

TALLINNA TEHNIKAÜLIKOOL

Infotehnoloogia teaduskond

Arvutiteaduste instituut

**Soovitussüsteemide rakendamine
tööpakkumisele sobiva töötaja leidmiseks
Töötukassa iseteeninduskeskkonnas**

bakalaureusetöö

Üliõpilane: Rauno Mänd
Üliõpilaskood: IAPB103702
Juhendaja: Evelin Halling

Tallinn
2015

Autorideklaratsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

(kuupäev)

(allkiri)

Annotatsioon

Käesoleva töö eesmärgiks on töötada Eesti Töötukassa Iseteenindusportaali jaoks välja uus soovitusüsteem, mis soovitab tööandja poolt sisestatud ning Töötukassa konsultandi poolt valiidselt tunnustatud tööpakkumisele vastavusse antud tööpakkumisega sobivaid isikuid. Sobiva kandidaadi soovitamisel võetakse arvesse Töötukassa konsultandi poolt tööpakkumisele määratud ISCO_TK standardi väärtust ning tööotsijate poolt Töötukassa Iseteenindusportaalis sisestatud CV-de ja töösoovide pealt saadavaid andmeid.

Töös vaadeldakse esmalt soovitusüsteemide mõistet ning esitatakse Eesti Töötukassa Iseteenindusportaali poolt praegu kasutusel oleva soovitusüsteem põhimõtted. Seejärel tuuakse välja viimase head ja vead ning analüüsitakse, mida ning kuidas annaks antud soovitusüsteemi näol täiustada. Ühtlasi esitletakse andmeid ja andmetalletusviise, mida oleks täiustatud soovitusüsteemis põhjendatud ning otstarbekas kasutada ning millised on andmemahutudega kaasnevad keerukused ja võimalused potentsiaalselt suurte andme- ning arvutusmahutude optimeerimiseks. Töö lõpptulemuseks on analüüs sellest, milline võiks autori arvates välja näha uus ning täiuslikum, mainitud eesmärki täitev, soovitusüsteem.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti viiekümne kahel leheküljel, koosnedes viiest peatükist, neljateistkümnest joonisest ning viiest tabelist.

Abstract

The major goal of this thesis is to work out a new recommendation system for the Estonian Unemployment Insurance Fund's self-service portal, which would recommend the best matching of unemployed people for the job offers that are entered by employer itself and validated by Estonian Unemployment Insurance Fund consultant.

Finding the eligible candidate, ISCO_TK standard value designated by Estonian Unemployment Insurance Fund consultant and CV-s, as well as data for job requests, entered by unemployed people through Estonian Unemployment Insurance Fund's self-service portal will be taken into consideration.

In thesis ab initio, the concept of the recommendation system will be reviewed after which the Estonian Unemployment Insurance Fund's self-service portal's recommendation system, currently in use, is being examined and its positive and negative traits identified, analyzed in an effort to improve and make recommendations for the current system. Additionally, data and data storage methods that would prove sufficient and functional to use in the improved recommendation system will be represented along with the identification of potential complications, necessary data size and different solutions for optimizing it.

Thesis conclusion will provide analysis of the authors ideas and concepts on the new and improved thesis recommendation system, which, will fullfill the major goal of this thesis.

The thesis is in Estonian and contains 52 pages of text, 5 chapters, 14 figures, 5 tables.

Lühendite ja mõistete sõnastik

EMPIS	<i>Estonian Employment Information System</i> Eesti Tööhõive Infosüsteem (Nortal – EMPIS, 14.03.2015).
ISCO	<i>International Standard Classification of Occupations</i> Rahvusvaheline ametinimetuste klassifitseerimisstandard (ILO – ISCO sõnastik, 12.02.2015).
ISCO_TK	ISCO standardi baasil Töötukassa infosüsteemile kohaldatud nn standard.
ITP	Töötukassa iseteenindusportaal
Kaasfiltreerimine	<i>Collaborative Filtering</i> Statistilisel korrelatsioonil põhinev sisu filtreerimise meetod (Prem Melville and Vikas Sindhwani - Recommender Systems, 23.03.2015).
Sisu-põhine filtreerimine	<i>Content-based Filtering</i> Otsestel parameetrite võrdlemisel põhinev sisu filtreerimise meetod (Prem Melville and Vikas Sindhwani - Recommender Systems, 23.03.2015).

Jooniste nimekiri

Joonis 1. ITP eraisiku avaleht.....	15
Joonis 2. Tööandja poolt sisestatava tööpakkumise samm 3 (näide vormi väljadest, kompaktsel kujul).	17
Joonis 3. Tööandja poolt Töötukassasse kontrollimisele saadetava tööpakkumise staatused tööandjale nähtava ITP kasutajaliidese läbi.	18
Joonis 4. CV täitmise samm 3 – Töökogemus.	19
Joonis 5. CV täitmise samm 5 - Täiendkoolitused.	21
Joonis 6. CV täitmise samm 6 - Oskused.	22
Joonis 7. Töösoovi lisamine – esimene samm.....	23
Joonis 8. Töösoovi lisamine – teine samm.	24
Joonis 9. Hierarhiline puu ning nende sõlmede vahel liikumine.....	26
Joonis 10. Tööpakkujale ITP kasutajaliidese vahendusel nähtav ja hallatav tööpakkumise raames süsteemi poolt leitavate sobivate kandidaatide nimekiri.	28
Joonis 11. ISCO_TK hierarhiline struktuur.....	36
Joonis 12. Struktuuriline esitus reaalsete ametinimetuste näitel.	36
Joonis 13. Ametinimetuste andmete struktuuri kirjeldav konseptuaalmudel.	37
Joonis 14. Ametinimetuste klassifikaatorite suunatud graaf	38

Tabelite nimekiri

Tabel 1 - Tööpakkumiste, töösoovide ning CV-de parameetrite ühisosa ülevaade	32
Tabel 2 - Ülevaade parameetrite mahtudest ITP süsteemis seisuga 23.03.2015	33
Tabel 3 - Hierarhilise struktuuri tabeliesitus Joonis 14 näitel.	38
Tabel 4 - Töö asukoha sobivusprotsentide jaotus.....	40
Tabel 5 - Parameetrite kaalud ja prioriteedid	46

Sisukord

Sissejuhatus	10
1. Teoreetiline ülevaade soovitussüsteemidest	12
1.1 Disain	12
1.2 Meetodid	12
2. Töötukassa iseteenindusportaali ülevaade	14
2.1 EMPIS	14
2.2 Töötukassa iseteenindusportaal	14
2.2.1 Iseteenindusportaali töövahenduse soovitussüsteemiga seotud moodulid	16
2.2.1.1 Tööpakkumine	16
2.2.1.2 CV	18
2.2.1.3 Töösoov	22
3. Tööturuameti Selveinfosüsteem	25
3.1 Hierarhiliste klassifikaatorite sobivusprotsendi arvutamine	26
3.2 Parameetrite komplektide protsentuaalne sobivus	26
3.3 Parameetrite liitmine ja terviksobivuse leidmine	27
4. Kasutusel oleva soovitussüsteemi lahenduse analüüs	28
5. Uue soovitussüsteemi lahenduse välja töötamine	32
5.1 Tööpakkumiste ja töösoovide ning CV-de parameetrite ühisosa	32
5.2 Andmemahud	33
5.3 Algoritmid	34
5.3.1 Ametinimetuste protsentuaalne sobivus	35
5.3.1.1 ISCO ja ISCO_TK	35
5.3.1.2 Klassifikaatori protsentuaalse sobivuse arvutamine	37
5.3.2 Töö asukoha protsentuaalne sobivus	39
5.3.3 Töökogemus aastates, protsentuaalne sobivus	41
5.3.4 Kutsetunnistuse protsentuaalne sobivus	42
5.3.5 Ametitunnistuse protsentuaalne sobivus	42
5.4 Terviksobivuse arvutamine	43
5.4.1 Sobivusprotsentide ette arvutamine	44
5.4.2 Parameetrite prioritiseerimine	45

Kokkuvõte	48
Summary.....	50
Kasutatud kirjandus	52

Sissejuhatus

Tööturuteenuste ja –toetuste seaduses sätestatud tööturuteenuste osutamist korraldab ja tööturutoetusi määrab Eesti Töötukassa. Eesti Töötukassa üheks seadusest tulenevaks kohustuseks on tööturuteenuste osutamine, mis peab soodustama isiku töölerakendumist ning tööandjal sobiva tööjõu saamist (Tööturuteenuste ja –toetuste seadus. – RT I, 04.06.2014, 8). Seaduse täitmise ühe väljundina on Eesti Töötukassa Iseteenindusportaali jaoks arendatud soovitusüsteem, mis suudab tööpakkumiste, kui töötute kohta käivate andmete põhjal tööandja tööpakkumise ja tööotsijad omavahel kokku viia. Küll on aga antud lahendus oma praktikas oldud aja vältel näidanud välja teatavaid puudujääke. Nende puudujääkide likvideerimine või ka leevendamine on kolme osapoole, nii Eesti Töötukassa, kui soovitusüsteemi tellija, tööandja, kui ka tööotsija seisukohast finantsilises, ajalises, kui ka emotsionaalses mõistes kriitilise tähtsusega.

