

# Using Natural Language Processing to Identify Pharmacokinetic Drug-Drug Interactions Described in Drug Package Inserts

**Richard Boyce, PhD**  
University of Pittsburgh  
5607 Baum Avenue  
Pittsburgh, PA 15206, USA  
[rdb20@pitt.edu](mailto:rdb20@pitt.edu)

**Gregory Gardner, MS**  
University of Pittsburgh  
5607 Baum Avenue  
Pittsburgh, PA 15206, USA  
[gag30@pitt.edu](mailto:gag30@pitt.edu)

**Henk Harkema, PhD**  
University of Pittsburgh  
6425 Penn Ave.  
Pittsburgh, PA 15206  
[hendrik.harkema@nuance.com](mailto:hendrik.harkema@nuance.com)

## Abstract

The package insert (aka drug product label) is the only publicly-available source of information on drug-drug interactions (DDIs) for some drugs, especially newer ones. Thus, an automated method for identifying DDIs in drug package inserts would be a potentially important complement to methods for identifying DDIs from other sources such as the scientific literature. To develop such an algorithm, we created a corpus of Federal Drug Administration approved drug package insert statements that have been manually annotated for pharmacokinetic DDIs by a pharmacist and a drug information expert. We then evaluated three different machine learning algorithms for their ability to 1) identify pharmacokinetic DDIs in the package insert corpus and 2) classify pharmacokinetic DDI statements by their modality (i.e., whether they report a DDI or no interaction between drug pairs). Experiments found that a support vector machine algorithm performed best on both tasks with an F-measure of 0.859 for pharmacokinetic DDI identification and 0.949 for modality assignment. We also found that the use of syntactic information is very helpful for addressing the problem of sentences containing both interacting and non-interacting pairs of drugs.

## 1 Introduction

Package inserts (PIs, aka drug product label) are the primary source of information for newly approved drugs and a potentially authoritative source of drug information from a medical-legal stand-

point (Marroum & Gobburu 2002). Among the information provided by PIs are *drug-drug interactions* (DDIs): known and predicted drug combinations that could lead to a clinically meaningful alteration in the effect of one of the drugs. The United States Federal Drug Administration (FDA) mandates that PIs for FDA-approved drugs include both observed and predicted clinically significant DDIs, as well as the results of pharmacokinetic studies that establish the absence of effect (FDA, 2010). Moreover, the PI is the only publicly-available source of information on DDIs for some drugs, especially newer ones (Dal-Ré et al. 2010). Hence, an automated method for identifying DDIs from drug PIs would be an important complement to methods for identifying DDIs from other sources such as the scientific literature. In this paper we describe the creation of a new corpus of FDA-approved drug package insert statements that have been manually annotated for pharmacokinetic DDIs. We then discuss how three different machine learning algorithms were evaluated for their ability to 1) identify pharmacokinetic DDIs in drug package inserts and 2) classify pharmacokinetic DDI statements by their modality (i.e., whether they report a DDI or that a drug pair does not interact).

## 2 Materials and Methods

### 2.1 The DDI Corpus and Schema

A corpus of annotated statements derived from FDA-approved drug PIs was created for use as training and test data while developing automated DDI extraction algorithms. The statements were derived from PIs using a strategy that ensured there

would be a representative sample of statements that 1) unambiguously identified interacting drug pairs, 2) unambiguously identified non-interacting drug pairs, and 3) included no mention of interacting drug pairs. Previous experience by our research group suggested that the manner in which DDI statements are described in PIs has changed over time in response to changing FDA regulations. Most notably, an FDA guidance document issued in 1999 was (to our knowledge) the first to explicitly suggest the inclusion of brief descriptions of pharmacokinetic DDI studies within specific sections of drug PIs (FDA, 1999). To account for this, investigators selected 64 PIs using a strategy that ensured the corpus would have a balanced sample of statements from drugs marketed before and after 2000. For the purpose of this study we designated all PIs for drugs marketed prior to 2000 as “older” and those for drugs marketed in or after 2000 as “newer.” PIs were downloaded from the DailyMed website,<sup>1</sup> and the entire “Drug Interactions” and “Clinical Pharmacology” sections were selected as text sources from “newer” PIs. For “older” PIs, which often lacked these two sections, investigators chose a section containing an apparent interaction statement and one randomly-selected section.

