AAAI Technical Report FS-12-05
Information Retrieval and Knowledge Discovery in Biomedical Text

Automatic Formalization of Clinical Practice Guidelines

Matthew S. Gerber and **Donald E. Brown**
Department of Systems and Information Engineering
University of Virginia
151 Engineer's Way
Charlottesville, VA 22904
{*msg8u,deb*}@*virginia.edu*

James H. Harrison
Department of Public Health Sciences
University of Virginia
Hospital West Complex, Suite 3000
Charlottesville, VA 22908-0717
jhh5y@virginia.edu

Abstract

Current efforts aim to incorporate knowledge from clinical practice guidelines (CPGs) into computer systems using sophisticated interchange formats. Due to their complexity, such formats require expensive manual formalization work. This paper presents a preliminary study of using natural language processing (NLP) to automatically formalize CPG recommendations. We developed a CPG representation using concepts from the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED–CT), and manually applied this representation to a sample of CPG recommendations that is representative of multiple medical domains and recommendation types. Using this resource, we trained and evaluated a supervised classification model that formalizes new CPG recommendations according to the SNOMED–CT representation, achieving a precision of 75% and recall of 42% ($F_1 = 54\%$). We have identified two important lines of future investigation: (1) feature engineering to address the unique linguistic properties of CPG recommendations, and (2) alternative model formulations that are more robust to processing errors. A third line of investigation – creating additional training data for the NLP model – is shown to be of little utility.

Introduction

Clinical practice guidelines (CPGs) were originally conceived as an information source to be used by physicians in the clinical decision making process (Field and Lohr 1990). CPGs provide recommendations for a wide variety of decisions including prevention, diagnosis, treatment, management, counseling, and others. CPG proponents expect compliance with these recommendations to improve patient health and reduce treatment and outcome variation (Field and Lohr 1990), and multiple studies in different medical domains have confirmed this expectation (Golub 2009; Lugtenberg, Burgers, and Westert 2009; Quaglini et al. 2004; Reker et al. 2002; Rutten et al. 2010). However, it remains difficult to introduce CPG information into daily practice in a way that reliably affects clinical decisions (Godemann et al. 2010; Grol 2001; Gross et al. 2001; Pathman et al. 1996; Tunis et al. 1994), largely because CPGs are formulated as free-text documents that cannot be used directly within clinical intelligence systems (e.g., decision support). Researchers have attempted to address this problem with computer-interpretable guideline interchange formats (e.g., GEM (Shiffman et al. 2000)), but the formalization process tends to be manual and complex, requiring experience with logical formalisms and knowledge engineering software.

Given the difficulties and expense associated with manual CPG formalization, researchers have begun to investigate the feasibility of automatically formalizing aspects of CPGs within particular medical domains. Seminal work in automatic CPG formalization was based largely on hand-crafted, non-probabilistic, domain-specific formalization rules (Kaiser 2005; Kaiser, Akkaya, and Miksch 2005; 2007; Kaiser and Miksch 2007; 2009). Results of these studies have shown promise; however, this work has focused primarily on hand-crafted formalization rules within a relatively simple medical domain (Kaiser 2005). Although it can be appropriate to focus initial investigations on particular domains and phenomena, we believe automated CPG formalization techniques will be more useful if investigated more generally.

In our study, we investigated the application of current NLP techniques to the task of automatic CPG formalization. We developed a recommendation representation using concepts from the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED–CT).[1] We drew CPG recommendations from the Yale Guideline Recommendation Corpus (YGRC), which covers a broad range of medical domains and recommendation types (Hussain, Michel, and Shiffman 2009). We then manually formalized these recommendations according to the SNOMED–CT representation. To our knowledge, this gold-standard formalization resource is unique. Finally, we trained and evaluated an automatic recommendation formalization model. In sum, this paper presents an initial exploration of automatic CPG formalization using current NLP techniques. Further progress in this area could lead to improved clinical intelligence applications (e.g., decision support) based on CPG knowledge. In the following sections, we describe our data collection effort and formalization experiments.

[1] http://www.nlm.nih.gov/research/umls/Snomed/snomed/main.html

Primary label	Secondary label	Example
ACTION	EVALUATION (129265001)	[computed tomography CT] should be used
	DRUG (410942007)	[oral risedronate] should not be used
	DOSAGE (260911001)	treat with oral risedronate [5 mg daily]
	PROCEDURE (71388002)	[red blood cell transfusion] is appropriate
EVIDENCE	STRONG (18669006)	[it has been shown to reduce the occurrence of NTDs]
	WEAK/NONE (41647002)	[there is insufficient evidence]
MODALITY	OBLIGATORY	computed tomography CT [should] be used
	NEVER	oral risedronate [should not] be used
	OPTIONAL	physician [may] choose
AGENT (223366009)		[physician] may choose
MORBIDITY (64572001)		prevent [preeclampsia]
POPULATION (385436007)		[obese women with gestational diabetes mellitus]
PURPOSE (288830005)		is used [to prevent osteoporotic fractures]
TEMPORAL (410669006)		[initial] treatment
TRIGGER		Diuretics are [recommended]

Table 1: CPG recommendation representation. The first column indicates a broad annotation label, which is specialized by labels shown in the second column. The third column provides examples of the primary/secondary label. Labeled text spans are indicated with square brackets. Numbers next to the labels indicate SNOMED–CT concept identifiers. Note that we have included representational elements for various modalities that do not have concepts in SNOMED–CT.

Data Collection

Our study treated automatic CPG formalization as a supervised classification task. Since no appropriate, gold-standard formalization corpora exist, we created our own. In the next section, we describe our data representation for CPG recommendations. Then we describe our manual formalization effort, which produced a set of CPG recommendations formalized according to the representation.

CPG Recommendation Representation

Our goal was to create a CPG representation that balances the need to express detailed knowledge with the need to automatically formalize guidelines according to the representation. This is an important tradeoff. At the extreme end are representations such as Asbru, which can express intricate actions, plans, and temporal dependencies (Peleg et al. 2003). Although this representation can enable complex reasoning within clinical intelligence systems, we do not believe that current NLP methods will support the automatic formalization of CPGs at this level of sophistication. On the other hand, an overly simplified representation would be easy to fill out using NLP techniques, but it might not support any useful applications.

In our study, we represented key CPG recommendation elements using concepts from SNOMED–CT, a comprehensive and widely used source of medical terminology. Consider the following recommendation from the Heart Failure Society of America's heart failure practice guideline (HFSA 2010), which has been marked up with SNOMED–CT concepts (indicated with subscripts):

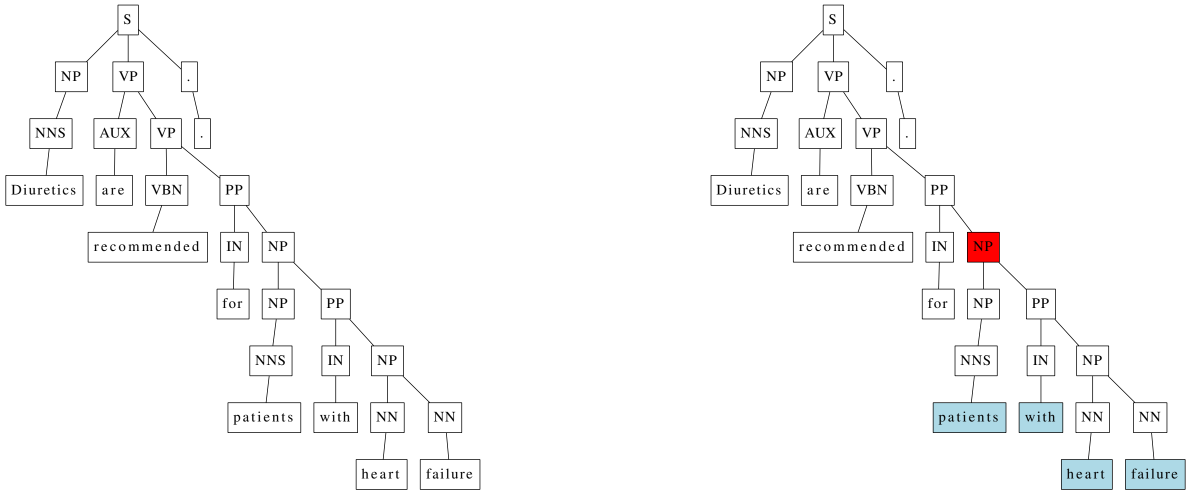
(1) [$_{DRUG}$ Diuretics] are recommended for [$_{POPULATION}$ patients with [$_{MORBIDITY}$ heart failure]].

As shown, SNOMED–CT concepts cover all major pieces of information contained in this recommendation. Table 1 shows our complete representation, which we believe strikes a useful balance between expressivity and NLP capabilities. Elements in this representation should support basic clinical intelligence applications, but should also be identifiable using current NLP techniques. As shown in Table 1, it was not always possible to match our representational elements to those in SNOMED–CT (e.g., for triggers, which provide an overt indication of a recommendation's presence).

CPG Recommendation Annotation

We applied the representation in Table 1 to a sample of CPG recommendations. We addressed the need for domain neutrality by basing our annotation effort on the Yale Guideline Recommendation Corpus (YGRC) (Hussain, Michel, and Shiffman 2009). To create this corpus, the authors first randomly sampled guidelines from the National Guideline Clearinghouse (NGC).[2] The authors then sampled a fixed number of recommendations from each guideline, where a recommendation was defined as "a statement whose apparent intent is to provide guidance about the advisability of a clinical action"(Hussain, Michel, and Shiffman 2009). Recommendations were sampled randomly from the guidelines to remove bias, as guidelines often present recommendations in particular orders (e.g., screening techniques followed by treatments). In total, the authors sampled 1,275 recommendations in this way. The authors quantified the medical domain distribution of the recommendations using the frequency of terms from a controlled vocabulary, and found that the domain distribution of the YGRC follows the domain distribution of the entire NGC corpus quite closely. Thus, the YGRC is an ideal data source for the creation and testing of automatic, domain-neutral recommendation formalization models.

[2] http://www.guideline.gov



(a) The syntactic parse tree for Example 2.

(b) Labeling the highlighted NP node labels all of the text that the node subsumes (i.e., "patients with heart failure").

Figure 1: Syntactic analysis for CPG recommendation labeling.

n's left/right siblings's syntactic category	Semantic type of *n*'s UMLS concept	*n*'s first/last word/POS
n's semantic argument positions within VerbNet	Number of *n*'s left/right siblings	Object head of following PP
n's parent's grammar rule and syntactic category	Whether or not *n* is followed by PP	*n*'s head word
Right-most NP's head word/POS if *n* is a PP	*n*'s left/right sibling's head word/POS	*n*'s synset within WordNet
Semantic type of *n*'s semantic head's UMLS concept	*n*'s left/right sibling's grammar rule	Whether or not *n* is a PP
Syntactic path to the passive verb nearest to *n*	*n*'s semantic argument positions	*n*'s syntactic category
Words surrounding *n*'s, within a three-word window	*n*'s grandparent's grammar rule	*n*'s head word if *n*'s parent is PP
n's left/right sibling's semantic head word	*n*'s semantic head's UMLS concept	Frequency of words within *n*
n's great grandparent's grammar rule	*n*'s left/right sibling's first/last word	*n*'s parent's head word's POS
		n's concept within the UMLS

Table 2: Features used for automatic CPG recommendation formalization. The variable *n* refers to the node being labeled.

For our study, we randomly selected 200 CPG recommendations from the YGRC and applied labels from our representation (Table 1) to each recommendation using the MMAX2 text annotation toolkit.[3] Our annotated recommendations look very much like the one shown in Example 1, with concept labels being applied to spans of text. In practice, we found our SNOMED–CT representation to be unambiguous (i.e., only a single label was applicable to a span of text) and efficient to apply. We used these 200 example formalizations to train and evaluate our automatic formalization model, which we describe below.

