

POSTER PRESENTATION

Open Access

Extracting drug-drug interactions from biomedical texts

Isabel Segura-Bedmar*, Paloma Martínez, César de Pablo-Sánchez

From Workshop on Advances in Bio Text Mining
Ghent, Belgium. 10-11 May 2010

Background

A drug-drug interaction (DDI) occurs when one drug influences the level or activity of another drug. The detection of drug interactions is an important research area in patient safety since these interactions can become very dangerous and increase health care costs. Although there are different databases supporting health care professionals in the detection of drug interactions,

this kind of resources is rarely complete. Drug interactions are frequently reported in journals of clinical pharmacology, making medical literature the most effective source for the detection of drug interactions. However, the increasing volume of the literature overwhelms health care professionals trying to keep an up-to-date collection of all reported DDIs. The development of automatic methods for collecting, maintaining and

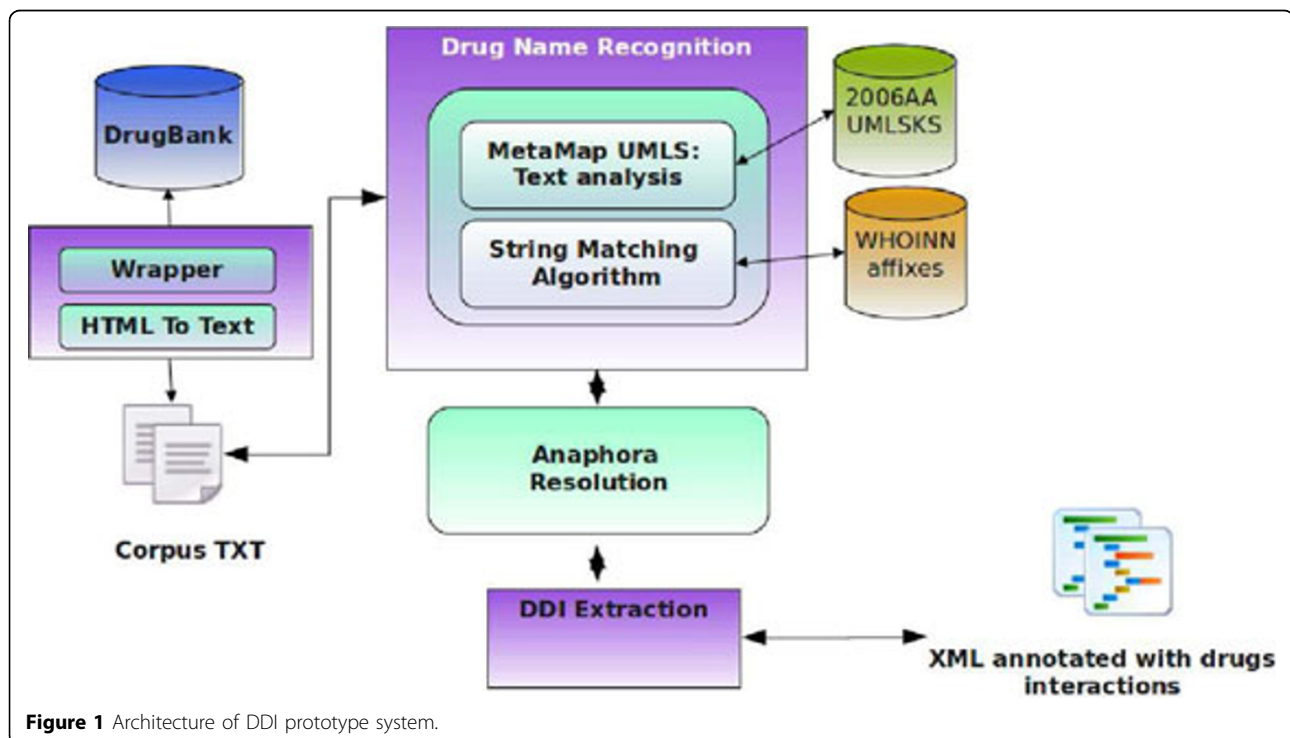


Figure 1 Architecture of DDI prototype system.

