growth hormone*, which suggests a negative examination result of the growth hormone stimulation test. The expression describes actual examination in an operational manner, presupposing scientific domain knowledge on the part of the intended readers. Although the substance name can be used to restore the name in this case, there are examinations with *proper name*, where no such hint is available in the result expression. For example, "fragility of red blood cells" for a patient of paroxysmal nocturnal hemoglobinuria (PNH)

Normal	Abnormal	Increase	Decrease	Positive	Negative	Sufficiency	Deficiency
normal sensitive effective	abnormal insensitive ineffective defective	increase increased increasing elevated elevation elevating rise risen rising	decrease decreased decreasing declined decline declining lowering diminished diminishing reduced	positive above high higher	negative below low lower	sufficiency sufficient excess excesses excessive	deficiency deficient insufficiency insufficient
3 items	4 items	9 items	10 items	4 items	4 items	5 items	4 items

Table 3: Simple dictionary for evaluative expressions

suggests a positive result for the *Ham test*. Because clinicians may perceive it in either way, there is a challenge for medical NLP to recognize the expressions of laboratory examination results as such.

3. Statistical analysis of result expressions

The 2x2 table is a hypothetical model, inductively built only with a limited number of samples. However, to the best of our knowledge, there exists no corpus that annotates expression of laboratory examination results with appropriate information to validate the model. Accordingly, as a preliminary study, this section attempts to estimate the frequency of each category, by building simple dictionaries for expressions of laboratory examination results. Because each type of expression in Table 2 requires an independent strategy to interpret them, the breakdown of the expressions would also contribute to guide future research efforts.

3.1. Dictionary for evaluative expressions

First, to extract expressions of laboratory examination results, a list of evaluative adjectives found in such expressions was built. To this end, we identified eight categories and 43 items in total, as shown in Table 3. Because expressions of examination results may be expressed in various forms, such as a noun phrase ("*elevation of transaminases*") or as a sentence ("*transaminases were elevated*"), a stemmer may be used in preprocessing for more efficient matching. However, for the simplicity of statistics, the terms were manually nominalized, verbalized, and paraphrased in advance.

3.2. Dictionary for examination names

Development of a dictionary for laboratory examination names may seem easy, as it can be done simply by extracting the examination names from an examination catalog, or a standardized terminology (International Health Terminology Standard Development Organisation, 1999). However, such a dictionary would be of little practical use for the detection of the expressions.

For example, an examination, called "human leukocyte antigen typing", is frequently used in medical literature to denote a type of disorder with genetic background. However, it appears as "HLA typing" and "HLA findings". It is possible to avoid the complication by adding the acronym, HLA, to the dictionary. However, because there are ambiguous acronyms, such as CSF, which is indicative of colony stimulating factor (CSF) or cerebrospinal fluid (CSF), such a simple strategy may not work as expected.

Result Category	# of records	# of synonyms	
Numerical	334	532	(1.6)
High/Low	202	368	(1.8)
Positive/Negative	163	391	(2.4)
Load/Function test	29	52	(1.8)
Typing/Genetic test	21	32	(1.5)
Fraction/Isozymes	18	40	(2.2)
Microscopical findings	14	22	(1.6)
Others	37	60	(1.6)
Total	818	1497	(1.8)

Table 4: Overview of the examination name dictionary

Another example is "creatine kinase isozymes" used to differentiate certain diseases, for which an examination to measure their levels is frequently used in clinical settings. However, there are a variety of descriptions found in the OMIM database (John Hopkins University, 1987) that refer to the result of the common examination. In the example below, BB, MB, and MM are the isozymes (types) of creatine kinase (CK).

- *The dimeric creatine kinase isozymes are involved in maintaining intracellular ATP levels, particularly in tissues that have high energy demands.*
- *The creatine kinase MM isozyme is found exclusively in striated muscle; the BB isozyme is found in smooth muscle, brain, and nerve; CKMB is found in human heart.*
- *demonstrated marked elevation of BB isozyme fraction of serum creatine kinase for male sibs with this disorder.*

Another type of difficulty arises from terminological variation. There is an examination that measures serum thyroid stimulating hormone (TSH) level, but another examination that measures "TSH receptor antibody" is also described as "antibody of TSH receptor", or "antibodies of TSH receptors", in actual literature. This circumstance also compromises simple dictionary matching strategies such as the longest match.