Methods

Our CPG recommendation formalization model begins with a syntactic analysis of the recommendation text using the statistical syntactic parser created by Charniak and Johnson (2005). Consider the following recommendation:

(2) Diuretics are recommended for patients with heart failure.

[3]One of the authors performed all of the annotation using the MMAX2 annotation environment (http://mmax2.net).

The syntactic analysis in Figure 1a shows the internal structure of the sentence in Example 2. The internal structure forms an inverted tree with a single root node S (for sentence) and many leaf nodes, each containing a single word from the recommendation text. The nodes in the tree allow us to identify phrases such as the highlighted noun phrase (NP) in Figure 1b, which subsumes the text "patients with heart failure". This rich phrase structure has proven to be extremely useful in many NLP tasks (e.g., semantic analysis (Punyakanok, Roth, and Yih 2008), machine translation (May and Knight 2007), and discourse parsing (Sagae 2009)).

We used multi-class logistic regression to apply labels from the representation in Table 1 to nodes within the syntactic parse tree of a CPG recommendation. For example, the logistic regression model might apply a label of POPULATION to the highlighted NP in Figure 1b. In addition to the labels contained in our representation, we defined a special label NULL. This label is required because each node in the syntax tree is given a label, but not all nodes are associated with an element of our representation. The AUX node in Figure 1b is an example of this.

Our logistic regression model uses the features in Table 2 to make labeling decisions. A full explanation of these features is beyond the scope of this paper, and we refer the interested reader to other work in which many of these features are described more fully (Gerber and Chai 2010). However, a few features in Table 2 involve the Unified Medical Language System (UMLS)[4] and are not documented elsewhere. Referring again to Figure 1b, note the NP node above "diuretics". The last feature in Table 2 characterizes each node by indicating the node's UMLS concept. In the case of the "diuretics" NP, the UMLS concept is C0012798. This concept includes the term "diuretics" as well as "water pills". The purpose of the UMLS feature is to provide the model with knowledge that can generalize to CPG text that it does not encounter during training. For example, even if the model is only exposed to "diuretics" during training, it will know something about "water pills" by virtue of the UMLS concept C0012798, which contains both terms. Identification of UMLS concepts was done by simple string matching within the UMLS database.

Results

CPG Annotation

As described above, we manually formalized 200 randomly selected recommendations from the YGRC according to the SNOMED–CT representation in Table 1. In total, we identified 834 representational elements in these recommendations. Table 3 shows how many of each element type were annotated in our recommendation set. Notably, elements such as AGENT and MORBIDITY are rarely observed in the recommendations. This is because such concepts are usually implied by the title of the CPG and the textual context of the recommendations.

To our knowledge, this is the only existing CPG dataset that contains textual annotations of this type, and we have made it freely available for research purposes.[5] Hopefully this will encourage further studies of automatic CPG formalization.

Model Evaluation

We evaluated our automatic formalization model using tenfold cross-validation over the manually formalized CPG recommendations described above. For each testing fold, we found the optimal logistic regression parameters using the LibLinear toolkit (Fan et al. 2008) and data contained in the training folds. Furthermore, we employed forward feature subset selection to identify the optimal subset of features for each training set (Pudil, Novovicova, and Kittler 1994). It is important to note that the feature selection process did not have access to information contained in the testing folds. Having trained the ten models, we evaluated each over its corresponding testing fold. We used the familiar metrics of precision, recall, and F_1 over the gold-standard representational elements. Table 3 presents the results. As shown, the model achieved an overall F_1 score of approximately 54% over 834 gold-standard representational elements. We discuss these results in the following section.

[4]http://www.nlm.nih.gov/research/umls

[5]http://ptl.sys.virginia.edu/ptl/projects/medical-informatics

Element	**#**	**P (%)**	**R (%)**	F_1 **(%)**
ACTION	301	74.5	26.2	38.7
MODALITY	158	71.9	73.0	72.4
POPULATION	140	83.7	56.6	67.5
TEMPORAL	53	28.6	1.1	2.2
TRIGGER	45	81.7	93.3	87.1
PURPOSE	43	58.3	16.3	25.5
EVIDENCE	38	83.3	13.2	22.7
AGENT	37	76.0	47.6	58.5
MORBIDITY	19	50.0	10.5	17.4
Overall	834	75.3	41.7	53.7

Table 3: Evaluation results. The first column indicates the representational element being evaluated. The second column indicates the number of elements present in the gold-standard recommendations. The final columns indicate precision, recall, and F_1 score for identifying the given elements using the NLP model.

