Planning safety solutions, models and algorithms for special databases

Fanni Magdolna Hízó[a], **Attila Kiss**[b]

[a]ELTE Eötvös Loránd University, Budapest
fanni.hizo@gmail.com

[b]ELTE Eötvös Loránd University, Budapest
kiss@inf.elte.hu

Abstract

DNA sequencing is the process of determining the order of nucleotides in DNA. The rapid speed of sequencing attained with modern DNA sequencing technology has been instrumental in the sequencing of complete DNA sequences, including the human genome. Nevertheless it is a sensitive data which needs safe but efficient storage methods.

The goal in this research was to analyze different models and algorithms to determine which is the most applicable for storage, and query considering the need of user permissions, and encryption.

In order to do this, we have examined different database management softwares how its works, is there any limitation which is worth considering. We have tested some algorithms in MySql and PostgreSQL, and compared the outcome to Python program results.

The next stage was to compare different species DNA sequence to the human genome. We analyzed the distribution of nucleotides with k-mers.

Finally we have looked into compression methods and encryption, compared the results to previous researches, and determined the future work.

The project has been supported by the European Union, co-financed by the European Social Fund (EFOP-3.6.3-VEKOP-16-2017-00002).

Keywords: bioinformatics, data science, scientific databases, data protection, relational databases, data compression, database applications