Presentation on a Method for Development of the Brazilian Health-related Content Web Search Portal

Felipe Mancini[a,b], Alex Esteves Jaccoud Falcão[a], Anderson Diniz Hummel[a], Fabio Teixeira[a], Fernando Sequeira Sousa[a], Thiago Martini Costa[a], Ivan Torres Pisa[c]

[a] *Postgraduate Program of Health Informatics, Federal University of São Paulo, São Paulo, Brazil*
[b] *Federal Institute of Education, Science and Technology of São Paulo, Guarulhos, Brazil*
[c] *Health Informatics Department, Federal University of São Paulo, São Paulo, Brazil*

Abstract

The increase in the amount of available information on the world wide web is inexorable, which, on one hand, provides the web user with more information. On the other hand, however, web searches become increasingly more difficult to handle due to the increasing number of retrieved documents. The present study is a proposal of development for a Brazilian search portal specific for health-related content. The aim of such development is to provide web users, mainly the non-specialist ones, with the largest number possible of web pages relevant to their search terms and inferred search intentions. The usefulness of a search portal for web pages with health-related content is potentially enormous, and the challenge of its implementation is motivating.

Keywords:

Internet, Health, Information storage and retrieval, Pattern recognition system.

Introduction

The world wide web content increases inexorably each day pushing and being pushed by the technological and economical development. Health-related activities are one contributor to this increase. In Brazil, 30% of web searches concern to information on health conditions related topics. However, adequacy and quality of available information are two relevant issues regarding health-related web content.

This paper shows a method that aims at implementing a Brazilian web portal for search and retrieval of health focusing on the needs and interests of the general public.

Materials and Methods

The architecture of this search portal, called In Health Search Model (InHealth), has a services-oriented architecture (SOA) - all modules to be incorporated to it should follow interoperation standards based on web-services. InHealth comprises fundamentally 3 modules – indexing module (IndHealth), inference module (InfHealth) and a graphical user interface module (IntHealth). The interaction of these 3 modules within InHealth architecture is outlined in Figure 1.
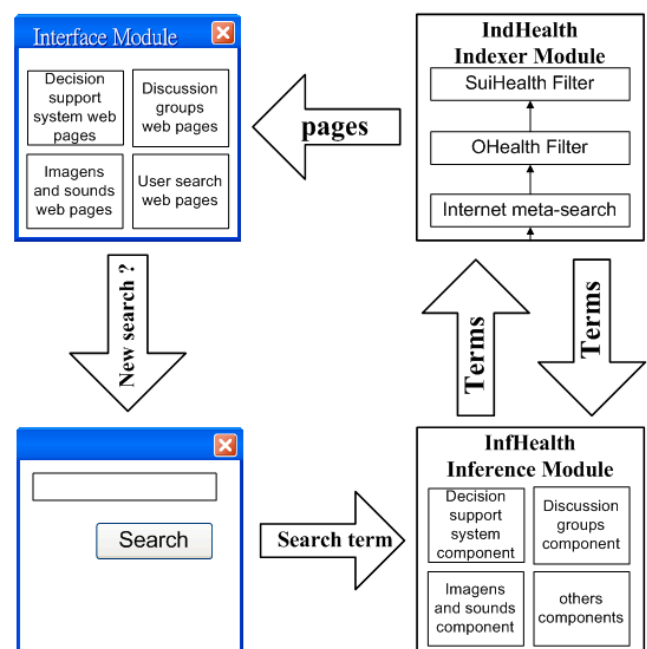

Figure 1- In Health Search Model modules interaction.

Preliminary results

Table 1 shows preliminary results of the IndHealth module for automated classification of health web content. Nearest-neighbor clustering presented best accuracy for this proposes.

Table 1-Sensitivity (SE) and specificity (SP) rates, and area under ROC curve (AUC) for nearest-neighbor clustering (NNC), artificial neural networks (ANN), and logistic regression (LR) classification algorithms, tested in OHealth filter.

	SE	SP	AUC
NNC	0.92	1.00	0.98
ANN	0.80	0.88	0.91
LR	0.92	1.00	0.98