Discussion

As shown in Table 3, the performance of the automatic formalization model varies across the representational elements. For some elements (e.g., POPULATION, MODALITY, and TRIGGER), the model exhibits performance that is on par with the current state of the art in related NLP tasks (e.g., semantic role labeling (Punyakanok, Roth, and Yih 2008)). Other representational elements are more difficult to identify. For example, ACTION elements are recovered with precision of 74.5% but recall of 26.2%. One could lower the prediction threshold to trade precision for recall; however, we doubt this would increase the overall F_1 score. Below, we discuss alternative strategies for increasing the model's performance.

Identify more informative features. The feature set used by our formalization model is not particularly customized to the CPG domain. We have borrowed many features from our prior work in semantic analysis of natural language text (Gerber and Chai 2010). We believe this feature set can be expanded to include features that specifically target the medical language used within CPGs. This would improve the model's ability to recognize clinical terminology (e.g., clinical actions) within CPG recommendations and would probably increase recall, which is a primary deficiency of the current model.

Change the model formulation. Our current recommendation formalization model has specific limitations that stem from its reliance on syntactic structures (i.e., the inverted trees shown in Figure 1). The arrangement of the internal nodes constrains the possible spans of text that can be labeled. For example, one cannot exactly label the text "are recommended" in the parse trees of Figure 1. When the

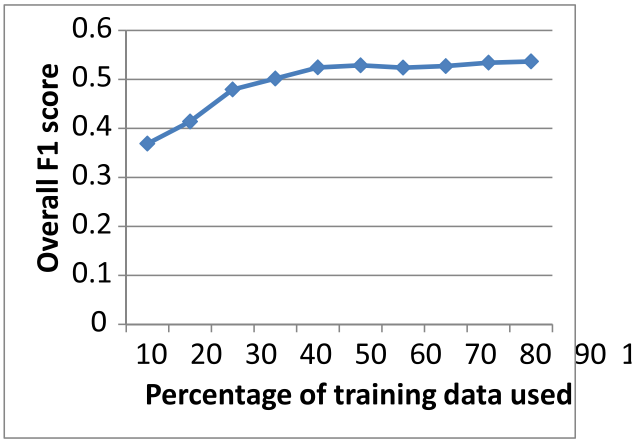


Figure 2: Effect of training set size on overall formalization performance. We reduced the training set size by randomly removing annotated recommendations. The y-axis shows the evaluation results using a constant evaluation set.

parse tree is correctly structured, this constraint prevents incorrect labelings; however, the parse tree structure is produced by an automatic process that occasionally makes mistakes. Thus, in some situations it might be impossible for the model to label the correct span of text. In our data, 158 (or 19%) of the 834 recommendation elements are not properly subsumed by single parse tree nodes. This limits the recall of our model to approximately 80%.

An alternative approach would be to label sequences of words instead of nodes in syntactic parse trees. This approach would allow the model to label arbitrary spans of text within a recommendation and would exhibit a recall upper bound of 100%. Such an approach has proven useful in related NLP tasks such as named entity identification (Bikel, Schwartz, and Weischedel 1999), and we believe it holds promise for recommendation formalization.

Annotate more data. In general, supervised statistical models like the one in our study tend to perform better when provided with larger training datasets. Thus, it might be possible to increase automatic formalization performance by manually formalizing additional CPG recommendations from the YGRC. Doing so could provide the model with better parameter estimates; however, the annotation process requires substantial time and should be justified by empirical evidence. To assess the potential gains from additional manual formalization, we systematically reduced the size of the existing training datasets and observed the changes in the evaluation metrics. Figure 2 presents the results. As shown, gains from using larger amounts of training data diminish substantially. If this curve were to be extrapolated beyond 100% (i.e., into the region covered by additional manual annotation effort), it does not appear that significant gains would be made.

