

**Анотация на дипломна работа  
към катедра “Информационни технологии”,  
ФМИ, СУ “Св. Кл. Охридски”**

**Дипломант: Марин Маринов Ножчев**

**Магистърска програма: Био- и медицинска информатика**

**Научен ръководител: ас. д-р Калин Георгиев, ФМИ**

**Тема на дипломната работа: *Анонимизация на клинична информация за пациенти***

Наличието на биомедицински данни за пациенти е критично за изследванията в медицината. Основен източник на данни са болниците, но здравните заведения нямат право по закон да ги разпространяват. Данните могат да се използват за изследвания само ако се трансформират по начин, който не позволява информацията да се свърже с конкретни пациенти. Този процес се нарича „медицинска анонимизация“.

Анонимизацията често се извършва тривиално чрез заменяне на имената на пациентите с уникални идентификатори. По този начин се запазва цялата релевантна за клинични изследвания информация като възраст и местопребиваване на пациентите. За съжаление, с разпространението на и полу-публични бази данни с лична информация на населението като Facebook, вече е възможно свързването на частично идентифицираща информация като година на раждане, местоживеене, пол с конфиденциални записи в клинични бази данни.

Основна пречка за пълната анонимизация са немаркираните лични данни особено в неструктурирани медицински текстове като анамнези и диагнози в амбулаторни листи. В тези документи често се споменават лични данни за пациентите дори извън съответните полета на формуляра. Например годината на раждане на пациента често се споменава в диагнозата, ако възрастта е релевантна към нея.

Целта на дипломната работа е анализ на state-of-the-art алгоритмите за анонимизация на клинични данни и реализация на алгоритъм, базиран на правила, който маркира лични данни в свободен медицински текст. Софтуерът е по-гъвкав от много от съществуващите решения.

Задачи, произтичащи от целта:

1. Обзор на стандарти алгоритми за анонимизация и тяхната софтуерна реализация.
2. Създаване на софтуер за автоматично откриване на потенциално идентифициращи данни в клинично документи базиран на статистически анализ.

Дата: 14.10.2013 г.

Дипломант:

Научен ръководител: