UNIVERSITY OF GENOA
DEPARTMENT OF INFORMATICS, BIOENGINEERING, ROBOTICS AND SYSTEMS ENGINEERING
MASTER'S PROGRAM IN BIOENGINEERING

Dibris

Thesis Project Form

Title (tentative): Progettazione, sviluppo e test di un sistema preliminare di de-identificazione volto ad anonimizzare reports clinici.

Thesis advisor(s): Giacomini Mauro, Sara Mora

E-mail: Mauro.Giacomini@dibris.unige.it

Address: Via Opera Pia 13

Phone: (+39) 010 33 56546

Description

Motivation and application domain

Nello scenario dei sistemi di cartelle cliniche e dei fascicoli sanitari elettronici, gli sforzi svolti negli ultimi decenni hanno avuto come obiettivo la raccolta di dati strutturati, come ad esempio i risultati degli esami di laboratorio, per la conduzione di studi clinici. Successivamente si Ã¨ notato che una grande parte potenzialmente utile delle informazioni Ã¨ perÃ² disponibile solo sotto forma di testo libero. Da ciÃ² Ã¨ nata lâ€™esigenza di sviluppare sistemi in grado di analizzare il â€œlinguaggio naturaleâ€• cosÃ¬ da ottenere in forma automatica e strutturata quelle informazioni che generalmente si trovano sotto forma non strutturata. Un prerequisito fondamentale per lâ€™uso di questi testi in ambito di ricerca clinica Ã¨ la loro anonimizzazione.

General objectives and main activities

La tesi si inserisce allâ€™interno di un progetto piÃ¹ ampio con lâ€™obiettivo primario di sviluppare un sistema che estragga automaticamente informazioni cliniche rilevanti da referti medici redatti in italiano, con testo libero, attraverso lâ€™applicazione di algoritmi di analisi del linguaggio naturale. Il workflow delle attivitÃ necessarie alla progettazione e sviluppo di un sistema preliminare di de-identificazione consisterÃ in:
â€¢ Suddivisione della frase in token
â€¢ Identificazione delle possibili Private Health Information (PHI), che possono appartenere a diverse categorie: nomi e cognomi, date e luoghi di nascita, ID, nome dellâ€™ospedale o del medico di riferimento, numeri di telefono, ...
â€¢ Costruzione di un database di appoggio da cui estrarre elementi casuali con cui sostituire i dati personali identificati
â€¢ Sostituzione dei dati personali con dati casuali
â€¢ Test del sistema sviluppato

Training Objectives (technical/analytical tools, experimental methodologies)

Nel corso della tesi proposta, lo studente apprenderÃ e utilizzerÃ le seguenti competenze:
â€¢ Strumenti di modifica e interrogazione di database (SQL)
â€¢ Linguaggiodi programmazione Python
â€¢ Introduzione alle tecniche di Natural Language Processing
â€¢ Approfondimento di approcci Ruled-Based
â€¢ Accenni di Machine Learning

Place(s) where the thesis work will be carried out: Laboratorio di Bioingegneria - DIBRIS

Additional information

Maximum number of students: 1