

Chapter 1.25

Information Extraction in Biomedical Literature

Min Song

Drexel University, USA

Il-Yeol Song

Drexel University, USA

Xiaohua Hu

Drexel University, USA

Hyoil Han

Drexel University, USA

INTRODUCTION

Information extraction (IE) technology has been defined and developed through the US DARPA Message Understanding Conferences (MUCs). IE refers to the identification of instances of particular events and relationships from unstructured natural language text documents into a structured representation or relational table in databases. It has proved successful at extracting information from various domains, such as the Latin American terrorism, to identify patterns related to terrorist activities (MUC-4). Another domain, in the light of exploiting the wealth of natural language documents, is to extract the knowledge or information from these unstructured plain-text files

into a structured or relational form. This form is suitable for sophisticated query processing, for integration with relational databases, and for data mining. Thus, IE is a crucial step for fully making text files more easily accessible.

BACKGROUND

The advent of large volumes of text databases and search engines have made them readily available to domain experts and have significantly accelerated research on bioinformatics. With the size of a digital library commonly exceeding millions of documents, rapidly increasing, and covering a wide range of topics, efficient and automatic

extraction of meaningful data and relations has become a challenging issue. To tackle this issue, rigorous studies have been carried out recently to apply IE to biomedical data. Such research efforts began to be called biomedical literature mining or text mining in bioinformatics (de Bruijn & Martin, 2002; Hirschman et al., 2002; Shatkay & Feldman, 2003). In this article, we review recent advances in applying IE techniques to biomedical literature.

MAIN THRUST

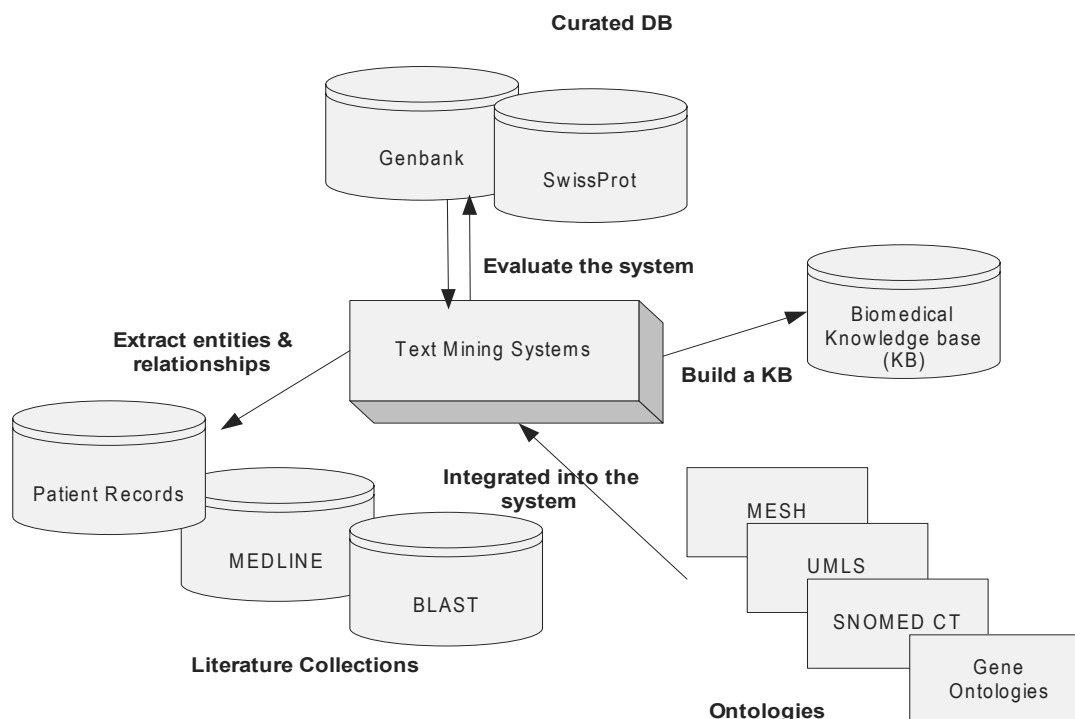
This article attempts to synthesize the works that have been done in the field. Taxonomy helps us understand the accomplishments and challenges in this emerging field. In this article, we use the following set of criteria to classify the biomedical literature mining related studies:

1. What are the target objects that are to be extracted?
2. What techniques are used to extract the target objects from the biomedical literature?
3. How are the techniques or systems evaluated?
4. From what data sources are the target objects extracted?

Target Objects

In terms of what is to be extracted by the systems, most studies can be broken into the following two major areas: (1) named entity extraction such as proteins or genes; and (2) relation extraction, such as relationships between proteins. Most of these studies adopt information extraction techniques using curated lexicon or natural language processing for identifying relevant tokens such as words or phrases in text (Shatkay & Feldman, 2003).

Figure 1. An overview of a biomedical literature mining system



6 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/information-extraction-biomedical-literature/26225

Related Content

Mobile Health Applications and New Home Care Telecare Systems: Critical Engineering Issues

Žilbert Tafa (2009). *Handbook of Research on Distributed Medical Informatics and E-Health* (pp. 305-324).

www.irma-international.org/chapter/mobile-health-applications-new-home/19942

Recordings of Impedance and Communication Between Defibrillator and Pacemaker Electrodes

Anders Jarløv, Anne Elisabeth Jarløvand Tim Toftgaard Jensen (2019). *International Journal of Biomedical and Clinical Engineering* (pp. 45-68).

www.irma-international.org/article/recordings-of-impedance-and-communication-between-defibrillator-and-pacemaker-electrodes/233542

Potential Evaluation of Electro Mechano Gram (EMG) for Osteoporosis Detection

Shashank Srivastava, Shipra Prakash, Suresh Bhalla, Alok Madan, Sunil Sharma, H. S. Chhabraand Jitesh S. Manghwani (2022). *International Journal of Biomedical and Clinical Engineering* (pp. 1-12).

www.irma-international.org/article/potential-evaluation-of-electro-mechano-gram-emg-for-osteoporosis-detection/309411

Privacy and Access to Electronic Health Records

Dick Whiddett (2009). *Medical Informatics: Concepts, Methodologies, Tools, and Applications* (pp. 1997-2005).

www.irma-international.org/chapter/privacy-access-electronic-health-records/26352

Compiling Medical Data into National Medical Databases: Legitimate Practice or Data Protection Concern?

Boštjan Bercicand Carlisle George (2008). *Ethical, Legal and Social Issues in Medical Informatics* (pp. 228-248).

www.irma-international.org/chapter/compiling-medical-data-into-national/18617