	OMIM		Orphanet	
Sentences with Examination Name (EN+)	13,343	(6.9%)	2,555	(7.8%)
Sentences with Evaluative Expression (EE+)	30,577	(15.8%)	4,566	(13.9%)
Sentences with both EN+ and EE+	4,742	(2.4%)	755	(2.3%)
Sentences with nominal laboratory results	4,836	(2.5%)	1,165	(3.6%)
Total # of sentences	193,687	(100.0%)	32,768	(100.0%)

Table 5: Number of sentences with examination names and evaluative expressions

		Evaluative expression		
		+	−	
Examination name	+	**2.3–2.4%**	4.5–5.5%	**6.9–7.8%**
	−	11.6–13.4%	**2.5–3.6%**	15.2–15.9%
		13.9–15.8%	7.0–9.1%	

Table 6: Number of sentences in the categories

As illustrated, automated generation of the dictionary is not a practical approach. Instead, we manually paraphrased examination names in the catalog for laboratory examinations (Fumimaro Takaku (ed.), 2009) and developed a dictionary by adding their synonyms and acronyms in succession, searching against the target document with the aim of better detection power. In this process, the catalog was used to extract laboratory examinations, simply because it is widely used in Japan and covers various categories of examinations for clinical purposes. Statistics of the resulting dictionary is shown in Table 4, with the number of synonyms compiled.

3.3. Dictionary for nominal laboratory result

As shown in Section 2., we found nominal expressions of examination results in medical literature. Such an expression can be tentatively defined as the expression of a finding or a medical state that is conclusively determined by a single laboratory examination. For example, *hyperkalemia* is identified by a test measuring serum potassium level, and thus such an expression is equivalent to "increased potassium". This circumstance suggests that the class of expressions may have characteristic prefixes, such as *hyper*- and *hypo*-, as well as a suffix, such as -*mia*.

To enumerate such expressions, the following operation was performed. First, the SNOMED CT vocabulary (International Health Terminology Standard Development Organisation, 1999) was used to extract terms that have the prefixes (*hyper*- and *hypo*-) and the suffix (-*mia*), resulting in 145, 102, and 476 terms, respectively. Second, SNOMED CT contains 2,285 concepts that have the label, "On examination", in their names, and 201 single-word concepts were selected for further processing. The resulting 924 terms contained findings that are simply discovered by physical examination, such as *hypodontia* (congenital absence of teeth), and disorders not diagnosed by a single laboratory finding alone, such as *ischemia*. Accordingly, the list was manually inspected to extract terms that denote clinical states determined only by a single laboratory test in ordinary clinical settings, and finally 244 terms were identified that matched the criteria.

3.4. Statistics and Analysis

Finally, items in the dictionaries were searched against the descriptions of diseases in public disease databases. To this end, the descriptions of diseases contained in OMIM (John Hopkins University, 1987) were downloaded, and then the GENIA Sentence Splitter (Sætre et al., 2007) was applied to the descriptions, resulting in 587,601 sentences, after the removal of duplications and predefined headings. Then, for 6,727 valid disease records, 193,687 sentences were selected for statistics, without further error correction. Preprocessing for Orphanet (INSERM SC11, 1997) was performed in a similar manner. Orphanet contained 6,442 disease records, but there were 3,073 records with an identical "under construction" message. The remaining 3,369 records were selected for further processing, and the sentence splitter yielded 32,768 sentences. Finally, items in the dictionaries were searched against the sentences, simply with the following command.

```
grep -i -w -f dictfile sentencefile
```

The result is shown in Table 5. Sentences with examination names accounted for 6.9–7.8% of the total sentences, and sentences with evaluative expression accounted for twice as many, accounting for 13.9–15.8% of the entire sentences. Sentences that contained both examination names and evaluative expressions accounted for 2.3–2.4%. In addition, nominal terms were searched in OMIM and Orphanet, and the expressions appeared in 2.5% and 3.6% of the sentences, respectively.