Käesoleva töö algseks eesmärgiks on analüüsida praegu kasutusel oleva Eesti Töötukassa Iseteenindusportaali tarvis välja töötatud soovitusüsteemi lahendust ning selle puudujääke. Teiseks töö eesmärgiks on analüüsida 2003. aastal koostatud Tööturuameti Selveinfosüsteemis kasutusel olnud töökuulutuse, töösoovide ja CV-de sobitamise arvutusmootorit, millest lähtuvalt annab töö autor ülevaate, milliseid meetodeid ning kuidas oleks võimalik selveinfosüsteemi jaoks välja töötatud lahendustest ära kasutada Töötukassa Iseteenindusportaali tarbeks. Kolmandaks ja ühtlasi töö põhieesmärgiks on pakkuda välja autoripoolne lahendus uue ning täiuslikuma soovitusüsteemi väljaarendamiseks, mis peaks tagama eelpool toodud kolmele osapoolele võidu nii ajalises, finantsilises, kui ka emotsionaalses mõttes.

Töö jaotub viide suuremasse osasse. Neist esimene annab teoreetilise ülevaate soovitusüsteemide mõistest kui ka olemusest. Töö teine osa kirjeldab põgusalt Töötukassa Infosüsteemi ning hiljem detailsemalt sellega liidestuvat Töötukassa Iseteenindusportaali. Kusjuures Töötukassa Iseteenindusportaali kirjeldamise peatükis tuuakse juba esile autori poolt välja töötava soovitusüsteemi raames olulised moodulid ning nendest uue soovitusüsteemi väljatöötamise rakendusliku osa sisendiks olevad andmeelemendid. Töö kolmandas osas kirjeldatakse ja analüüsitakse praegu kasutusel olevat lahendust koos oma eeliste ja puudustega.

Töö neljas osa keskendub juba eespool, eesmärkide osas mainitud, 2003. aastal koostatud Tööturuameti Selveinfosüsteemis kasutusel olnud töökuulutuse, töösoovide ja CV-de sobitamise arvutusmootori analüüsile, millest lähtuvalt annab töö autor ülevaate, milliseid meetodeid ning kuidas oleks võimalik selveinfosüsteemi jaoks välja töötatud lahenduste hulgast ära kasutada Töötukassa Iseteenindusportaali infosüsteemi tarbeks. Juba mainitud töö osad ning nende tulemid on ühtlasi sisendiks töö viimasele, viiendale suuremale osale, kus autor keskendub detailsemalt enda poolt pakutava, täiuslikuma soovitusüsteemi väljaarendamise kirjeldamisele ning analüüsile.

1. Teoreetiline ülevaade soovitusüsteemidest

Soovitusüsteemi põhieesmärk on genereerida süsteemi lõppkasutajale tähendusrikkaid soovitusi mõne toote, teenuse või miks-mitte ka inimese kohta. Reaalse elu näited sellest on raamatusoovitus elektroonses ostukeskkonnas Amazon või sõbrasoovitus elektroonses sotsiaalvõrgustikus Facebook. (Prem Melville and Vikas Sindhwani - Recommender Systems, 23.03.2015)

1.1 Disain

Soovitusüsteemi lahenduse disain sõltub mitmetest erinevatest asjaoludest. Näiteks sellest, millisesse domeeni soovitused kuuluvad. Domeeni all mõeldakse seda, mida soovitakse, ehk eelpool mainitud tooteks saab olla mõni kindel ese, näiteks raamat. Toode saab olla ka informatsioon, näiteks uudised või hotellihalvustused. Inimeste soovitamise all saame aga lisaks eelnevalt mainitud Facebooki sõbrasoovitustele, mõelda ka näiteks sobivate kandidaatide soovitamist tööpakkumisele. (Prem Melville and Vikas Sindhwani - Recommender Systems, 23.03.2015)

Lisaks domeenile on olulisteks teguriteks veel näiteks eesmärk, nendest kõige levinum on müük. Samas eesmärgiks võib olla ka lojaalsus või näiteks kogukonna ehitamine, et populariseerida mõnd hobi või lahendada mõni suurem probleem nagu näiteks vähiravimi leiutamine. (Joseph Konstan and Michael Ekstrand – Intro. Recommender Systems, 03.02.2015)

Viimasena võiks soovitusüsteemi disaini juures välja tuua saadaval olevate andmete iseloomu ja struktuuri. Ehk, kui palju andmeid on, mida need andmed kirjeldavad, kuidas neid hoitakse ning mis on nende privaatsustase jne. (Prem Melville and Vikas Sindhwani - Recommender Systems, 23.03.2015)

1.2 Meetodid

Soovitusüsteemid baseeruvad soovituste genereerimisel enamasti ühele küsimusele. Nimelt, „Kas kasutajale U meeldib element X“. Kaks levinuimat meetodit sellele vastuse leidmiseks on:

- Vaadata milline on U ning seejärel vaadata, kas X on sarnane.
- Vaadata kellele meeldib X ning seejärel vaadata, kas U on talle sarnane.

Esimene meetod põhineb otsesel parameetrite võrdlemisel ning soovitusüsteemide vallas kutsutakse seda sisupõhiseks filtreerimiseks. Teine meetod põhineb statistilisel korrelatsioonil ning soovitusüsteemide vallas kutsutakse seda kaasfiltreerimiseks. Ühtlasi kasutatakse ka eelmainitute kombinatsioone. (Prem Melville and Vikas Sindhwani - Recommender Systems, 23.03.2015)

Autor keskendub oma töös esimesele, otsesel parameetrite võrdlemisel põhinevatele meetoditele. Töö edenedes selguvad meetodi detailid ning see, kuidas viimast on võimalik praktikas kasutada.

2. Töötukassa iseteenindusportaali ülevaade

Antud peatükis annab autor mainitust detailse ülevaate just nimelt soovitusüsteemide mõistes. Esmalt käsitletakse põgusalt Töötukassa keskset infosüsteemi, ehk EMPIS-t. Seejärel võetakse vaatluse alla käesoleva töö üks olulisematest objektidest - Töötukassa iseteenindusportaal ehk ITP. Viimane on EMPIS-ega integreeruv osa. Peale Iseteenindusportaali üldisele vaatlemisele tuuakse esile uue soovitusüsteemi välja töötamise juures olulist rolli mängivad kolm peamist elementi, Iseteenindusportaali moodulit, milleks on tööpakkumised, CV-d ja töösoovid.

2.1 EMPIS

EMPIS (Estonian Employment Information System, Eesti Tööhõive Infosüsteem) on veebipõhine menetlusinfosüsteem, mis võimaldab kiiret ja mugavat töötute registreerimist, neile töötushüvitiste määramist, erinevate teenuste pakkumist, tööpakkumiste vahendamist, lepingute allkirjastamist asutustega, hangete menetlust, jne. Süsteem teostab ka andmevahetust teiste riiklike infosüsteemidega ja Töötukassa siseste rakendustega (dokumendihaldusrakendus, raamatupidamistarkvara, Töötuskindlustuse Infosüsteem ehk TKIS, Iseteenindusportaal ehk ITP). EMPIS lihtsustab menetlust nii töötute kui ka Töötukassa ametnike jaoks, tehes ametnike ülesandeid lihtsamaks ja tõhusamaks. (Nortal – EMPIS, 14.03.2015)

2.2 Töötukassa iseteenindusportaal

Üheks EMPIS-e suuremaks ja olulisemaks tugisüsteemiks võib nimetada sellega liidestuvat süsteemi - Töötukassa Iseteenindusportaali (edaspidi ITP). Seisuga 21.05.2015 pääseb Töötukassa iseteeninduskeskkonda interneti aadressilt www.tootukassa.ee/iseteenindus. Keskkonda pääsemiseks on vaja esmalt ennast identifitseerida. Identifitseerimiseks on võimalik valida kahe meetodi vahel: ID-kaart või Mobiil-ID. Pärast õnnestunud identifitseerimist avaneb standardtingimustel eraisiku vaade (vt. Joonis 1).

Minu avaleht

- ▶ Kokkulepitud tegevused
- ▶ Töötamispäevik
- ▶ Minu tegevuste ajalugu

Minu andmed

Minu CV-d

Minu volitused

Minu töopakumised

Avaldused ja otsused

Töötõtsingud

Vajalikud teenused

i Olete sisseloginud eraisikuna. Ettevõttega seotud toiminguteks vahetage rolli. Seost ettevõttega saate vaadata "Minu volitused" alt

Minu tegevused - Maie Omapere Kaks

Teated

Teil ei ole ühtegi teadet.

Individaalses töötõtsimiskavas kokkulepitud tegevused (ITK)

Teil ei ole ühtegi töötõtsukassaga kokkulepitud tegevust.

Joonis 1. ITP eraisiku avaleht

ITP on süsteem, kus klientidel on iseseisvalt võimalik otse EMPIS-ega suhelda. Tavapärastelt on ITP klientideks, kas töötõ, tööd otsiv inimene või tööd pakuv ettevõte. Töötule või tööd otsivale inimesele pakub ITP põhifunktsionaalsuste seast, sõltuvalt vajadusest, kas näiteks digitaalsel kujul Töötõtsukassale erinevate avalduste või taotluste esitamise, koolitustele registreerimise, tööpakkumistele kandideerimise ja töösoovide ning CV-de esitamise ja sisestamise võimalust. Kõige harilikumad avaldused, mida läbi ITP esitatakse on töötõna või töötõtsijana arvelevõtmise ning töötõtoetuse avaldused. Ettevõõtjale on ITP poolt pakutavaks põhifunktsionaalsuseks tööpakkumiste loomine ning neile kandideerinute ning potentsiaalsete kandidaatide haldamise võimalus. Samuti on ettevõõttele võimalik läbi ITP esitada avaldusi ja teavitusi. Lisaks on koolitõtsukaarti omavatel ettevõõtetel võimalus läbi ITP koolitõtsi ning nende registreerunuid hallata.

Et süsteemist on soovitusõtsüsteemi väljatõõtamise raames fookuse all just ettevõõttele tööpakkumist ning sellele sobivate kandidaatide leidmist puudutav funktsionaalsus, vaadeldakse järgnevalt lähemalt just nendega seonduvat.

2.2.1 Iseteenindusportaali töövahenduse soovitusüsteemiga seotud moodulid

Töötukassa iseteenindusportaalil on kasutusel mitmeid olulisi mooduleid, millest anti ülevaade ka eelmises peatükis, kuid antud alapunktis vaatleb autor vaid analüüsitava loodava soovitusüsteemi kontekstis olulisi. Sestap tuuakse olulistena välja tööpakkumiste, CV-de ning töösoovide moodulid. Oluline on ära märkida ka see, et eespool nimetatud moodulid on iseteenindusportaalil omakorda seotud kahe rolliga, milleks on tööpakkuja ning kandidaat ehk töötu või tööd otsiv isik. Tööpakkujaga seondub tööpakkumiste ning kandidaadiga seonduvad kaks ülejäänud, CV-de ning töösoovide, moodulit. Järgnevates alapunktides on nimetatud moodulid täpsemalt lahti kirjeldatud.

2.2.1.1 Tööpakkumine

Tööpakkumise mooduli näol on tegemist tööpakkuja rolli kuuluva funktsionaalsusega. Tööpakkumise sisestamisel on tööandjal vaja korrektselt läbida kolm sammu, mille käigus sisestatud andmete näol tekib terviklik tööpakkumine. Need sammud on oma üldnimetuse ja küsimise järjekorra poolest järgmised (vt. Ühtlasi Joonis 2): „Tööandja andmed“, „Tööpakkumise kirjeldus“, „Nõuded kandidaadile“.

Tööandja andmed > Tööpakkumine > **Nõuded kandidaadile** > Tööpakkumise eelvaade ja esitamine

Nõuded kandidaadile

Haridustase

Nõutav varasem töökogemus samal ametialal

Sertifikaat, pädevustunnistus, mootorsõidukijuhi ametikoolituse läbimise kutsetunnistus

ADR (ohtlike ainete vedu) luba Bussijuht

Elektriku pädevustunnistus A Elektriku pädevustunnistus B

Kutsetunnistus

Kutsetunnistus + Lisa

Keeleoskus + Lisa uus

Keel	Kõnes	Kirjas	Muuda	Kustuta
Ühtegi Keeleoskust ei ole lisatud!				

Juhiload

A1 B1 C1 D1 BE C1E D1E B ADR

Arvutioskused

e-posti kasutamine interneti kasutamine kontoritarkvara

maiandustarkvara inseneritarkvara arvutigraafika

Joonis 2. Tööandja poolt sisestatava tööpakkumise samm 3 (näide vormi väljadest, kompaktsel kujul).

Kui tööandja on Tööpakkumise sammud edukalt ära täitnud, järgneb sellele koostatud tööpakkumise Töötukassale kontrollimiseks saatmise samm (vt. Joonis 3). See samm väldib seda, et tööpakkumine, mis Töötukassa portaalis avaldatakse poleks oma olemuselt ebakvaliteetne. Kontrollimisele läinud tööpakkumise valideerib ning korrigeerib Töötukassa poolt konsultant. Konsultant saab kas tööpakkumise tagasi lükata, kui tegemist näiteks ebakvaliteetse eksemplariga või siis lisada tööpakkumisele külge teda iseloomustav ISCO ning seejärel tööpakkumine avaldamiseks sobivaks kinnitada. Kui tööpakkumine on avaldamiseks sobivaks kinnitatud, saab ta kõigile nähtavaks ning töötule ja tööotsijatele kandideeritavaks.

- 1) Edukalt täidetud tööpakkumine saadetakse tööandja poolt Töötukassasse valideerimisse.
- 2) Töötukassa konsultant valideerib tööpakkumist, mille tulemusena ta kas
 - 2.1) lükkab tööpakkumise tagasi VÕI
 - 2.2) lisab tööpakkumisele ISCO ja kinnitab tööpakkumise. Viimase tulemina muutub tööpakkumine niinimetatult "Aktiivseks", ehk avalikuks ning töötule või tööotsijatele kandideeritavaks

Joonis 3. Tööandja poolt Töötukassasse kontrollimisele saadetava tööpakkumise staatused tööandjale nähtava ITP kasutajaliidese läbi.

Tööpakkumise mooduli raames tööandjalt nõutavatest ja sisestatavatest parameetritest on autori poolt välja töötatava soovitusüsteemide kirjelduse ja realiseerimise juures olulisteks parameetriteks järgmised: ametinimetus, töö asukoht, haridustase, hariduse valdkond, kutsetunnistus, ametitunnistus, keeleoskus, tööaeg, töötasu, juhiloa, arvutisokkus, töösuhte kestus, töötasu tüüp, tööleasumise aeg, töökogemus aastates.

2.2.1.2 CV

Antud peatükk kirjeldab töötaja CV loomise protsessi. CV loomise protsess koosneb seitsmest üldistatud nime kandvast sammust: „Üldandmed“, „Isikuandmed“, „Töökogemus“, „Hariduskäik“, „Täiendkoolitused“, „Oskused“ ja „Muu“. Töö käigus välja töötatava soovitusüsteemi jaoks mitte-olulised punktid jäetakse täpsemalt kirjeldamata, sõltumata sellest, kas ITP hetkel toimivates soovitusüsteemides neid kasutab või mitte.

