



TALLINNA TEHNIKAÜLIKOOL
INSENERITEADUSKOND

TTÜ Tartu kolledž

MAA-AMETI KINNISVARA TEHINGUTE STATISTIKA ANDMEKAEVE TÖÖRIISTA VÄLJATÖÖTAMINE

CREATING THE DATA-MINING TOOL FOR STATISTICS ON REAL PROPERTY
TRANSACTIONS OF ESTONIAN LAND BOARD

RAKENDUSKÕRGHARIDUSETÖÖ

Üliõpilane: Jarmo Kurs

Üliõpilaskood: 154582NDFR

Juhendaja: MSc Rinaldo Rüütli, insener

Tartu, 2019

KOKKUVÕTE

Tänapäevases andmeterikkas maailmas on üha suurenev konkurents uute andmete väljaselgitamiseks, sest andmed on saanud väga väärtuslikeks. Andmeteaduses luuakse aina enam spetsiaalseid tööriistu, et kätte saada olulist ja kasulikku informatsiooni, mida enda tarbeks võimalikult hästi ära kasutada. Käesoleva töö eesmärk oligi luua andmekaeve tööriist ning kontrollida selle toimivust Maa-ameti hinnastatistika alusel. Eesmärgi täitmiseks tegi autor esiteks ülevaate andmekaeve olemusest, eesmärkidest, arengust tänaseni ja erinevatest andmeallikate vormidest. Järelduste tegemisele andsid olulise aluse andmekaevest tulenevate ohuallikate uurimine, mitte ainult Maa-ameti kui asutuse kontekstis, vaid ka kogu ühiskonna aspektis.

Eesmärgi täitmiseks valmistas autor andmekaeve tööriista, mis sisaldas endas mitmeid skripte. Andmete salvestamiseks loodi skript `dbfconverter.py`, mis avas `dbf` laiendiga andmefaili ja kirjutas vajaliku info MongoDB andmebaasisüsteemi dokumentide kogusse „katastrid“. Järgmine skript nimega `hinnaparing.py` ülesandeks oli pöörduda Maa-ameti päringukeskkonna poole läbi URL-i aadressi, võttes aluseks EHAK koodid ja valitud ajavahemiku ning võrrelda üksteisele järgnevatel päevadel toimunud tehingute arvu. Kui tehingute arvus oli vahe, kirjutati saadud info koos katastriüksuste arvu, pindala, hinna, EHAK koodi ja kuupäevaga MongoDB tarkvaras asuvasse „Maa-ameti“ andmebaasi dokumentide kogusse „hinnad“. Kolmanda skripti nimega `tehinguinfo.py` ülesandeks oli kogust „hinnad“ välja võtta read, kus katastrite arv on üks. Sellisel juhul otsis skript kogust „katastrid“ välja vastava katastritunnuse ja salvestas selle dokumentide kogusse „tehinguinfo“. Juhul, kui dokumentide kogus „hinnad“ oli katastrite arv enam kui üks, siis oli võimalik tuvastada autori poolt välja toodud „täpsustava algoritmiga“, millised katastriüksused antud haldusüksuses müüdi, kuid igaühe täpset hinda polnud võimalik tuvastada.

Analüüsi tulemust iseloomustabki kõige paremini dokumentide kogu „tehinguinfo“, mis sisaldab hulga ridu, milles igaühes on üks katastritunnus, selle pindala, tehingu hind ja kuupäev. Need ongi andmekaeve protsessi poolt väljastatud uued, huvitavad ja kasulikud andmed. Seega võib andmekaevet ja tööks püstitatud eesmärgi täitmist üsna edukaks pidada. Tööriista peamiseks tugevuseks võib pidada selle kergesti jälgitavat ent informatiivset ja põhjalikku sisu. Skriptide töökindlust on korduvalt testitud erinevate lähteandmetega. Nõrgaks kohaks on juhtumid, kus tööriist ei suuda välja tuua konkreetset katastritunnust, kuna sama EHAK koodiga piirkonnas on mitu sarnase pindalaga maatükki. Siin näeb ka autor tööriista täiendamise võimalust tulevikus.

Analüüsi käigus tuvastatud andmeid saaksid informatiivselt enda tarbeks kasutada kinnistute kokkuostjad, saades oma konkurentide ees tugeva eelise. Kahjuks aga lähevad sellised teadmised

vastuollu ühiskonnas soovitule, kui avalikuks saavad paljude tehingute väärtused. Kuigi Maa-ameti hinnangul peaksid tehingu hindade info kätte saama vaid loa saanud isikud, siis käesolevas töös loodud tööriista abil jõuaksid selleni paljud inimesed. Autor tegi mõningad soovitused Maa-ametile eesmärgiga nendepoolset jagatavat infot rohkem kontrolli all hoida, et tehingute andmed poleks sellisel viisil kättesaadavad ja andmete kaevandamist tuleks edaspidiselt teatavate meetmetega piirata.