Conclusions and Future Work

The study presented in this paper focused on the automatic formalization of CPG recommendations using NLP methods. Prior to this study, automatic CPG formalization had only been investigated within specific medical domains and for a limited number of recommendation types. Our study broadens the scope of automatic CPG formalization through the creation of a new, gold-standard formalization resource. This resource, which is freely available, is the result of manually identifying SNOMED–CT concepts within a sample of 200 CPG recommendations covering a broad range of medical domains and recommendation types. To our knowledge, this is the only existing set of manually formalized CPG recommendations that seeks to enable automatic NLP methods.

We have investigated the use of supervised classification and standard NLP feature sets for automatic recommendation formalization. Our method (i.e., labeling nodes in a syntax tree) is not novel within the NLP community, but this is the first time it has been applied to the CPG domain. Judging from our current results and analyses, we believe this is a promising approach. Our current formalization model exhibits precision levels that are on par with state-of-the-art methods in related NLP tasks; however, the model's recall levels are not sufficient for reliable CPG recommendation formalization. We have identified a number of areas for future work, in particular feature engineering and model formulation. We have made the important observation that, for the current approach, additional annotation effort is not likely to improve formalization performance.

Our long-term goal is to provide a method for accurate, automatic formalization of CPGs from multiple medical domains. This would enable CPG libraries such as the National Guideline Clearinghouse to augment each textual CPG with a formalized version. The formalized versions could be generated quickly as new guidelines and guideline updates are developed. Furthermore, the formalized versions could be automatically consumed by a broad range of client systems such as intelligent clinical decision support and performance analysis. Our study will serve as a baseline for future work in this area.

References

Bikel, D. M.; Schwartz, R.; and Weischedel, R. M. 1999. An algorithm that learns what's in a name. *Mach. Learn.* 34(1-3):211–231.

Charniak, E., and Johnson, M. 2005. Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 173–180.

Fan, R.-E.; Chang, K.-W.; Hsieh, C.-J.; Wang, X.-R.; and Lin, C.-J. 2008. LIBLINEAR: A Library for Large Linear Classification. *Journal of Machine Learning Research* 9:1871–1874.

Field, M., and Lohr, K. 1990. *Clinical practice guidelines: directions for a new program.* Natl Academy Pr.

Gerber, M., and Chai, J. 2010. Beyond NomBank: A study of implicit arguments for nominal predicates. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 1583–1592. Uppsala, Sweden: Association for Computational Linguistics.

Godemann, F.; Blittersdorf, K.; Poschenrieder, M.; Klimitz, H.; Hauth, I.; and Gutzmann, H. 2010. Guideline compliance in the treatment of schizophrenic patients. introduction of a computer-assisted treatment pathway. *Nervenarzt* 81(5):584–593.

Golub, R. 2009. Clinical practice guidelines and scientific evidence. *JAMA* 302(2):142–147.

Grol, R. 2001. Successes and failures in the implementation of evidence-based guidelines for clinical practice. *Medical Care* 39(8):1146–1154.

Gross, P. A.; Greenfield, S.; Cretin, S.; Ferguson, J.; Grimshaw, J.; Grol, R.; Klazinga, N.; Lorenz, W.; Meyer, G. S.; Riccobono, C.; Schoenbaum, S. C.; Schyve, P.; and Shaw, C. 2001. Optimal methods for guideline implementation: Conclusions from Leeds Castle meeting. *Medical Care* 39(8):1185–1192.

HFSA. 2010. HFSA 2010 comprehensive heart failure practice guideline. *Journal of Cardiac Failure* 16:475–733.

Hussain, T.; Michel, G.; and Shiffman, R. N. 2009. The Yale guideline recommendation corpus: A representative sample of the knowledge content of guidelines. *I. J. Medical Informatics* 78:354–363.

Kaiser, K.; Akkaya, C.; and Miksch, S. 2005. Gaining process information of clinical practice guidelines using information extraction. In *Proceedings of the 10th Conference on Artificial Intelligence in Medicine in Europe, AIME 2005*, volume 3581 of *LNAI*, 181–190. Aberdeen, UK: Springer Verlag.