Kui kandidaat on Töötukassa iseteeninduskeskkonda sisse loginud on tal võimalus luua ja hallata endale kuuluvaid CV-sid. Uue CV loomisel avaneb esimese sammuna CV üldandmete vaade, kus soovitusüsteemi jaoks oluliseks punktiks on CV staatus. Sobivate kandidaatide

nimekirja koostamisel võetakse arvesse ainult CV-d, mille staatus on kas „Avalik“ või „Anonüümne“. Siinkohal avaneb kandidaadil võimalus reguleerida enda osalemist soovitusüsteemis.

Täitnud üldandmete sammus kõik korrektselt, jõutakse töötaja CV isikuandmete sammu, kus välja töötatava lahenduse mõistes olulisi punkte välja tuua ei saa, kuna Töötukassa ärireeglitest lähtudes ei tohi näiteks soo jms alusel tööle kandideerijat valida.

Kolmandaks sammuks on töökogemuse sisestamise samm. Töökogemuste sammus saab kandidaat sisestada kogu oma senise töökogemuse. Uue töökogemuse lisamiseks tuleb vajutada nupule „Lisa uus“, misjärel avaneb standardvorm mida kirjeldab ka Joonis 4.

The screenshot shows a web form titled "Töökogemus" (Work Experience) within the Töötukassa system. The form is divided into several sections:

- Tööandja ***: A text input field with a placeholder "Ettevõtte nimetus, kus te töötasite."
- Ametinimetus ***: A text input field.
- Ametikoha valdkond**: A dropdown menu with the option "- vali -". Below it, a note says "Näiteks raamatupidaja ametikoha valdkond on "finants, raamatupidamine, statistika"."
- Riik ***: A dropdown menu with the option "Eesti".
- Ajavahemik ***: A date range selector with dropdowns for month and year, and input fields for "aasta" and "kuni" (year and month).
- Tööülesanded ***: A large text area with a placeholder "kuni 2000 tähemärki".
- Saavutused**: A text area with a placeholder "kuni 100 tähemärki".

At the bottom of the form, there are two buttons: "Salvesta" (Save) and "Loobu" (Cancel).

Joonis 4. CV täitmise samm 3 – Töökogemus.

Väljatöötatava soovitusüsteemi mõistes on oluliseks elementideks ametikoha valdkond ning töötatud ajavahemik. Teised väljad on oma ebastandardsuse tõttu kahjuks raskesti kasutatavad ning jäätakse seeläbi kõrvale.

Neljandas, hariduskäigu sammus, saab kandidaat sisestada informatsiooni oma senise hariduse kohta. Hariduskäiku võimaldab ITP sisestada ühe kaupa jaotatuna koolide ja haridustasemete järgi. Uue hariduskäigu lisamiseks tuleb vajutada nupule „Lisa uus“, misjärel avaneb vaade, kus kandidaadil on võimalik sisestada hariduskäigu detailsed andmed ühe kooli ja haridustaseme piires. Loodava soovitusüsteemi puhul on oluliseks elemendiks „Haridustase“. Kui haridustase on muu, kui puudulik, alg-, põhi- või keskkharidus, siis võetakse arvesse ka haridustasemega omandatud eriala ja valdkond, sest vastasel korral haridustase eriala omandamist ette ei näe. Et rohkem välju antud juhul andmetes standardiseeritud kuju ei oma, jäävad nad ka loodava süsteemi puhul kõrvale.

Viiendal, täiendkoolituse sammul, on kandidaadil võimalik sisestada informatsiooni täiendkoolituste, ametitunnistuste ja kutsetunnistuste kohta (vt. Joonis 5). Täiendkoolituse kohta lisatavad andmed pole taaskord standardiseeritud ning jäävad vaatluse alt kõrvale. Ametitunnistuse lisamiseks tuleb, sammus täiendkoolitus, vastavas sektsioonis vajutada nupule „Lisa uus“, millele järgneb ametitunnistuse andmeid kandidaadilt küsiva vormi avanemine. Loodava soovitusüsteemi raames on olulised järgmised standardiseeritud andmeid sisaldavad väljad: „Tunnistus“ ja „Kehtib kuni“.

CV-le kutsetunnistuse lisamiseks peab kandidaat andma nõusoleku andmete importimiseks Eesti kutseregistrist Kutsekoda, vajutades vastavas sektsioonis nupule „Võta andmed kutsekojast“. Kutsetunnistuse näol on samuti tegemist loodava soovitusüsteemi jaoks olulise elemendiga.

* Kohustuslikud andmed

Täiendkoolitus

[+ Lisa uus](#)

Asutus, koolitus	Maht	Aeg	Muuda	Kustuta
------------------	------	-----	-------	---------

Ametitunnistused

[+ Lisa uus](#)

Sertifikaat, pädevustunnistus, mootorsõidukijuhi ametikoolituse läbimise kutsetunnistus.

Tunnistus, number	Väljastaja	Kehtib kuni	Muuda	Kustuta
-------------------	------------	-------------	-------	---------

Kutsetunnistused kutseregistrist

[Võta andmed kutsekojast](#)

Kui kutseregistrist ei tule infot teie kutsetunnistus(t)e kohta, võtke palun ühendust kutsekojaga:
<http://www.kutsekoda.ee/et/kutsekoda/kontakt>.

[Salvesta ja jätk](#) [Loobu](#) [Tagasi](#)

Joonis 5. CV täitmise samm 5 - Täiendkoolitused.

Kuuendas, oskuste kirjeldamise sammus on enamus kandidaadi käest küsitav informatsioon standardiseeritud kujul ning sisend võetakse ühtlasi ka loodava soovitussüsteemi raames arvesse. Arvesse võetavateks elementideks on kandidaadi poolt enda CV-sse lisatud keeleoskused, omatavad mootorsõidukite juhtimislubade kategooriad, isikliku auto kasutamise võimalus, arvutioskused (vt. Joonis 6).

* Kohustuslikud andmed

Keeleoskus
+ Keeleoskus

Keel
Kõnes
Kirjas
Muuda
Kustuta

Juhiload

A1 B1 C1 D1 BE C1E D1E R ADR
 A B C D AM CE DE T Tõstuk

Isikliku auto kasutamise võimalus

Kui te saate tööks kasutada isiklikku autot, tehke siia linnuke.

Kui märgite, et saate kasutada isiklikku autot, kontrollige üle, et oleks märgitud ka juhiload

Arvutioskused

e-posti kasutamine inseneritarkvara andmebaaside haldus
 interneti kasutamine arvuti graafika riistvara haldus ja nälgaldus

Joonis 6. CV täitmise samm 6 - Oskused.

CV loomise viimases, seitsmendas sammus olevaid andmeid ei saa nende ebastandardsuse tõttu kasutada ning jäetakse loodava soovitusüsteemi raamest välja.

2.2.1.3 Töösoov

Töösoov on ITP funktsionaalsus, mis võimaldab töötajal piltlikult Töötukassale öelda, milliseid töökohti ta enim eelistab. Töösoov on töötaja profiilis, CV kõrval, oluline lisa, kuna CV-s kajastatud kogemused ja oskused ei peegelda alati töötaja eelistusi. Töösoov seda aga teeb, mistõttu tegemist on loodava soovitusüsteemi jaoks väga olulise funktsionaalsusega.

Töösoov on oma olemuselt veebivorm, mida kasutajal on võimalik täita ning mis läbi saab töötaja töösoovid standardsel kujul.

Töösoovi loomine koosneb sisuliselt kahest sammust, kus esimeses sammus sisestab töötaja enim soovitud ametikoha standardsel kujul. Standardne kuju tähendab seda, et kasutaja peab valiku tegema süsteemis sisalduvate ametinimetuste hulgast (vt. Joonis 7).

Töösoovi lisamine

Ametikoht * autoj

- laohoidja-autojuht
- rahvusvaheliste vedude autojuht
- sõiduautojuht
- veoautojuht

Salvesta Loobu

Joonis 7. Töösoovi lisamine – esimene samm.

Seejärel pärast „Salvesta“ nupu vajutamist luuakse töösoov ning kasutaja suunatakse ümber töösoovi detailsema vormi peale. Võttes arvesse esimeses sammus tehtud valikut, määrab süsteem soovitud ametile valdkonna. Vastava valdkonna piires on tööotsijal võimalus teises sammus lisada töösoovile veelgi täiendavaid ametikohti. Lisaks ametikohtadele on tööotsijal võimalus töösoovi märkida kuni linna või valla täpsusega soovitud töö piirkonnad. Veel on tööotsijal võimalik märkida soovitud tööaeg ning viimasena töötasu (vt. Joonis 8). Kõik nimetatud elemendid on loodava soovitusüsteemi raames olulised.