DDIs are typically classified as occurring by either pharmacodynamic or pharmacokinetic mechanisms. A pharmacodynamic DDI involves the additive or synergistic amplification of a drug’s effect. In a pharmacokinetic (PK) DDI, one drug, called a *precipitant*, affects (inhibits or induces) the absorption, distribution, metabolism, or excretion of another drug, called the *object*. To simplify our task, we decided to focus specifically on PK DDIs. Prior to annotating the PI statements, a schema was created for the entities that the investigators considered important components of a PK DDI. The schema modeled drugs as having two characteristics, *type* and *role*. The *type* of drug could be active ingredient (e.g., simvastatin), drug product (e.g., Zocor), or metabolite (e.g., beta-OH-simvastatin). Drugs annotated as metabolite also referred to the active ingredient parent compound. The *role* of a drug could be either an object or a precipitant. Two other properties were provided to model each PK DDI: 1) whether the statement from which the DDI was identified suggested an observed effect or a

lack of an observed effect between two coadministered drugs (i.e., positive vs negative *modality* statements), and 2) whether the statement included quantitative or qualitative data in describing an interaction or non-interaction between a drug pair (i.e., quantitative vs qualitative statements). Finally, the segment of text in which the interaction claim was made was annotated as an *interaction phrase*. With the corpus and schema in place, drugs and PK DDIs present in the PI statements were then annotated by two independent reviewers using Knowtator, an annotation tool integrated with the Protégé ontology editor (Ogren 2006).

One annotator was a pharmacist and DDI expert, and the other a librarian specializing in drug information retrieval. To help the annotators, co-investigator RB ran the NCBO Annotator (Jonquet, Shah & Musen 2009) over the corpus using the RxNorm drug terminology (Nelson et al. 2011) to pre-annotate as many active ingredients and drug products as possible. The annotators reviewed these “pre-annotations” while identifying entities that missed during the pre-annotation process. Co-investigator HH used Knowtator to calculate inter-annotator agreement statistics from the annotators’ initial annotation sets. RB then worked with the two annotators to achieve consensus on the final corpus of annotated DDI statements.

## 2.2 Setting up the DDI statement extraction experiment

Once the set of DDI annotations was compiled, we devised two machine learning tasks. The first task was to determine whether two drugs mentioned in a statement taken from a PI are noted as either interacting or not interacting with each other by pharmacokinetic mechanisms (i.e., *does the statement report a PK DDI with the drug pair of either a positive or negative modality?*). The second task was to determine the modality of a given PK DDI. The first task *did not* include determining the roles of the drugs if an interaction is found, i.e., which member of the pair of drug mentions is the precipitant and which one is the object. To enable the exploration of the performance of multiple machine learning methods, we divided two-thirds of the annotated PI statements into a *development* set and one-third into a blind *test* set. PI statements annotated as reporting DDIs were stratified within the

<sup>1</sup> <http://dailymed.nlm.nih.gov/>

two sets using a random selection method that ensured a representative balance of sentence distance between drug mentions, DDI modality, DDI type, and drug PI age designation (see above). Statements not containing an interaction were stratified by sentence distance between drug mentions, and PI age designation. Stratification was done on the level of statements. Thus, statements taken from the same package insert may have been distributed over the development and test set.

We observed that 99% of corpus statements annotated as a PK DDI mentioned an interacting drug pair within a three sentence region. Thus, we created a *baseline* dataset by iterating through PI statements in the development set and identifying all drug pair mentions that occurred within a three-sentence span. Throughout the remainder of this paper we refer to the statements identified by this process as *instances*.

Instances containing drug pairs that were manually annotated as participating in an interaction (either with positive or negative modality) were labeled as *positive* instances for the extraction task; all other pairs were labeled as *negative* instances. Prior to generating features for machine learning, each instance was pre-processed. Numbers (e.g. “1”, “34”, “5.2”, etc.) were replaced by the string “num” to make them more meaningful to a learning algorithm across instances. This allowed the algorithm to associate numerical references with each other using a general pattern, instead of learning phrases with specific numbers (e.g. the phrase “num mg” may be significant, whereas “10 mg” may be less significant). Similarly, to abstract away from specific names, the names of drug products, active ingredients, and metabolites in each statement were replaced by the string “drugname”. This forces the learning algorithm to generalize over the participants of interactions, preventing it from identifying interactions based on the identity of the participants.

In the baseline dataset, each instance’s pre-processed sentence text was translated to bigrams using TagHelper, a text analysis program written on top of the Weka machine learning software (Hall et al. 2009; Rosé et al. 2008). Bigrams are a comprehensive set of consecutive word pairs that appear in a sentence. Words in bigrams were stemmed by TagHelper to facilitate learning more general concepts conveyed by phrases. For example, the commonly occurring phrases “increases

auc” and “increased auc” are stemmed to “increase auc” and then merged to the bigram. The baseline set of instances was loaded into Weka and three models were built using three different machine learning algorithms. The three algorithms were a rule learner (“JRip”), a decision tree (“J48”), and an SVM algorithm (“SMO”). Algorithm parameters were left at Weka defaults and 10-fold cross-validation was used to develop each model.

Exploration of Weka predictions from the baseline dataset showed that a major source of confusion for the machine learning algorithms was an inability to distinguish between pairs of drugs that do and do not interact within the same sentence. A frequent source of this kind of occurrence in the package insert text was *coordinate structures* such as “Drug A interacts with Drugs B and C”, where “B and C” is a coordinate structure. For such sentences, the baseline dataset contains the interacting pairs (A,B) and (A,C), along with the non-interacting pair (B,C). However, because all three pairs are represented by the same set of bigrams, it is obvious that information from bigrams alone is insufficient to distinguish which pairs interact and which simply co-occur within the sentence.

Another problem was that of multiple mentions of the same drug within an instance’s sentence span, as, for example, in the sentence “Co-administration of A and B leads to increased AUC levels for B.” Because the annotators had identified only one drug mention per annotated interaction, the algorithms incorrectly considered other mentions of the same drug as part of a non-interacting pair. Two solutions were implemented to help alleviate these problems. First, the dataset was condensed to a set of instances with unique drug pairs and sentence spans. If any of the baseline instances contributing to the condensed instance contained interactions, the condensed instance was said to contain an interaction. In this way, multiple drug mentions within a sentence span containing an interaction would translate to a single instance representing an interaction between the two drugs.

Second, two natural language dependency parsers were used to extract extra features from the sentence text for each instance: the Stanford NLP Parser (Klein & Manning 2003) and ClearParser (Choi 2011). Following approaches to relation extraction proposed in other domains e.g., (Bunescu & Mooney 2005), the dependency structure produced by each parser was searched for the

shortest path between the pair of drug mentions of the instance. The words on this path were stemmed using the Stanford NLP Tools stemmer (Stanford NLP 2011), and added to the dataset as the instance’s “syntactic path”.

Once a statement is classified as describing a PK DDI between two drugs, it is important to know if there is an observed effect or a lack of effect between two coadministered drugs (i.e., *positive vs negative* modality statements). To present the learning algorithms with the most relevant training data, modality prediction was treated as a separate task from interaction prediction. Development and test sets were created in the same manner as for interaction prediction, however instances that did not represent interactions were excluded. Only bigram features were used for modality prediction. Model training and testing proceeded in the same manner as for interaction prediction.

### 3 Results

A total of 208 multi-sentence sections were extracted from 64 PIs. Prior to consensus, inter-annotator agreement between the two annotators on PK DDI, active ingredient, drug product, metabolite mentions and was found to be 60%, 96.3%, 99.5%, and 60.8% respectively. The majority of disagreements about DDIs were due to a tendency of one annotator to incorrectly annotate some pharmacodynamic DDIs as PK DDIs. Also, one annotator incorrectly assumed that all metabolites had been pre-annotated and so did not actively attempt to annotate metabolite entities. These and other minor issues were corrected and full consensus was reached by both annotators. The final drug package insert PK DDI corpus contains 592 PK DDIs, 3,351 active ingredient mentions, 234 drug product mentions, and 201 metabolite mentions.<sup>2</sup>

Tables 1 and 2 provide more details on the modality and drug types present in the 592 consensus PK DDI statements. Table 1 shows that 388 state-

Modality	Interaction Type		Total
	Qualitative	Quantitative	
Negative	202	2	204
Positive	186	202	388
Total	388	204	592

Table 1. PK DDI statement modality shown by interaction type.

<sup>2</sup> <http://purl.org/NET/nlprepository/PI-PK-DDI-Corpus>

ments indicated that a PK DDI would occur between a drug pair, while 204 statements indicated that an interaction would not occur. The table also shows that 204 statements reported quantitative measures while 388 did not. Table 2 shows that the majority (86%) of PK DDI statements reported interactions by stating the two active ingredients involved in the DDI, with a much smaller proportion using a drug product in the description. Also, 35 DDI statements reported an effect on a drug metabolite.

A total of 11,048 PI instances were generated for the baseline dataset. This was reduced to 5,015 instances after condensing the instances down to unique drug pairs and sentence spans. In the final dataset, about a third of instances were drug pairs within the same sentence (1,583). The rest were split between drug pairs in adjacent sentences (1,717), and drug pairs with two sentences of separation (1,715). The dataset included 542 interactions of which 493 included the drug pair within a single sentence. 355 interactions were positive modality and 187 negative; 360 were qualitative, 182 quantitative. 1,636 instances were categorized as “new” based on drug release data while 3,379 were classified as “old”.

Results for interaction and modality prediction are shown in Table 3. For both the interaction and modality prediction tasks, the SVM algorithm (SMO) outperformed the rule learner (Jrip) and decision tree (J48). On the test set which was not used in training, the SVM classifier identified PK DDIs with an F-measure of 0.859 vs 0.762 for the rule learner and 0.802 for the decision tree algorithm. All algorithms performed quite well on the modality classification task but the SVM algorithm performed best with an F-measure of 0.949 vs 0.929 (rule learner) and 0.917 (decision tree).

### 4 Discussion

The automatic identification of DDIs in unstructured text is a topic that is gaining much interest. This work makes an important contribution to the field by being the first to demonstrate that machine learning can be applied quite effectively to the task of extracting PK DDIs from FDA-approved PIs.

Precipitant Type	Object Type			Total
	Active ingredient	Drug product	Metabolite	
Active ingredient	506	14	34	554
Drug product	37	-	1	38
Total	543	14	35	592

Table 2. A summary of consensus annotated PK DDIs by precipitant and object type.

As our work focuses on extracting PK DDIs, it is most similar to that of Karnik *et al.* (Karnik et al. 2011) who explored the performance of an “all paths” graph kernel (Airola et al. 2008) on a corpora of PK DDIs derived from 219 MEDLINE abstracts. The best performing algorithm in their experiments had an F-measure of 0.658 which is considerably less than the F-measure of 0.859 that our SVM achieved. However, the two results are not directly comparable because of unknown differences between the corpora. For example, it may be that PIs use more standard language patterns to report PK DDIs than what is found in MEDLINE abstracts. In future work we will explore how well the SVM algorithm performs over MEDLINE abstracts and contrast any differences between the two DDI sources that might affect NLP.

The only other project we are aware of that focused explicitly on extracting PK DDIs from unstructured text is that of Tari *et al.* (Tari et al. 2010), who evaluated a rule-based algorithm for extracting PK DDIs from papers and abstracts in the scientific literature. In this study the authors distinguished between *explicit* DDIs (statements indicating a direct observation of a PK effect from a give drug combination) and *implicit* DDIs (DDIs that can be inferred based on claims about drug metabolic properties extracted from scientific texts). The algorithm was ran over ~17 million MEDLINE abstracts and the output DDIs were compared with a reference standard set of 494 DDIs identified manually from 265 DrugBank drug pages. The algorithm’s recall of DrugBank interactions was only 12%. However, a manual inspection of the results found that 78% of the DDIs extracted by the algorithm were valid based on the source texts, even though they were not present in their reference standard. These results are important because they suggest that the set of DDIs present in DrugBank are *incomplete* and highlight the need for corpora derived from other text sources such as the one we developed from drug PIs for this study.