Kaiser, K.; Akkaya, C.; and Miksch, S. 2007. How can information extraction ease formalizing treatment processes in clinical practice guidelines? *Artif. Intell. Med.* 39:151–163.

Kaiser, K., and Miksch, S. 2007. Modeling treatment processes using information extraction. In Yoshida, H.; Jain, A.; Ichalkaranje, A.; Jain, L. C.; and Ichalkaranje, N., eds., *Advanced Computational Intelligence Paradigms in Healthcare – 1*, volume 48 of *Studies in Computational Intelligence (SCI)*. Springer Verlag. 189–224.

Kaiser, K., and Miksch, S. 2009. Versioning computer-interpretable guidelines: Semi-automatic modeling of "living guidelines" using an information extraction method. *Artificial Intelligence in Medicine* 46(1):55–66.

Kaiser, K. 2005. *LASSIE - Modeling Treatment Processes Using Information Extraction*. Ph.D. Dissertation, Vienna University of Technology.

Lugtenberg, M.; Burgers, J.; and Westert, G. 2009. Effects of evidence-based clinical practice guidelines on quality of care: a systematic review. *Quality and Safety in Health Care* 18(5):385–392.

May, J., and Knight, K. 2007. Syntactic re-alignment models for machine translation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 360–368. Prague, Czech Republic: Association for Computational Linguistics.

Pathman, D. E.; Konrad, T. R.; Freed, G. L.; Freeman, V. A.; and Koch, G. G. 1996. The awareness-to-adherence model of the steps to clinical guideline compliance: The case of pediatric vaccine recommendations. *Medical Care* 34(9):pp. 873–889.

Peleg, M.; Tu, S.; Bury, J.; Ciccarese, P.; Fox, J.; Greenes, R. A.; Hall, R.; Johnson, P. D.; Jones, N.; Kumar, A.; Miksch, S.; Quaglini, S.; Seyfang, A.; Shortliffe, E. H.; and Stefanelli, M. 2003. Comparing computer-interpretable guideline models: a case-study approach. *Journal of the American Medical Informatics Association : JAMIA* 10(1):52–68.

Pudil, P.; Novovicova, J.; and Kittler, J. 1994. Floating search methods in feature selection. *Pattern Recognition Letters* 15:1119–1125.

Punyakanok, V.; Roth, D.; and Yih, W. 2008. The importance of syntactic parsing and inference in semantic role labeling. *Comput. Linguist.* 34(2):257–287.

Quaglini, S.; Cavallini, A.; Gerzeli, S.; and Micieli, G. 2004. Economic benefit from clinical practice guideline compliance in stroke patient management. *Health Policy* 69(3):305–315.

Reker, D. M.; Duncan, P. W.; Horner, R. D.; Hoenig, H.; Samsa, G. P.; Hamilton, B. B.; and Dudley, T. K. 2002. Postacute stroke guideline compliance is associated with greater patient satisfaction. *Arch Phys Med Rehabil* 83(6):750–756.

Rutten, G.; Degen, S.; Hendriks, E.; Braspenning, J.; Harting, J.; and Oostendorp, R. 2010. Adherence to clinical practice guidelines for low back pain in physical therapy: Do patients benefit? *Physical therapy* 90(8):1111–1122.

Sagae, K. 2009. Analysis of discourse structure with syntactic dependencies and data-driven shift-reduce parsing. In *Proceedings of the 11th International Conference on Parsing Technologies (IWPT'09)*, 81–84. Paris, France: Association for Computational Linguistics.

Shiffman, R.; Karras, B.; Agrawal, A.; Chen, R.; Marenco, L.; and Nath, S. 2000. GEM: A proposal for a more comprehensive guideline document model using XML. *Am. Med. Informatics Assoc.* 7:488–498.

Tunis, S.; Hayward, R.; Wilson, M.; Rubin, H.; Bass, E.; Johnston, M.; and Steinberg, E. 1994. Internists' attitudes about clinical practice guidelines. *Annals of internal medicine* 120(11):956–963.