Töösoov: sõidukijuhtimine, tõstukijuhtimine

Ametikohad valdkonnas

Lisage vasakul olevast kastist sobivaid ameteid oma töösoovile juurde. Mida rohkem ameteid valite, seda nähtavam olete töopakujatele.

Võimalikud valikud:

- kraana-, tõstuki- vm tõsteseadmete juht
- farmi- ja metsatöömasina juht
- maaparandus-, teemasina vms juht
- bagerijuht, süvendusmasinajuht
- trammijuht
- pindamismasina juht
- trollibussijuht
- autokraanajuht

Valitud valikud:

- rahvusvaheliste vedude autojuht

Töö piirkond

kogu Eesti + Lisa uus

Riik	Asukoha lisainfo
Ühtegi asukohta pole lisatud	

Sobiv tööaeg

hooajaline osaline täistööaeg vahetustega

Töötasu

Netopalk kuus alates EUR

Salvesta Tagasi

Joonis 8. Töösoovi lisamine – teine samm.

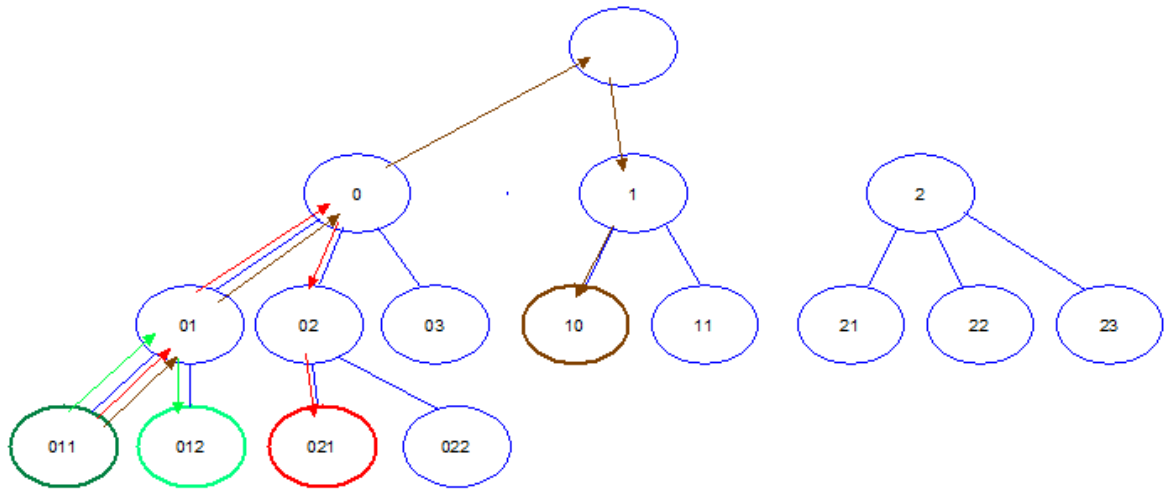
3. Tööturuameti Selveinfosüsteem

Tööturuamet oli aastatel 2005 - 2009 Sotsiaalministeeriumi valitsemisalasse kuuluv valitsusasutus, mille tegevusvaldkondadeks olid riigi tööturupoliitika elluviimine, tööturuteenuste osutamise, tööturutoetuste maksmise korraldamine, töötute ja tööotsijate üle arvestuse pidamine jms (Tööturuameti põhimäärus. - RTL 2005, 121, 1894). Tulenevalt Sotsiaalministri määruste muutmisest 2009. aastal - Tööturuameti ja Eesti Töötukassa funktsioonid liideti ning nende ühiseks nimeks sai Eesti Töötukassa (Sotsiaalministri määruste muutmine. – RTL 2009, 36, 481).

Selveinfosüsteem näol oli tegemist Tööturuameti kasutusel olnud infosüsteemiga, mille põhieesmärk oli infotehnoloogiliselt süstematiseerida Tööturuameti igapäevatööd. Selveinfosüsteemi põhifunktsionaalsuste hulga kuulus ka töövahendus, mille raames olid välja töötatud mõned tööpakkumise ja kandidaadi sobitamise algoritmid (Selveinfosüsteem, 12.03.2015).

Töö autoril on ligipääs dokumendile, kus on kirjeldatud mõningaid selveinfosüsteemis kasutusel olnud sobitusalgoritme. Järgnevalt tuuakse välja kolm, dokumendis kirjeldatud sobitusalgoritmi. Autori hinnangul on välja toodavad algoritmid kasulikud uue soovitusüsteemi välja töötamise juures. Kuidas täpsemalt, seda kirjeldab autor järgnevates peatükkides.

3.1 Hierarhiliste klassifikaatorite sobivusprotsendi arvutamine



Joonis 9. Hierarhiline puu ning nende sõlmede vahel liikumine.

Igale joonisel oleva puu sõlmele vastab unikaalne klassifikaatori väärtus (vt. Joonis 9). Kaugust klassifikaatori kahe väärtuse vahel võib ette kujutada sammude arvuga, mis tuleb sooritada liikudes ühest sõlmest teise. Liikuda saab ainult vertikaalselt, st näiteks sõlmest 011 sõlme 012 liikumine toimub kahe sammuga: 011 – 01 – 012 (joonisel kujutatud roheliste nooltega). Analoogiliselt sõlmest 011 sõlme 10 liikumiseks tuleb sooritada viis sammu: kolm sammu üles ning kaks alla.

Seega saab kahe klassifikaatori väärtuse protsentuaalset sobivust väljendada valemiga:

$$P = \left(1 - \frac{k_u n_u + k_d n_d}{(N + 1)(k_u + k_d)}\right) * 100\%$$

kus k_u on üles sooritatud sammu kaal, n_u - üles sooritatud sammude arv, k_d - alla sooritatud sammu kaal, n_d - alla sooritatud sammude arv, N - tasemete arv hierarhias. Teguri N asemel kasutatakse tegurit $N + 1$ selleks, et sobivus ei oleks kunagi null.

3.2 Parameetrite komplektide protsentuaalne sobivus

Parameetrite komplektide H_1 ja H_2 protsentuaalset sobivust saab väljendada valemitega:

$$P = \frac{N(H_1 \cap H_2)}{N(H_1)} * 100\%$$

kus $N(H)$ tähistab elementide arvu hulgas H .

3.3 Parameetrite liitmine ja terviksobivuse leidmine

Kui üksikute parameetrite sobivused P_1, P_2, \dots, P_n on määratud, siis terviksobivus määratakse summana:

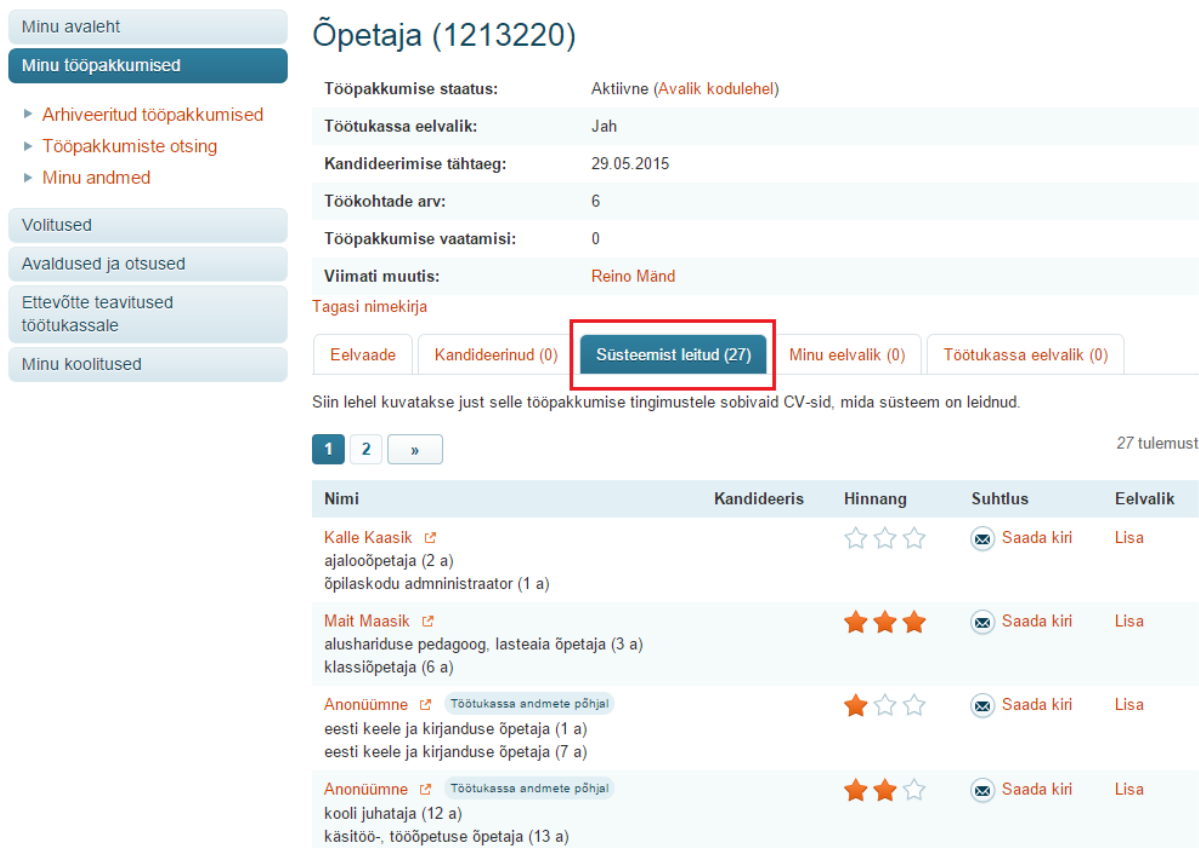
$$S = \sum k_i P_i$$

kus parameetrite kaalude k_i suhtes kehtib seos

$$\sum k_i = 1$$

4. Kasutusel oleva soovitusüsteemi lahenduse analüüs

Tööpakkumisele leitud sobivate kandidaatide nimekiri on kasutajaliideselt tööpakkujale kuvatav sarnaselt Joonis 10 nähtuga.



The screenshot shows a user profile for 'Õpetaja (1213220)' with various search filters and a list of candidates. The 'Süsteemist leitud (27)' button is highlighted with a red box.

Õpetaja (1213220)

Toöpakkumise staatus: Aktiivne (Avalik kodulehel)

Töötukassa eelvalik: Jah

Kandideerimise tähtaeg: 29.05.2015

Tookohtade arv: 6

Toöpakkumise vaatamisi: 0

Viimati muutis: Reino Mänd

Tagasi nimekirja

Eelvaade | Kandideerinud (0) | **Süsteemist leitud (27)** | Minu eelvalik (0) | Töötukassa eelvalik (0)

Siin lehel kuvatakse just selle tööpakkumise tingimustele sobivaid CV-sid, mida süsteem on leidnud.

1 2 » 27 tulemust

Nimi	Kandideeris	Hinnang	Suhtlus	Eelvalik
Kalle Kaasik ajalooõpetaja (2 a) õpilaskodu administraator (1 a)		☆☆☆	✉ Saada kiri	Lisa
Mait Maasik alushariduse pedagoog, lasteaia õpetaja (3 a) klassiõpetaja (6 a)		★★★	✉ Saada kiri	Lisa
Anonüümne <small>Töötukassa andmete põhjal</small> eesti keele ja kirjanduse õpetaja (1 a) eesti keele ja kirjanduse õpetaja (7 a)		☆☆☆	✉ Saada kiri	Lisa
Anonüümne <small>Töötukassa andmete põhjal</small> kooli juhataja (12 a) käsitöö-, tööõpetuse õpetaja (13 a)		★★☆	✉ Saada kiri	Lisa

Joonis 10. Tööpakkujale ITP kasutajaliidese vahendusel nähtav ja hallatav tööpakkumise raames süsteemi poolt leitavate sobivate kandidaatide nimekiri.

Tööpakkumisele sobivate kandidaatide nimekiri koostatakse kompleksse, dünaamiliselt kokku genereeritud SQL päringu alusel, mille loomine koosneb sisuliselt kolmest sammust. Esimeses sammus võetakse tööpakkumise vormilt nimekirja koostamise aluseks olevad parameetrid, mida tööpakkuja sisestas. Viimased saavad dünaamiliselt koostatava SQL päringu ühtedeks sisendparameetriteks. Teise sammuna leitakse parameetrid, millega tööpakkumise parameetreid omavahel võrdlema hakatakse. Nendeks on kasutaja CV ning töösoovide vormidele sisestatud parameetrid. Kolmas samm on mainitud parameetrite omavaheline võrdlemine. Millest tuleneb aga SQL päringu dünaamika, illustreerib järgnev. Kui tööpakkumise vormil on näiteks küsitud, kas kandidaadil on isikliku auto kasutamise

võimalus, siis kasutaja jaatava või eitava vastuse korral kantakse tingimus dünaamiliselt ka vastavalt SQL päringusse sisse või jäetakse sealt välja. Kui sisestatud tööpakkumise parameeter on võrreldav töösoovide või CV parameetriga, siis ainult sellisel juhul lisatakse antud parameetreid SQL tingimustesse.

Vaatlemaks kõne all olevat SQL päringut täpsemalt, jagab töö autor lihtsama mõistetavuse huvides päringu elemendid parameetrite järgi plokkideks. Vastavalt sellele, mis parameetrid on tööpakkumises nõutud, on ka plokk vastavalt kas siis esindatud või mitte. Kui plokk on esindatud, kvalifitseerub kandidaat nimekirja juhul, kui plokis kirjeldatud tingimus on täidetud. Kõiki plokkide ühendab päringus loogiline jaatus. See tähendab seda, et nimekirja pääseb ainult see kandidaat, kes kõik esindatud plokid ning omakorda plokkides kirjeldatud tingimused ära täidab.

Kõiki võimalikke plokkide ja plokki tingimusi kirjeldab järgmine:

Ametinimetus:

Kandidaadi töösoovi mistahes ametinimetus peab vastama vähemalt ühele tööpakkumisega seotud ametinimetusele.

VÕI

Kandidaadi vähemalt üks töösoovi ametinimetuse valdkond peab vastama tööpakkumise valdkonnale.

Töö asukoht:

Kui tööpakkumisel on töö asukohaks märgitud välisriik siis kvalifitseeruvad kandidaadid kellel on töösoovis märgitud soovitud töö asukohaks välisriik.

VÕI

Kui tööpakkumisel on märgitud, et töö asukohaks on kogu Eesti, siis kvalifitseeruvad kandidaadid, kes töösoovis on samuti märkinud valiku „kogu Eesti“.

VÕI

Kui tööpakkumisel on töö asukohad täpsemalt ära märgitud, siis kvalifitseeruvad need kandidaadid, kes on sobivateks töö asukohtadeks märkinud kogu Eesti või kui töösoovis kattub vähemalt üks aadressi element töö asukoha aadressi elemendiga.

Töökogemus (aastates):

Kandidaadi töökogemus nõutud valdkonnas peab olema sama pikk või pikem, kui tööpakkumises märgitud. Tingimuse kontrollimiseks liidetakse kandidaadi CV-l märgitud töötatud aastad nõutud eriala piires kokku ja võrreldakse tööpakkumisel märgitud nõudega.

Haridus:

Kandidaadi CV mistahes hariduse valdkond peab vastama vähemalt ühele tööpakkumises nõutud valdkonnale.

JA

Kandidaadi CV kõrgeim haridustase, eelnevalt leitud valdkonnas, peab olema sama või kõrgem tööpakkumises nõutud haridustasemest.

Kutsetunnistus:

Kvalifitseeruvad ainult need kandidaadid, kellel tööpakkumisel nõutud kutsetunnistus on CV-le märgitud ja on kehtiv.

Juhiload:

Kvalifitseeruvad ainult need kandidaadid kellel on kõik töökuulutusel nõutud juhilubade kategooriad CV-le märgitud.

Isiklik auto:

Kvalifitseeruvad ainult need kandidaadid, kes on oma CV-se ära märkinud isikliku auto kasutamise võimaluse positiivselt.