* Correspondence: isegura@inf.uc3m.es
Computer Science Department, University Carlos III of Madrid, Leganés,
28921, Spain
Full list of author information is available at the end of the article

interpreting this information is crucial for achieving a real improvement in their early detection. Information Extraction (IE) techniques can provide an interesting way of reducing the time spent by health care professionals on reviewing the literature. Nevertheless, no approach has been carried out to extract drug-drug interactions from biomedical texts.

Materials and methods

We have conducted a detailed study on various IE techniques applied to biomedical domain. Based on this study, we have proposed two different approximations for the extraction of DDIs from texts. The first approximation proposes a hybrid approach, which combines shallow parsing and pattern matching to extract relations between drugs from biomedical texts. A pharmacist defined a set of lexical patterns (12) to capture the various language constructions used to express DDIs in pharmacological texts. The second approximation is based on a supervised machine learning approach, in particular, a kernel-based approach that uses Support Vector Machines (SVM) presented by [1]. In addition, we have created and annotated the first corpus annotated with DDIs, DrugDDI, which allows us to evaluate and compare both approximations. The corpus consists of 579 documents from the DrugBank database and contains a total of 3,160 DDIs. We have also defined three auxiliary processes to provide crucial information, which are used by the aforementioned approximations. These auxiliary tasks are as follows: a process for text analysis based on the UMLS MetaMap Transfer tool (MMTx) to provide shallow syntactic and semantic information from texts, a process for drug name recognition and classification which is based on a set of nomenclature rules recommended by the World Health Organization, and a process for drug anaphora resolution. Finally, we have developed a pipeline prototype (see Figure 1) which integrates the different auxiliary processes. The pipeline architecture allows us to easily integrate these modules with each of the approaches proposed in this work: pattern-matching or kernels.

Results

While the first approximation based on pattern matching achieves low performance (Precision=48.7%, Recall=25.7%, F-measure=33.6%), the approach based on kernel-methods achieves better performance, especially better recall (Precision=55.1%, Recall=82.3%, F-measure=66.0%). The variability of natural language makes it difficult for our first approach to accurately detect all drug-drug interactions occurring in texts since sentences conveying the same relation may be lexically and syntactically composed in different ways. Inversely, sentences that are lexically common may not necessarily convey

the same relation. Therefore, the set of lexical patterns proposed by our pharmacist is not enough to identify many of the interactions. Performance achieved by the kernel-based approach is comparable to studies which have carried out a similar task such as the extraction of protein-protein interactions.

Conclusions

To the best of our knowledge, this work has proposed the first integral solution for the automatic extraction of DDI from biomedical texts. We hope that our proposal and the DrugDDI corpus contribute to the development of useful tools to assist healthcare professionals in the early detection of DDIs.

Acknowledgements

This work has been partially supported by the Spanish research projects: MAVIR consortium (MA2VICMR S2009/TIC-1542, <http://www.mavir.net>), a network of excellence funded by the Madrid Regional Government and TIN2007-67407-C03-01 (BRAVO: Advanced Multimodal and Multilingual Question Answering). The authors are grateful to María Segura Bedmar, manager of the Drug Information Center of the Mostoles University Hospital, Spain, for her valuable assistance in the annotation of the corpus and evaluation of the system.

Published: 6 October 2010

Reference

1. Giuliano C, Lavelli A, Romano L: **Exploiting shallow linguistic information for relation extraction from biomedical literature.** *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics (EACL-2006)* 2006, 5-7.

doi:10.1186/1471-2105-11-S5-P9

Cite this article as: Segura-Bedmar et al.: Extracting drug-drug interactions from biomedical texts. *BMC Bioinformatics* 2010 **11**(Suppl 5):P9.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

