

Abdominal USG Raporlarında Bilgi Kalıpları

Ergin SOYSAL^a, Nazife BAYKAL^b

^a Tıp Eğitimi ve Bilişimi AD, Ankara Üniversitesi Tıp Fakültesi, Ankara

^b Enformatik Enstitüsü, Ortadoğu Teknik Üniversitesi, Ankara

Information Patterns in Abdominal USG Reports

Abstract

Purpose: Natural language is the oldest and natural route of information storage and sharing. This endlessly flexible form of storage severely cripples machine evaluation of this information. Aim of this study is to examine the patterns encapsulating the information in abdominal USG reports in the form of free text.

Material and Method: 756 reports of abdominal USG examination performed in a period of two years in Radiodiagnostic Department of Ankara University Medical School were analyzed at words, phrases and sentence patterns levels.

Results: Each report is composed of 15.18 ± 0.25 sentences and 107.23 ± 1.8 words. 7% of all words was misspelled. 92.8% of all misspelled words contain only one simple error. Reports consist of 578 different words stems, of those, organ names, organ regions, property names, verbs and adjectives constitute 10.73%, 20.33%, 9.6%, 9.6%, 16.95% respectively. Three major sentence patterns are detected, expressing the entities, properties of entities and relations among them.

Conclusion: Radiological reports are considered as free text from computational aspect. However, information is presented in limited number of distinct grammatical patterns. From this point of view, these reports are machine processible to a larger extent.

Key Words

Natural language; Radiodiagnostic reports; Language patterns

Özet

Amaç: Doğal dil, enformasyon depolama ve paylaşımının en eski yoludur. Bu sınırsız esneklikteki depolama formatı, bu bilginin bilgisayar tarafından değerlendirilmesini oldukça zor bir hale getirir. Bu çalışmanın amacı, serbest metin formatındaki abdominal USG raporlarında enformasyonu barındıran kalıpları incelemektir.

Materyal ve Metod: Ankara üniversitesi Tıp Fakültesinde son iki yılda yapılan 756 abdominal USG incelemesine ait raporlar; kelime, deyim ve cümle seviyelerinde incelendi.

Bulgular: Her rapor $15,18 \pm 0,25$ cümle ve $107,23 \pm 1,8$ sözcükten meydana geliyordu. Sözcüklerin yazılımda 7% oranında hata mevcuttu. Bu hatalı kelimelerin %92,8'i tek bir basit hata içermektedir. Raporlar, 578 değişik kökten gelişen kelimeler içermektedir: organ adları (%10,73), organ bölgeleri (%20,33), nitelik isimleri (%9,6), fiiller (%9,6) ve sıfatları (%16,95). Organlar, nitelikler ve bunlar arasındaki ilişkileri ifade etmekte kullanılan üç majör cümle kalıbı tespit edildi.

Sonuç: Radyoloji raporları bilgisayar ile işleme açısından serbest metin olarak düşünülür. Ancak, enformasyon, sınırı sayıdaki gramer kalıpları içinde sunulmaktadır. Bu bakış açısından, bu raporlar geniş ölçüde bilgisayar tarafından işlenebilir durumdadır.

Anahtar Kelimeler

Doğal dil, Radyodiyagnostik raporları, Dil kalıpları