Täiendavad piirangud:

Päringu tulemustesse sobivaid kandidaate otsitakse hulgast kellel leidub aktiivne CV ning leidub vähemalt üks töösoov

Sorteerimine:

Nimekirjas on esimesed kõik need kandidaadid, kelle vähemalt üks tööpakkumisega seotud ametinimetus kattub vähemalt ühe töösoovis märgitud ametinimetusega. Neile järgnevad kandidaadid, kes on kõige kauem töötanud staatuses olnud. Kandidaadid, kes ei ole töötanud staatuses jäävad sorteerimisel nimekirja lõppu.

Koostatava päringu plusse ja miinuseid on kirjeldatud lähtudes autori arvamusest ning mitmeaastasest kogemusest projektmeeskonna liikmena ning põhinedes reaalsele kliendi tagasisidele. Sellest lähtuvalt on plussid ja miinused järgmised:

Plussid:

- Tulemused tulevad kiirelt/vahetult
- Kandidaadid, kes päringus kehtestatud kriteeriumid täidavad sobivad suhteliselt hästi, sest see tähendab, et ka tööpakkumises kehtestatud kriteeriumid on enamjaolt täidetud.

Miinused:

- Pole konfigureeritav. Praeguses lahenduses parameetritel puutuvad konstantsed konfigureeritavad kaalud ja prioriteedid, mille järgi üht parameetrit teisest esile tõsta või vastupidi.
- Keeruline täiustada. Päringu dünaamika sisemiselt väga keeruline ja raskesti loetav.
- Kõiki tööpakkumises sisestatud nõudeid on kandidaatide leidmisel kohustuslik täita - ehk kandidaate välistatakse küllaliski kergekäeliselt. Näiteks, kui kandidaadil sobivad kõik parameetrid peale ühe täiuslikult, siis teda tulemusse ei valita. Välistamise põhjus võib olla reaalses elus aga sageli kergelt ületatav.
- Ei võeta arvesse kõiki parameetreid, mida tegelikkuses töösoovide ja CV vormidele sisestada saab. Puuduvad näiteks ametitunnistuse, keeleoskuse, tööaja, töötasu ning arvutioskuse võrdlus.
- Sorteerimisel võetakse arvesse ainult kahte parameetrit, mis aga ei taga alati ootuspärast kandidaatide järjestust.

5. Uue soovitusüsteemi lahenduse välja töötamine

Kuna antud töö keskendub otsesel parameetrite võrdlemisel põhinevatel meetoditel, jätame me kõrvale kaudsed statistilisel korrelatsioonil põhinevad meetodid.

5.1 Tööpakkumiste ja töösoovide ning CV-de parameetrite ühisosa

Varasemates peatükkides läbi viidud süsteemi analüüsi käigus anti ülevaade uue lahenduse jaoks olulistest parameetritest. Veelgi parema ülevaate saamiseks koondab töö autor need alljärgnevasse tabelisse.

Tabel 1 - Tööpakkumiste, töösoovide ning CV-de parameetrite ühisosa ülevaade

Parameetrid	Tööpakkuja	Kandidaat	
	Tööpakkumine	CV	Töösoov
Standardne ametinimetus/ametivaldkond			
Töö asukoht			
Töökogemus (aastates)			
Kutsetunnistus			
Hariduse valdkond			
Ametitunnistus			
Haridustase			
Keeleoskus			
Juhiload			
Tööaeg			
Töötasu			
Arvutioskus			
Isiklik auto			
Töösuhte kestus			
Töötasu tüüp			
Tööle asumise aeg			

	esindatud
	mitte esindatud

Tabel 1 on välja toodud parameetrid, mida oleks võimalik soovitusüsteemi mõistes parameetrite võrdlemismeetodites ära kasutada. Tabel näitab, milliseid parameetreid võimaldab süsteem sisestada tööpakkujal ja milliseid kandidaadil. Näiteks „Tööpakkumine“

veerus olev parameeter „Kutsetunnistus“ on märgitud roheliseks ning seda ka „CV“ veerus. Samas „Töösoov“ veerus on reaalse vastava veeru väärtus märgitud valgeks. Viimane tähendab seda, et parameeter on esindatud tööpakkumise ning „CV“ moodulis, kuid mitte „Töösoov“ moodulis.

Tabelis on näha ka, et osade parameetrite osas tööpakkuja ja kandidaadi vaheline ühisosa puudub. Töö autor peab neid parameetreid küll olulisteks, kuid kuna ühisosa praeguses ITP poolt realiseeritud lahenduses puudub ja suurt pilti mõjutab see üldjoontes vähe, jätab autor kõnealused parameetrid antud töös käsitlemata.

5.2 Andmemahud

Tabel 2 - Ülevaade parameetrite mahtudest ITP süsteemis seisuga 23.03.2015

	Tööpakkumine: 193 550	CV: 51 550	Töösoov: 211 777
Standardne ametinimetus/ametivaldkond	236 580	161 392	309 856
Töö asukoht	202 242	51 550	355 818
Töökogemus (aastates)	193 550	161 392	
Kutsetunnistus (kutsekojast)	193 550	3 675	
Hariduse valdkond	9 026	65 308	
Ametitunnistus	2 110	2 186	
Haridustase	193 550	65 308	
Keeleoskus	297 093	395 265	
Juhiload	23 238	70 983	
Tööaeg	103 154		340 223
Töötasu	193 550		211 777
Arvutioskus	64 438	136 706	
Isiklik auto	193 550	51 550	

Tabel 2 on välja toodud, seisuga 23.03.2015, kõiki süsteemi maha salvestatud parameetrite arv, mis on olulised tööpakkumisele sobivate kandidaatide leidmise protsessis. Tabeli päises välja toodud numbrid tähistavad põhielementide arvu. Näiteks „Tööpakkumine: 193 550“ tähendab, et andmebaasi on maha salvestatud kokku 193 550 tööpakkumist. Tabeli vasakusse veergu on loetletud põhielementidega seotud parameetrid ning tabeli sisus kuvatavad numbrid tähistavad põhielemendiga seotud parameetrite arvu. Oluline on veel kord ära märkida, et tegemist on kõikide parameetritega, ehk sisse on arvutatud ka arhiveeritud tööpakkumised, mitteaktiivsed CV-d ning samuti töösoovid millele ei leidu vastavusse aktiivset CV-d. Sisse

arvutatud on need tulenevalt infosüsteemi andmebaasi struktuurist, kus ka mitteaktiivsed kirjed asuvad samades andmetabelites aktiivsetega ning mis teatud päringute puhul vajavad rohkemal või vähemal määral siiski protsessimist.

Kuna antud töö keskendub eeskätt otsestele parameetrite võrdlemisele, on sellest tulenevalt mitteaktiivsed kirjed esialgu väheolulised. Selle tulemusel tegi autor uued päringud, millest saame teada, et mitte-arhiveeritud tööpakkumisi on ligikaudu 1500, aktiivses staatuses CV-sid ligikaudu 28 000 ning aktiivseid töösoove ligikaudu 87 000. Sama suhtega on vähem ka nendega seotud parameetreid.

Kuna ühel kandidaadil saab töösoove olla mitu siis järgneva analüüsi huvides toob autor välja ka võrdluse, mitu aktiivset CV-d omavat kandidaati omavad ka vähemalt ühte töösoovi. Vastava andmebaasipäringu tulemuseks sai autor ligi 26 000 kirjet. Mis tähendab, et 28 000 aktiivse CV-ga on seotud 26 000 töösoovi, mis protsentuaalselt on ligikaudu 92 protsenti.

5.3 Algoritmid

Autor leiab, et parima tööpakkumisele sobiva kandidaadi leidmise tulemuse saavutamiseks tuleks igale kandidaadi parameetrile välja arvutada protsentuaalne sobivus tööpakkumises nõutud parameetritega. Seejärel tuleks välja arvutada iga kandidaadi terviksobivus tuginedes parameetrite sobivusprotsentidele, mille tagajärjel on võimalik moodustada kandidaatidest pingerida, mille tipus olevad kandidaadid sobivad hinnanguliselt antud tööpakkumisega kõige paremini.