A larger body of research exists for the task of extracting DDIs of any type (i.e., PK *or* pharmacodynamic DDIs). Ten research papers were presented at the recent “Challenge Task on Drug-Drug Interaction Extraction” held at the 2011 SemEval Conference (Segura-Bedmar, Martinez & Sanchez-Cisneros 2011). All systems in this challenge were tested against the “DrugDDI corpus”; a set of 579 documents from the DrugBank database with 3,160 manually-annotated DDIs (Segura-Bedmar, Martinez & Pablo-Sanchez 2010). The best performing system in this challenge utilized an ensemble learning approach (Thomas et al. 2011) and produced an F-measure of 0.657. The second best performing method utilized composite kernels, a method that combines feature-based and kernel-based methods, and was found to perform with an F-measure of 0.64 (Chowdhury et al. 2011). Airola *et al.*’s “all paths” graph kernel (mentioned above) performed much more poorly on the DrugDDI corpora than on the Karnik’s PK-DDI corpus (F-measure 0.16 vs 0.658). The authors note that there were significant differences between in the two corpora with regards to the length and complexity of the sentences reporting DDIs .

To the best of our knowledge, only one other NLP study that has focused specifically on drug interactions reported in drug product labeling (Rubricchi & Quaglini 2012). The investigators compared the ability of an SVM classifier and a conditional random fields (CRF) classifier for assigning 13 semantic labels to Italian language text present in the interaction section of “Summary of Product Characteristics” documents (the Italian equivalent of PIs). The investigators explored the influence of a range of features on classifier performance, including orthographical, neighboring word, syntactic, parts of speech, and dictionary features. When all features were employed, the SVM had slightly better performance than the CRF classifier (micro-averaged F-measure: 91.41 vs 91.13, macro-averaged F-measure: 84.99 vs 80.83).

Model (dataset)	Jrip			J48			SMO		
	Prec	Recall	F	Prec	Recall	F	Prec	Recall	F
Baseline (development)	0.588	0.656	0.62	0.584	0.573	0.578	0.639	0.677	0.658
Stanford Parser (development)	0.762	0.68	0.719	0.809	0.804	0.807	0.851	0.815	0.833
ClearParser (development)	0.787	0.793	0.79	0.822	0.791	0.806	0.828	0.887	0.856
Stanford Parser (test)	0.778	0.665	0.717	0.828	0.832	0.83	0.843	0.838	0.84
ClearParser (test)	0.764	0.76	0.762	0.85	0.76	0.802	0.836	0.883	0.859
Modality (test)	0.963	0.897	0.929	0.887	0.948	0.917	0.941	0.957	0.949

Table 3. Results for interaction prediction on the baseline, development, and blind test set. Also shown are results for modality prediction for the blind test set (results over the development set are similar but not shown).

One key difference between the Rubrichi study and ours is that the task of tagging unstructured text with semantic elements that describe a DDI is not the same as classifying whether or not a statement containing a drug pair is reporting a DDI between the drugs. The difference is especially apparent when considering coordinate structures such as “Drug A interacts with Drugs B and C”, Semantic tagging would be useful for identifying the drug entities but is not useful (on its own) for identifying which of the three drug pairs interact with each other.

It is interesting to note that most recent work on DDI extraction had not made the distinction between PK and pharmacodynamic DDIs that is standard in the fields of pharmacology and pharmacy. This distinction might be relevant to DDI extraction because the two types of interactions are discovered in distinct ways that might lead to significant differences in how they are described in scientific documents. For example, there is a fairly standard set of *in vitro* experiments and clinical trials that have been a routine part of drug development for more than a decade (FDA, 1999). The same is not true for pharmacodynamic DDIs, which are more challenging to study because they involve additive and synergistic effects that are not necessarily related to a drug’s dose or clearance. Since it is reasonable that the methods used to investigate a DDI strongly influences its description, we think future work should examine if PK and pharmacodynamic DDI descriptions are different enough to warrant distinct DDI extraction efforts.

An error analysis of the final dataset suggested some reasons for cases where the machine learning algorithms misclassified instances. Instances that were not interactions, but were classified as such, contained a large number of sentences with de-

scriptions of studies or biochemical processes and measurements. These types of statements may share a number of features with actual interactions (e.g. numerical data, changing levels of drug, etc.) without containing an interaction. There also remain cases where several drug names occur and the classifiers were unable to differentiate between the interacting pair and non-interacting pairs. Unfortunately, no such clear pattern was apparent for instances that described interactions, but were classified as containing no interaction statement. A number of large sentences were observed in these instances, suggesting sentence complexity may play a role, increasing the difficulty of natural language parsing.

Analysis of the attribute weights assigned by the SVM algorithm (SMO) after training for interaction prediction shows some commonality regardless of whether the data was processed by the Stanford Parser or the ClearParser. For example, 19 out of the 20 most significant features identified by the algorithm from the dataset when processed by the Stanford Parser were words on the syntactic path; one less than when the dataset was processed by the ClearParser. Common significant features include words such as “coadminister”, “auc”, “pharmacokinetic”, and “absorption”. The algorithm placed greater importance on the words “increase” and “decrease” when the dataset was processed by the Stanford Parser, while the words “reduce” and “enhance” received greater attribute weights when the data was processed by the ClearParser. A similar analysis of the SVM algorithm developed for PK DDI modality prediction shows that bigrams with the words “no” or “not” are clearly the features of most importance to the model.

We also note that the algorithm's performance on the test set of PI statements is very similar to the algorithm's performance over the development set (see Table 3). We think that this finding is largely due to the careful stratification approach we used when creating the development and test sets. It might also be possible that the features in the unstructured PI text do not vary greatly between PIs regardless of their age. However, Table 2 shows that our PK DDI corpora had considerable variation in terms of quantitative vs qualitative and positive vs negative DDI statements. Thus, we anticipate that the SVM algorithm's performance will be maintained when ran against a much larger PI corpus and future work will test how well the algorithm generalizes to other sets of PIs.

## 5 Conclusion

We created a new, publically available, corpus of FDA-approved drug PI statements that have been manually annotated for PK DDIs by a pharmacist and a drug information expert. Also, we evaluated three different machine learning algorithms for their ability to 1) identify PK DDIs in drug PIs and 2) classify PK DDI statements by their modality (i.e., whether they report a DDI or no interaction between drug pairs). Experiments found that an SVM algorithm performed best on both tasks with an F-measure of 0.859 for PK DDI identification and 0.949 for modality assignment. We found that the use of syntactic information is very helpful for addressing the problem of sentences containing both interacting and non-interacting pairs of drugs. The strong performance of our algorithm for PK DDIs suggests that approaching pharmacokinetic and pharmacodynamic interactions as different NLP tasks is a potentially promising approach for advancing automated DDI extraction. Given the marked difference in performance between our extraction methods and previous work, we are planning further experiments to establish whether this difference reflects the comparative simplicity of the extraction task represented by our corpus, some specific strength of the applied extraction methods, or some other factor.

## Acknowledgement

This project was funded by grant K12-HS019461 from the Agency for Healthcare Research and

Quality (AHRQ). The content is solely the responsibility of the authors and does not represent the official views of AHRQ. We also thank John Horn, PharmD (University of Washington) and Mr. Rob Guzman (University of Pittsburgh) for their work annotating the corpus and identifying related research.

## References

- Airola, Antti, Sampo Pyysalo, Jari Björne, Tapio Pahikkala, Filip Ginter & Tapio Salakoski. 2008. All-paths graph kernel for protein-protein interaction extraction with evaluation of cross-corpus learning. *BMC Bioinformatics* 9(Suppl 11). S2. doi:10.1186/1471-2105-9-S11-S2 (3 May, 2012).
- Bunescu, Razvan C. & Raymond J. Mooney. 2005. A shortest path dependency kernel for relation extraction. *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 724–731. (HLT '05). Stroudsburg, PA, USA: Association for Computational Linguistics. doi:10.3115/1220575.1220666. <http://dx.doi.org/10.3115/1220575.1220666> (2 May, 2012).
- Choi, Jinho. 2011. ClearParser GoogleCode page. *clearparser*. <http://code.google.com/p/clearparser/> (10 December, 2011).
- Chowdhury, Md. Faisal Mahub, Asma Ben Abacha, Alberto Lavelli & Pierre Zweigenbaum. 2011. Two Different Machine Learning Techniques for Drug-Drug Interaction Extraction. *1st Challenge task on Drug-Drug Interaction Extraction (DDIExtraction 2011)*, 19–26. Huelva, Spain.
- Dal-Ré, R., A. Pedromingo, M. García-Losa, J. Lahuerta & R. Ortega. 2010. Are results from pharmaceutical-company-sponsored studies available to the public? *European Journal of Clinical Pharmacology* 66(11). 1081–1089. doi:10.1007/s00228-010-0898-y (5 August, 2011).
- FDA. 1999. *FDA Guideline: In Vivo Drug Metabolism/Drug Interaction Studies – Study Design, Data Analysis, and Implications for Dosing and Labeling*. Rockville, MD: Food and Drug Administration. <http://www.fda.gov/downloads/Drugs/Guidance/ComplianceRegulatoryInformation/Guidances/ucm072119.pdf>.
- FDA. 2010. CFR - Code of Federal Regulations Title 21.

