Electronic Thesaurus to Recover Information on Mammography

Paulo Roberto Barbosa Serapião, Paulo Mazzoncini de Azevedo Marques

Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo - Brazil

Abstract and Objective

The goal of this study is to propose, investigate and develop a system which can represent electronic knowledge –thesaurus- in Portuguese in order to facilitate mammography report reading for the specialist and also to improve the recovery of existing information in them. In order to do so, the technique of finding out information in databases, the thesaurus structure and development, and web language were employed.

Keywords:

Thesaurus, Mammography, Text mining

Introduction

The Brazilian medical practice, which is based on Portuguese, has difficulties using information patterns, which are usually in English, in its daily tasks. A huge national effort has been made in order to translate such patterns. There is also difficulty adapting the use of such patterns throughout the country. These are some examples of information patterns mentioned above: RadLex™, Bi-Rads™, Menelas®, and OpenGalen® among others. This article presents the research, which tries to solve that negative status of the Brazilian "lusofona" Community.

Methods

A detailed conceptual analysis of mammography reports from the Radiology Information System at the Ribeirão Preto Medical School Hospitals das Clinicas (RIS – HCFMRP) from 01/03/2000 to 06/03/2006 was performed to create a Bi-Rads™ thesaurus; after that it was possible to access 22,535 definitive reports. The following methodological division was made in order to have a longitudinal view of the basis and a similar comprehension of RIS information:

01/03/2000 – 03/03/2000 (430 reports)
01/03/2002 – 03/03/2002 (576 reports)
01/03/2004 – 03/03/2004 (622 reports)
01/03/2006 – 03/03/2006 (643 reports)

The conceptual information was taken from the following fields of the documents: description of the report, conclusion of the report and in some cases from the suggestion made by the doctor in charge. Text mining Eureka 3, 0 (beta) was used during this stage. This study used a shareware developed at the Federal University in Rio Grande do Sul (Brazil) because some text mining programs use natural language to remove some words during the procedure. The thesaurus was created using ISO 2788:1986 rules. The electronic interface was developed in html language.

Results

Through that relationship structure and its construction, surfing the terminology is a simple task, with no waste of time and no additional practice for the user. All information from different RIS tables, in different time frames and from random patients have the same structure and provides an organized, available and active surfing in the clinical data so that the professional can make his/her medical decision using the Bi-Rads™ thesaurus. As stated, Bi-Rads™ can be accessed in a web environment and in a near future it will be available on the Internet.