

Biyomedikal Metinlerde Bilgi Keşfi

Pınar YILDIRIM^{a,1}, Antonio Jose Jimeno YEPES^b, Osman SAKA^c, Dietrich Rebholz SCHUHMANN^b

^a *Bilgisayar Mühendisliği Bölümü, Çankaya Üniversitesi, Ankara*

^b *EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK*

^c *Biyostatistik ve Tıp Bilişimi AD, Akdeniz Üniversitesi, Antalya*

Knowledge Discovery in Biomedical Texts

Abstract. The scientific biomedical literature reports on clinical studies and medical research. The literature is increasingly processed automatically to answer specific medical problem and to discover relevant but unnoticed information. Automatic knowledge identification from the literature requires biomedical text mining solutions that consist of information retrieval information extraction techniques and employ natural language processing techniques. In this study, text mining solutions were applied to Medline abstracts to analyse in detail the treatment of liver-specific parasites. In the first step, Medline abstracts have been tagged with species mention (Entrez NCBI Taxonomy) and all abstracts have been selected that report on parasites leading to damages in the liver tissue according to the relevance feedback from medical doctors. Parasites with the highest occurrence frequency have been selected for further analysis. In the next step, a systematic analysis of treatment mentions from the context of the Medline abstracts was performed (drug names from Drugbank) and time series for the treatment of each parasite with different drugs was calculated. In addition, drugs' therapy classes were used to distinguish anthelmintic and antiprotozoal drugs. The time series show how drug treatments evolve over time, leading to the replacement of existing treatments by newer drugs and possibly demonstrating the emergence of resistance against well established drugs. We hope that these analyses allow both healthcare providers and pharmaceutical initiatives to see some changes in drug usage over time and compare them in different time periods.

Keywords. Biomedical Text Mining, Time Series Analyses, Liver, Parasites.

Özet. Bilimsel biyomedikal literatürler klinik çalışmaları ve tıbbi araştırmaları raporlarlar. Literatürler özel tıbbi problemi ve ilişkili ama fark edilmemiş bilgileri keşfetmek için yoğun otomatik işlemlere tabi tutulurlar. Bu literatürden otomatik bilgi tanımlanması bilgi çekip getirme (information retrieval), bilgi çıkarma yöntemleri (information extraction techniques) ve doğal dil işleme (natural language processing) gibi yöntemleri içeren metin madenciliği çözümlerini gerektirir. Bu çalışmada, karaciğerle ilgili parazitlerin tedavisindeki detayları analiz etmek için metin madenciliği çözümleri Medline özetlerine uygulanmıştır. İlk aşamada, Medline özetleri canlı isimleri (Entrez NCBI Taxonomy) ile

¹ Sorumlu Yazar: Çankaya Üniversitesi, Mühendislik Mimarlık Fakültesi Bilgisayar Mühendisliği Bölümü, Öğretmenler Cd. No:14 06530 Ankara/Türkiye; E-posta adresi: pinar@cankaya.edu.tr.

işaretlenmiştir ve tıp doktorlarından alınan görüşler doğrultusunda karaciğer dokusuna zarar veren parazitlere ait bütün özetler seçilmiştir. En yüksek frekanslı parazitler daha detaylı analiz için seçilmiştir. Bir sonraki aşamada, Medline özetlerinin içeriğinden tedaviye yönelik sistematik bir analiz gerçekleştirilmiştir (Drugbank'dan ilaç isimleri) ve her bir parazitin farklı ilaçlarla tedavisine ait zaman serisi analizi oluşturulmuştur. Ayrıca ilaçların tedavi sınıfları anthelmintic ve antiprotozoal ilaçları ayırt etmek için kullanılmıştır. Zaman serileri ilaç tedavilerinin zamanla nasıl geliştiğini, yeni ilaçların var olan ilaçların yerini almasını ve kabul görmüş ilaçlara karşı muhtemel direnç geliştirilmesini göstermektedir. Biz umuyoruz ki bu analizler hem sağlık hizmeti sunan kişiler ve hem de ilaç girişimcileri için zaman içinde ilaç kullanımındaki değişiklikleri görmelerine ve bunların farklı zaman dönemlerinde karşılaştırılabilmelerine olanak sağlar.

Anahtar Kelimeler. Biyomedikal Metin Madenciliği; Zaman Serisi Analizi; Karaciğer; Parazitler.