

Yapılandırılmamış Kulak Burun Boğaz Epikriz Notlarının Yapılandırılmış Formata Dönüştürülmesi

Başak OĞUZ^a, Uğur BİLGE^a, M. Kemal SAMUR^a

^a Biyoistatistik ve Tıp Bilişimi AD, Akdeniz Üniversitesi, Antalya

Transforming Unstructured Otolaryngology Discharge Notes Into A Structured Form

Abstract

The explosive growth of databases in almost every area of human activity has created a great demand for new, powerful tools for turning data into useful knowledge. Text mining is a tool that saw an increasing interest in the 2000s, for enabling people to find unknown information and facts from the free-text data. Information extraction is one of the effective approach to text mining that transforms the unstructured text to the structured form. Much of the information that exists in patient clinical records is difficult to access, as they are often in unstructured text formats. Accessing accurate data and information about patients is important, and the time that physicians spend while making decisions about patients is valuable.

The aim of this study is to develop a domain based system to transform unstructured otolaryngology discharge notes to structured forms to make easy access, search and information extraction. Firstly, the discharge notes, gathered from Otolaryngology department, are applied preprocessing steps (separating texts into tokens by using whitespaces, removing the punctuation marks (periods, comma, etc.) from the texts) to decrease the dimension and improve performance of the system. Secondly, in order to identify common section in the discharge notes, including patient history, age, problems, and diagnosis etc., several word lists was constituted. Only predefined words and word groups exist in the problems section, can be tagged by the system, due to it being work in progress. After word lists are completed for each predefined sections and the system is able to tag them in the texts, we will place them into a database and then use traditional data mining tools to identify patterns in this extracted data.

Key Words:

Otolaryngology; Text Mining; Information Extraction

Özet

Veritabanlarının, neredeyse insan aktiviteleriyle ilgili tüm alanlarda çok büyük oranlarda büyümesi, veriyi kullanışlı bilgiye dönüştüren yeni ve güçlü araçlara talep yaratmıştır. Metin madenciliği, özellikle 2000 yılından sonra daha fazla ilgi gören, serbest formatta bulunan metinlerin içindeki daha önceden bilmediğimiz bilgileri ortaya çıkarmamızı sağlayan işlemler

bütünüdür. Bilgi çıkarma sistemleri ise yapılandırılmamış metinleri yapılandırılmış formata dönüştüren, metin madenciliği için geliştirilen etkili yaklaşımlardan biridir. Hasta klinik kayıtlarında bulunan bilginin büyük çoğunluğu genellikle yapılandırılmamış metin formatında bulunduğu için bu bilgilere erişim zordur. Hastalarla ilgili doğru veri ve bilgiye erişmenin ve hekimlerin hasta ile ilgili kararlar verirken harcadığı zamanın önemi göz önünde bulundurulduğunda tıp alanında bu tür araçlara olan ihtiyaç artmıştır.

Bu çalışmada, bilgiye erişimi kolaylaştırmak için yapılandırılmamış formatta bulunan Kulak Burun Boğaz (KBB) epikriz notlarını yapılandırılmış formata dönüştüren, alan bilgisine dayalı olarak geliştirilmekte olan bir sistem anlatılmaktadır. Sistemin performansını artırmak ve girdi metinlerinin boyutunu azaltmak için KBB'den alınan epikriz notlarında bazı ön işlem adımları (kelime gövdeleme, boşluk kullanılarak metin içindeki kelime ve kelime gruplarını ayırma, noktalama işaretlerini kaldırma vb.) uygulanmıştır. Girdi metinlerindeki hikaye, yaş, şikayet gibi genel bölümleri tanımlamak için çeşitli kelime listeleri oluşturulmuştur. Fakat, sistem geliştirilme aşamasında olduğu için sadece "şikayet" bölümündeki, önceden tanımlanan kelime ve kelime grupları sistem tarafından etiketlenmektedir. Kelime listeleri tamamlandıktan ve bu listelerdeki kelime ve kelime grupları sistem tarafından etiketlenmeye başladıktan sonra, elde edilen verilerin geliştirilen bir veritabanına aktarılması ve bu veriler üzerinde geleneksel veri madenciliği yöntemlerinin uygulanması planlanmaktadır.

Anahtar Kelimeler:

Kulak Burun Boğaz; Metin Madenciliği; Bilgi Çıkarma

Sorumlu Yazarın Adresi

Başak Oğuz, Akdeniz Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Dumlupınar Bulvarı, Kampus, Antalya 07059 E-posta: basakoguz@akdeniz.edu.tr