

Master thesis annotation
Department of Information Technology,
FMI, Sofia University "St. Kl. Ohridski"

Student: Marin Marinov Nozhchev

Master course: Bio- and medical informatics

Scientific advisor: asst. Kalin Georgiev, PhD, FMI

Master Thesis Topic: *Anonimization of clinical patient data*

The availability of clinical data is critical for data-driven medical research. The primary source of such data are hospital, but all public health institutions are forbidden from distributing private data. The data can be used for research only after they are transformed in a way, which doesn't allow the result to be related to individual patients. That process is defined as „medical Anonimization“.

Anonimization is frequently implemented as trivial replacement of patient names and addresses with unique identifiers. This way all relevant clinical information is retained in the text, including age and nationality of patients. Unfortunately, large semi-public databases with personal data, like Facebook, are increasingly available, which makes it possible to link trivially anonymized records to the actual people by combining individually harmless data points like year of birth, city of residence and gender.

The major obstacle to complete anonimization is the presence of unstructured medical texts like treatment history and diagnosis. In those documents, there are frequent mentions of patient private data outside dedicated form fields. For example, the age of the patient is frequently mentioned if it is relevant to the diagnosis.

The goal of the master thesis is an overview and analysis of the state-of-the-art algorithms for anonimization of clinical data and the implementation of a rule-based algorithm that marks and removes private data in unstructured clinical text. The software is significantly more flexible and easy to integrate in larger systems than existing solutions.

14.10.2013

Student: