DDIExtractor: A Web-Based Java Tool for Extracting Drug-Drug Interactions from Biomedical Texts

Daniel Sánchez-Cisneros, Isabel Segura-Bedmar, and Paloma Martínez

University Carlos III of Madrid, Computer Science Department
Avd. Universidad 30, 28911 Leganés, Madrid, Spain
{dscisner,isegura,pmf}@inf.uc3m.es
http://labda.inf.uc3m.es/

Abstract. A drug-drug interaction (DDIs) occurs when one drug influ-ences the level or activity of another drug. The detection of DDIs is an important research area in patient safety since these interactions can be-come very dangerous and increase health care costs. Although there are several databases and web tools providing information on DDIs to pa-tients and health-care professionals, these resources are not comprehen-sive because many DDIs are only reported in the biomedical literature. This paper presents the first tool for detecting drug-drug interactions from biomedical texts called DDIExtractor. The tool allows users to search by keywords in the Medline 2010 baseline database and then detect drugs and DDIs in any retrieved document.

Keywords: Drug-Drug Interactions, Biomedical Information Extraction.

1 Background

Drug-drug interactions are common adverse drug reactions and unfortunately they are a frequent cause of death in hospitals. The management of DDIs is a critical issue due to the overwhelming amount of information available on them. We think that Information Extraction (IE) techniques can provide an interesting way of reducing the time spent by health care professionals on reviewing the literature. There are some examples of IE applications in different biomedical subdomains. iHOP (information Hyperlinked Over Proteins)[1] [1] is a web service that automatically extracts key sentences from MedLine documents where genes, proteins and chemical compounds terms are annotated and linked to MeSH terms by machine learning methods. Another example is the Reflect Tool [2], which is a free service tag of gene, protein, and small molecule names in any web page. The NCBO Resource Index system[2] allows to annotate and index texts with ontology concepts from more than twenty diverse biomedical resources.

[1] www.ihop-net.org/
[2] http://www.bioontology.org/resources-index

This paper describes a web-based tool for searching relevant documents from Medline 2010 database and analyzing them in order to highlight drugs and DDIs occurring in them. Drug name recognition and DDIs detection are performed by the DrugDDI system [3] which is based on a Shallow Linguistic Kernels to extract relationships between entities. The paper is organized as follows: Section 2 describes the functionality and the architecture of this tool; and the on-going and future work is drawn in Section 3.

2 DDIExtractor: A Web Tool to Detect Drugs and DDI

To make the reading of the biomedical literature easier and faster, we have developed a web-based java tool called DDIExtractor[3]. DDIExtractor allows users to find relevant documents entring specific keywords queries. Then, the tool can process any retrieved document highlighting its drug names and DDIs. For each drug, the tool can also show a popup including its description, drug categories and drug interactions among other useful information from Drugbank databases[4]. The tool works as is shown in Figure 1.

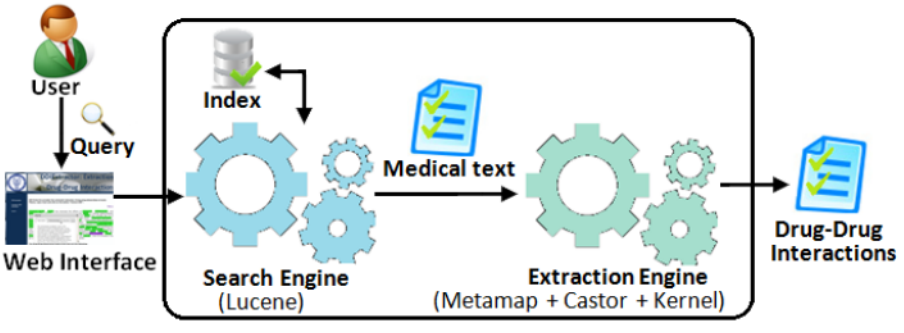

Fig. 1. Architecture of DDIExtractor

2.1 Search Engine

First, DDIExtractor allows searching relevant documents from the Medline 2010 database. This database contains a total of 18,502,916 records grouped by publication years. We used the Apache Lucene engine[5], which is a free open source information retrieval(IR) API originally implemented in Java.

This engine is a tradicional IR system that provide structured and unstructured information to the user, retriving a list of results related to the query searched (e.g.: Aspirin). Each result will specify its title, journal, authors, publication date, Medline identificator(PMID), etc. Then the user can click on a result to see the entire abstract or even to extract its drug-drug interactions.

[3] http://163.117.129.57:8080/ddiextractorweb/
[4] http://www.drugbank.ca/
[5] http://lucene.apache.org/

Fig. 2. The application highlights the drugs found in the text and shows a popup including related information for each drug

2.2 Extractor

To extract DDIs, the tool uses the DrugDDI system based on the tool java Simple Relation Extraction (JSRE)[6]. The following list shows the main steps to extract DDIs:

- Step 1: The tool allows users to select a text from the retrieved list of documents or introduce a new text.
- Step 2: The text is analyzed by the MMTx[7] tool which provides syntactic and semantic information. This information is converted into an input XML format compatible with our DrugDDI system. Then, the XML format is processed using the XML data binding framework Castor[8] in order to represent the information of the whole sentence as well as the local contexts around the drugs by their Part-of-Speech (PoS) tags, lemmas and drug names. A detailed description of this process can be found in [4]. Each pair of drugs in the same sentence is a possible DDI. Then, each candidate DDI is classified by the DrugDDI system in order to detect those pairs of drugs that are actually DDIs.
- Step 3: Drug names and DDIs are highlighted in texts using a different color according to their semantic types provided by the MMTx tool. For each

[6] http://hlt.fbk.eu/en/technology/jSRE
[7] http://metamap.nlm.nih.gov/
[8] http://www.castor.org/

drug identified, the tool can also show a popup including its definitions, drug categories and drug interactions among other useful information (see Figure 2).

3 On-Going and Future Work

DDIExtractor is an online tool that highlights drugs and DDIs from texts and represents a first step making the reading and interpretation of pharmacological texts easier and faster.

While several drug databases and web resources provide users structured information on DDIs, this tool retrieves structured and unstructured information in order to detect its DDIs.

Ideally, prescribed information about a drug should list its potential interactions, joined to the following information about each interaction: its mechanism, related doses of both drugs, time course, seriousness and severity, the factors which alter the individual's susceptibility to it, and the probability of its occurrence. However, in practice this information is rarely available. Our future work will focus on the improvement of our system for extracting DDIs, especially in the detection of these additional features. Also, we would like to design an evaluation methodology oriented toward user acceptance.

Acknowledgments. This work is supported by the projects MA2VICMR (S2009/TIC-1542) and MULTIMEDICA (TIN2010-20644-C03-01).

References

1. Hoffmann, R., Valencia, A.: Implementing the iHOP concept for navigation of biomedical literature. Bioinformatics 21(suppl. 2) (2005)
2. Pafilis, E., O'Donoghue, S., Jensen, L., Horn, H., Kuhn, M.: Reflect: augmented browsing for the life scientist. Nature Biotechnology 27(6), 508–510 (2009)
3. Segura-Bedmar, I., Martínez, P., de Pablo-Sánchez, C.: Using a shallow linguistic kernel for drug-drug interaction extraction. Journal of Biomedical Informatics (2011) (accepted)
4. Segura-Bedmar, I.: Application of information extraction techniques to pharmacological domain: extracting drug-drug interactions. PhD thesis, Universidad Carlos III de Madrid, Leganés, Madrid, Spain (2010)