

## 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

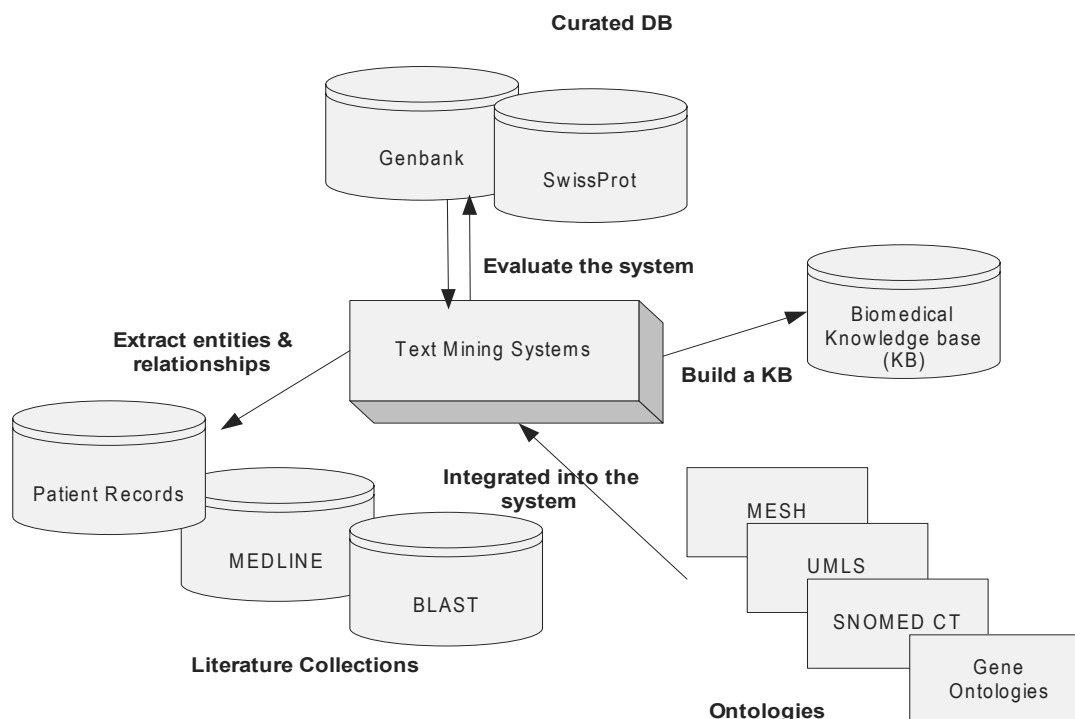
This article attempts to synthesize the works that have been done in the field. Taxonomy helps us understand the accomplishments and challenges in this emerging field. In this article, we use the following set of criteria to classify the biomedical literature mining related studies:

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](http://www.igi-global.com/chapter/information-extraction-biomedical-literature/26225)

## Related Content

---

### Electrical Conductivity of Skin Compared to Skin Perfusion Recordings

Anders Jarløvand Tim Toftgaard Jensen (2017). *International Journal of Biomedical and Clinical Engineering* (pp. 1-17).

[www.irma-international.org/article/electrical-conductivity-of-skin-compared-to-skin-perfusion-recordings/189117](http://www.irma-international.org/article/electrical-conductivity-of-skin-compared-to-skin-perfusion-recordings/189117)

### A Romance of the Three Kingdoms: Biotechnology Clusters in Beijing, Shanghai and Guangdong Province, China

Petr Hanel, Jie He, Jingyan Fu, Susan Reidand Jorge E. Niosi (2017). *Comparative Approaches to Biotechnology Development and Use in Developed and Emerging Nations* (pp. 67-129).

[www.irma-international.org/chapter/a-romance-of-the-three-kingdoms/169515](http://www.irma-international.org/chapter/a-romance-of-the-three-kingdoms/169515)

### Music Therapy for Dementia Patients: Tuned for Culture Difference

Yuki Tanaka, Hiroki Nogawaand Hiroshi Tanaka (2011). *Early Detection and Rehabilitation Technologies for Dementia: Neuroscience and Biomedical Applications* (pp. 257-279).

[www.irma-international.org/chapter/music-therapy-dementia-patients/53447](http://www.irma-international.org/chapter/music-therapy-dementia-patients/53447)

### Development of Tactile Display Devices Using fMRI under High Magnetic Fields

Masayuki Kitazawa (2011). *Early Detection and Rehabilitation Technologies for Dementia: Neuroscience and Biomedical Applications* (pp. 287-292).

[www.irma-international.org/chapter/development-tactile-display-devices-using/53449](http://www.irma-international.org/chapter/development-tactile-display-devices-using/53449)

### Innovative Hospital Management: Tracking of Radiological Protection Equipment

Holger Fritzsche, Elmer Jeto Gomes Ataide, Afshan Bi, Rohit Kalva, Sandeep Tripathi, Axel Boese, Michael Friebeand Tim Gonschorek (2020). *International Journal of Biomedical and Clinical Engineering* (pp. 33-47).

[www.irma-international.org/article/innovative-hospital-management/240745](http://www.irma-international.org/article/innovative-hospital-management/240745)