These numbers in Table 5 were then transformed into a 2x2 table (Table 6), for comparison with Table 2. In the table, bold items are measured in the analysis, and other items are calculated from the measured numbers. Table 6 suggests that there is a gap between the number of expressions that include (6.9–7.8%) and do not include (15.2–15.9%) examination names. A breakdown of the numbers indicates a further gap between evaluative expressions that contain (2.3–2.4%) and do not contain (11.6–13.4%) examination names.

The dictionary for evaluative expressions includes prepositions such as *above* and *below* that appear often in ordinary English writings, and adjectives such as *normal* and *abnormal* may be used for clinical findings, as well as for laboratory examination results. Accordingly, it is likely that statistics overestimates the number of evaluative expressions for laboratory result expressions (11.6–13.4%). In contrast, the number of expressions with neither examination names nor evaluative expressions (2.5–3.6%) accounted for only limited vocabulary, and the actual figure was underestimated. For example, medical literature occasionally contains acronyms such as *B27* and *Rh(+)*, which suggest that laboratory examinations for HLA typing and blood typing were performed. Because comprehensive listing of such expressions requires far greater costs, these cases are not included in this study, resulting in the underestimation. Combined together, although the simple statistics is not sufficiently accurate to draw a firm conclusion, the number (15.2–15.9%) suggests undocumented classes of expressions that merit further investigation.

4. Discussion

Although the estimation in the last section was rudimentary, the result suggested that there are expressions of laboratory examination results with and without their names in the expressions. This section discusses the findings of the preliminary analysis, in the light of future corpus design to interpret expressions of laboratory examination results.

First and foremost, the newer corpus has to address the omission of examination names, identified in the expressions of laboratory examination results. Existing annotation schemes might recognize examination names and results in expressions independently, and their relationship would be identified (Roberts et al., 2007). However, such a strategy can cover expressions that have simple relationship between examination names and their results. Accordingly, in the future corpus, several factors must be taken into consideration. Examinations with quantitative results can be expressed in a descriptive manner, with evaluative adjectives and without explicit reference to the examination name. In this class of expressions, the most complex case is exemplified in the *Ham test* example, where the restoration of the examination name requires scientific knowledge and inference. Examinations with qualitative results can also be expressed without their formal examination name; for example, those with microscopical identification. We also identified a nominalized form of laboratory result expression, such as *hypercholesterolemia*. For recognition of these expressions, they must be annotated as such, preferably associated with their restored examination names.

Second, we analyzed evaluative expressions (Table 3) in medical literature and proposed eight categories for evaluative expressions that represent certain aspects of such expressions. However, in clinical settings, “slight increase” and “severe increase” may have different meanings. For example, a “slight increase” of WBC may be observed even in healthy person, whereas a “severe increase” certainly suggests abnormality in the body. Accordingly, the *degree of change* must be taken into account in the annotation scheme.

Third, expressions of qualitative results also demand special handling, for example, results for various typing tests and genetic analysis. The nominal expressions of laboratory results, and the *Ham test* example might be included here. In this area, paraphrasing technologies (Androutsopoulos and Malakasiotis, 2010) may help to convert the expressions into more tractable form, given a sufficient number of annotated documents.

Lastly, complexity in the expressions of laboratory expressions must be addressed, in addition to the processing of names and evaluative expressions. This preliminary study revealed that only one-third of sentences with examination names contained an evaluative expression, suggesting that sentences displaying a simple relation between an examination name and an examination result are in the minority. Some sentences included multiple evaluative expressions and conjunctions, further complicating the interpretation process. Besides, the result expressions might deliver generalized information, as well as description of cases. The corpus needs to provide necessary information to appropriately handle such cases.

5. Related Work

The processing of expressions for laboratory examination results in free-text format has been studied in a very limited context. We performed systematic survey on the PubMed (National Library of Medicine, 1996) and the ACL Anthology (The Association for Computational Linguistics, 2002), and found just a few items in this problem domain.

Zhang (2006) suggested a tag, *medical test*, for the expressions of examination results in their corpus of clinical case reports. The paper randomly selected case reports in a database and removed reports that described groups of patients with the aim of extracting descriptions of individual patients. However, the report was not accompanied by in-depth analysis of the expressions, and the corpus is not available for research use. Okumura et al. (2013) presented a small corpus with similar tags for laboratory examination results in medical literature, and suggested the need for knowledge extraction of laboratory examination results, also without detailed analysis.

To recognize expressions of laboratory results, corpora for medical records may be reused, such as the CLEF corpus (Roberts et al., 2007) and the i2b2 corpus (Uzuner et al., 2011). However, as suggested throughout this paper, there is an unignorable amount of expressions for laboratory test results, that do not contain an explicit reference to examination names. Accordingly, existing corpora are unusable for knowledge acquisition of that type of results. Kang and Kayaalp (2013) shares the same limitations, which attempted to recognize four elements, (name of specimens, analytes, units of measures and detection limits) that explicitly appear in descriptions of laboratory tests. An exception is a study conducted by Bhatia et al. (2010) that attempted to interpret examination results for the management of diabetes mellitus patients, including weight, blood pressure, low-density lipoprotein (LDL), high-density lipoprotein (HDL), creatinine, total cholesterol (TC), glucose, fasting glucose, and glycated hemoglobin (HbA1C) levels, in medical records. The study reported 80.0–98.3% recall and

88.8–100% precision, which is impressive. However, descriptions for weight, blood pressure, and HbA1C levels are simple enough to be processed only with predefined rules, and the remaining tests contain very limited variations in their expressions. Accordingly, they are not applicable to a more general expression.

The interpretation of examination results appear in guidebooks for laboratory result interpretation (Wallach, 2007; Pagana and Pagana, 2009), but only common diseases are mostly covered, and a comprehensive knowledge source even for rare diseases is not yet available. Okumura and Tateisi (2013) analyzed the logical structure of such a guidebook and investigated the efficient acquisition and representation of the knowledge in the book. There are ontologies for laboratory examinations and their results, which were developed for the integration and exchange of laboratory data across institutions, but they are not designed for the interpretation of the results described in free format. (Baorto et al., 1997; Baorto et al., 1998; McDonald et al., 2003; Khan et al., 2006).

6. Conclusion

The relationship between laboratory examination results and their causes are available in guidebooks for laboratory examinations. However, the causes listed in the literature are common diseases, and rare diseases are seldom covered. Accordingly, the development of a disease knowledge base with broad coverage necessitates the processing of expressions for laboratory examination results, available in various medical literatures such as case reports and review articles. However, a detailed study of such expressions has not been explored in the field of medical NLP.

This study identified that expressions of laboratory examination results can be classified into two major categories, that with and without examination names. The expressions without examination names may seem counterintuitive, but it is natural for domain experts to omit formal names if there is more essential information than examination names in the context and if the description carries the information necessary to infer the names. The processing of such expressions requires scientific knowledge of laboratory examinations and an inference mechanism.

This study also identified evaluative expressions that tend to appear in quantitative expressions of laboratory examination results. The expressions are used for differentiating quantitative and qualitative results, each of which requires an independent strategy to interpret expression of laboratory examination results. This axis, coupled with the examination name axis, should define four classes of expression of laboratory examination results that must be identified to extract the necessary information.

Knowledge about the causal relationship between diseases and their laboratory findings is a key component of clinical decision support systems, and automated acquisition of this knowledge would contribute to their improved quality. However, existing corpora in medical NLP studies have been mostly designed to interpret clinical findings in medical records. Although the methodology used in this article is still preliminary and the data volume is limited, the results suggest that the interpretation of laboratory results expression in medical literature requires unique considerations that have not yet been covered by already published studies. To promote further research, it would be desirable to develop an appropriate corpus designed for the interpretation of medical literature.

7. References

Ion Androutsopoulos and Prodromos Malakasiotis. 2010. A survey of paraphrasing and textual entailment methods. *Journal of Artificial Intelligence Research*, 38(1):135–187, May.

Alan R. Aronson. 2001. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In *AMIA Annual Symposium*, pages 17–21.

David M. Baorto, James J. Cimino, Curtis A. Parvin, and Michael G. Kahn. 1997. Using logical observation identifier names and codes (loinc) to exchange laboratory data among three academic hospitals. In *Proceedings of the AMIA Annual Fall Symposium*, pages 96–100. American Medical Informatics Association.

David M. Baorto, James J. Cimino, Curtis A. Parvin, and Michael G. Kahn. 1998. Combining laboratory data sets from multiple institutions using the logical observation identifier names and codes (loinc). *International journal of medical informatics*, 51(1):29–37.

Ramanjot S. Bhatia, Amber Graystone, Ross A. Davies, Susan McClinton, Jason Morin, and Richard F. Davies. 2010. Extracting information for generating a diabetes report card from free text in physicians notes. In *the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents*, pages 8–14. Association for Computational Linguistics.

Carol Friedman, Stephen B Johnson, Bruce Forman, and Justin Starren. 1995. Architectural requirements for a multipurpose natural language processor in the clinical environment. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, page 347. American Medical Informatics Association.

Fumimaro Takaku (ed.). 2009. *Laboratory examinations databook 2009-2010*. Igaku-Shoin. (in Japanese).

INSERM SC11. 1997. Orphanet. http://www.orpha.net/.

International Health Terminology Standard Development Organisation. 1999. SNOMED CT. http://snomed.org/.

John Hopkins University. 1987. OMIM: Online Mendelian Inheritance in Man. http://www.ncbi.nlm.nih.gov/omim.

Yanna Shen Kang and Mehmet Kayaalp. 2013. Extracting laboratory test information from biomedical text. *Journal of pathology informatics*, 4.

Agha N. Khan, Stanley P. Griffith, Catherine Moore, Dorothy Russell, Arnulfo C. Rosario Jr, and Jeanne Bertolli. 2006. Standardizing laboratory data by mapping to loinc. *Journal of the American Medical Informatics Association*, 13(3):353–355.

Clement J. McDonald, Stanley M. Huff, Jeffrey G. Suico, Gilbert Hill, Dennis Leavelle, Raymond Aller, Arden Forrey, Kathy Mercer, Georges DeMoor, John Hook, et al. 2003. Loinc, a universal standard for identifying laboratory observations: a 5-year update. *Clinical chemistry*, 49(4):624–633.

National Library of Medicine. 1996. PubMed. http://www.ncbi.nlm.nih.gov/pubmed.

Takashi Okumura and Yuka Tateisi. 2013. Interpretation of laboratory examination results and their simple representation. In *IEEE International Symposium on Computer-Based Medical Systems (CBMS2013)*, June.

Takashi Okumura, Eiji Aramaki, and Yuka Tateisi. 2013. Classification and characterization of clinical finding expressions in medical literature. In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM2013)*, December.

Kathleen D. Pagana and Timothy J. Pagana. 2009. *Mosby's manual of diagnostic and laboratory tests*. Elsevier Health Sciences.

Angus Roberts, Robert Gaizauskas, Mark Hepple, Neil Davis, George Demetriou, Yikun Guo, Jay Subbarao Kola, Ian Roberts, Andrea Setzer, Archana Tapuria, et al. 2007. The CLEF corpus: semantic annotation of clinical text. In *AMIA Annual Symposium Proceedings*, volume 2007, pages 625–629. American Medical Informatics Association.

Rune Sætre, Kazuhiro Yoshida, Akane Yakushiji, Yusuke Miyao, Y Matsubyashi, and Tomoko Ohta. 2007. Akane system: protein-protein interaction pairs in the biocreative2 challenge, ppi-ips subtask. In *Proceedings of the Second BioCreative Challenge Workshop*, pages 209–212.

Guergana K. Savova, James J. Masanz, Philip V. Ogren, Jiaping Zheng, Sunghwan Sohn, Karin C. Kipper-Schuler, and Christopher G. Chute. 2010. Mayo clinical text analysis and knowledge extraction system (cTAKES): Architecture, component evaluation and applications. *Journal of the American Medical Informatics Association*, 17(5):507–513.

The Association for Computational Linguistics. 2002. ACL Anthology. http://aclweb.org/anthology/.

Özlem Uzuner, Brett R South, Shuying Shen, and Scott L DuVall. 2011. 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association*, 18(5):552–556.

Jacques B. Wallach. 2007. *Interpretation of diagnostic tests*. Lippincott Williams & Wilkins.

Yitao Zhang and Jon Patrick. 2006. Extracting patient clinical profiles from case reports. In *the 2006 Australasian Language Technology Workshop (ALTW2006)*, pages 167–168.