- <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=201.57> (7 June, 2011).
- Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann & Ian H Witten. 2009. The WEKA data mining software: an update. *SIGKDD Explorations* 11(1). 10–18.
- Jonquet, Clement, Nigam H Shah & Mark A Musen. 2009. The open biomedical annotator. *Summit on Translational Bioinformatics 2009*. 56–60. (10 December, 2011).
- Karnik, Shreyas, Abhinata Subhadarshini, Zhiping Wang, Luis M Rocha & Lang Li. 2011. Extraction Of Drug-Drug Interactions Using All Paths Graph Kernel. *1st Challenge task on Drug-Drug Interaction Extraction (DDIExtraction 2011)*. Huelva, Spain.
- Klein, Dan & Christopher D Manning. 2003. Fast Exact Inference with a Factored Model for Natural Language Parsing. (Ed.) S Thrun S Becker & Keditors Obermayer. *Science* 15. 3–10.
- Marroum, P.J. & J. Gobburu. 2002. The product label: how pharmacokinetics and pharmacodynamics reach the prescriber. *Clinical Pharmacokinetics* 41(3). 161–169. (7 June, 2011).
- Nelson, Stuart J, Kelly Zeng, John Kilbourne, Tammy Powell & Robin Moore. 2011. Normalized names for clinical drugs: RxNorm at 6 years. *Journal of the American Medical Informatics Association: JAMIA* 18(4). 441–448. doi:10.1136/amiajnl-2011-000116 (10 December, 2011).
- Ogren, Philip V. 2006. Knowtator: a Protégé plug-in for annotated corpus construction. *Proceedings of the 2006 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, 273–275. Morristown, NJ, USA: Association for Computational Linguistics. doi:<http://dx.doi.org/10.3115/1225785.1225791>.
- Rosé, Carolyn, Yi-Chia Wang, Yue Cui, Jaime Arguello, Karsten Stegmann, Armin Weinberger & Frank Fischer. 2008. Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning* 3(3). 237–271. doi:10.1007/s11412-007-9034-0 (10 December, 2011).
- Rubrichi, S & S Quaglini. 2012. Summary of Product Characteristics content extraction for a safe drugs usage. *Journal of Biomedical Informatics* 45(2). 231–239. doi:10.1016/j.jbi.2011.10.012 (3 May, 2012).
- Segura-Bedmar, Isabel, Paloma Martinez & Cesar Pablo-Sanchez. 2010. Extracting drug-drug interactions from biomedical texts. *Workshop on Advances in Bio Text Mining*, vol. 11 Suppl 5, 9. Madrid, Spaim: BMC Bioinformatics. <http://www.biomedcentral.com/1471-2105/11/S5/P9>.
- Segura-Bedmar, Isabel, Paloma Martinez & Daniel Sanchez-Cisneros (eds.). 2011. *Proceedings of the First Challenge Task: Drug-Drug Interaction Extraction 2011*. Huelva, Spain. <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-761/> (9 December, 2011).
- Stanford NLP. 2011. The Stanford NLP (Natural Language Processing) Group. <http://nlp.stanford.edu/software/> (10 December, 2011).
- Tari, Luis, Saadat Anwar, Shanshan Liang, James Cai & Chitta Baral. 2010. Discovering drug-drug interactions: a text-mining and reasoning approach based on properties of drug metabolism. *Bioinformatics (Oxford, England)* 26(18). i547–553. doi:10.1093/bioinformatics/btq382 (9 December, 2011).
- Thomas, Philippe, Mariana Neves, Illes Solt, Domonkos Tikk & Ulf Leser. 2011. Relation Extraction for Drug-Drug Interactions using Ensemble Learning. *1st Challenge task on Drug-Drug Interaction Extraction (DDIExtraction 2011)*, 11–18. Huelva, Spain.