Kuna andmeid on palju näeb autor ette teatavaid ressursiprobleeme. Nende olemusest ja võimalikest lahendustest tuleb juttu töö edenedes. Esmalt aga annab aga autor detailsema ülevaate kuidas parameetreid võrrelda ja kuidas arvutada parameetrite sobivusprotsente.

Autor keskendub antud töös olulisematele parameetritele, mis tähendab, et mõningate vähem oluliste parameetrite sobivusprotsentide arvutusmeetodid jäetakse kirjeldamata. Milliseid parameetreid peab autor olulisteks, selgub aga antud peatüki järgmistes alapunktides.

Autor mainib ära veel selle, et tööpakkumisele kandideerimiseks peab kandidaadil olema olemas CV, millest tulenevalt saame soovitusüsteemi välja töötamisel tugineda eeldusele, et kandidaadil eksisteerib CV. Töösoov ei ole kohustuslik, kuid selle puudumine nõrgestab

oluliselt kandidaadi positsiooni soovitude nimekirjas. Tuginedes statistikale on hetkel süsteemis ligikaudu 92 protsenti aktiivsetest CV-dest seotud ka vähemalt ühe töösooviga.

Kuigi CV sisaldab piisavalt palju andmeid, millele soovitus genereerimisel baseeruda, leiab autor, et esmase versiooni väljatöötamisel tuleb kasuks CV-sid ja töösoove koos vaadata. Seetõttu seab autor eelduseks selle, et ka töösoov peab kandidaadil eksisteerima. Ehk antud töö raames jäetakse soovitusnimekirjast välja kandidaadid, kellel töösoov puudub.

5.3.1 Ametinimetuste protsentuaalne sobivus

5.3.1.1 ISCO ja ISCO_TK

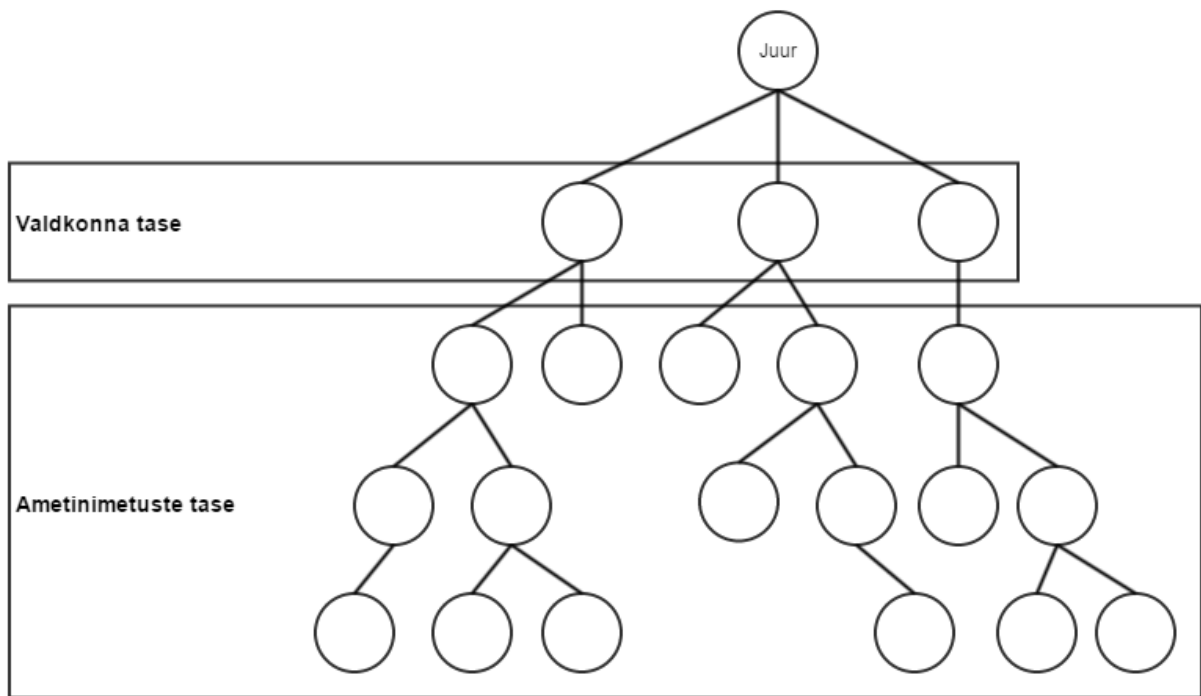
ISCO (ingl k International Standard Classification of Occupation) standard on üks peamisi klassifikaatoreid, mille eest International Labour Organization (ILO) on vastutav. See kuulub rahvusvahelisse majandus ja sotsiaalklassifikaatorite hulka. ISCO on tööriist organiseerimaks ameteid selgelt defineeritud gruppide moodustuvasse hulka, kusjuures gruppide moodustamisel võetakse arvesse grupeeritavate ametite poolt läbi viidavaid tööülesandeid ning ametil lasuvaid kohustusi. ISCO standardi eesmärk on võimaldada:

- luua ühtne baas rahvusvaheliste raportite ja võrdluste koostamiseks ning statistiliste ja administratiivsete andmete vahetamiseks;
- luua mudel, mille põhjal arendada riikliku ja regionaalset ametite klassifikatsioone;
- luua süsteem, mida saab otseselt kasutada riikides, mis ei ole välja töötanud endi riiklikke ametite klassifikatsioone.

Standard on ühtlasi loodud eesmärgiga, et ta oleks kasutatav statistilistel, kui ka hulgal erinevatel kliendile orienteeritud aplikatsioonidel. Kliendile orienteeritud aplikatsioonid sisaldavad seejuures tihtipeale töötajate ja vabade töökohtade ühendamist, lühi- ja pikaajalist töötajate rahvusvahelist migratsiooni. (ILO – ISCO 12.02.2015)

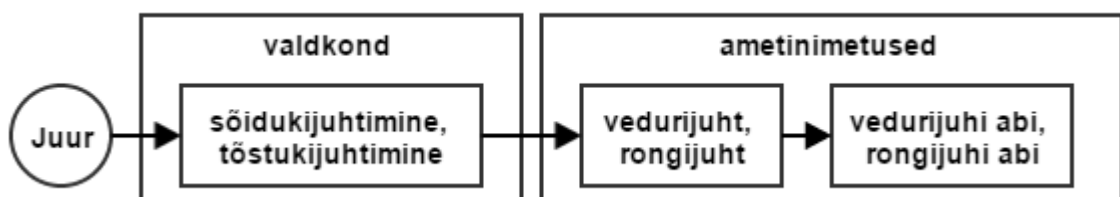
Töötukassa infosüsteem ei kasuta otseselt ISCO standardit vaid on loonud ISCO standardi põhjal oma niinimetatud standardi, et paremini sobituda süsteemi vajadustega. Sellest tulenevalt on Töötukassal võimalus vastavalt vajadusele ametinimetusi lisada, muuta ja kustutada. Nimetame seda standardit edaspidi kui ISCO_TK.

ISCO_TK standardi struktuuri kirjeldab Joonis 11.



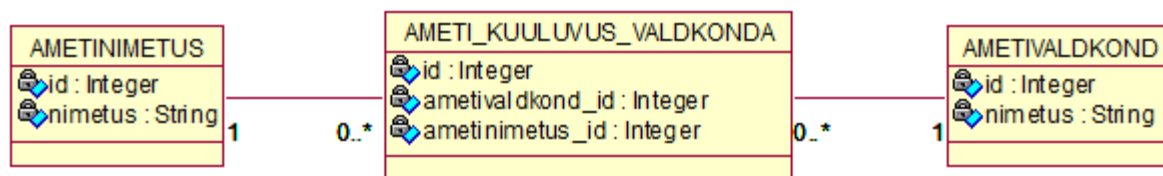
Joonis 11. ISCO_TK hierarhiline struktuur.

ISCO_TK standard on sarnaselt ISCO standardile, klassifikaatoritest koosnev hierarhia. ISCO_TK hierarhias on kokku viis taset. Esimene tase on juure tase, mis kõik klassifikaatorid omavahel kokku seob. Teine tase on valdkonna tase, mis kõik ametinimetused suurematesse ametivaldkonnadesse jagab. Kolmas, neljas ja viies tase on ametinimetuste tasemed, mille hulgast juba realselt valitakse klassifikaatoreid tööpakkumistele ja töösoovidele. (vt. Joonis 11)



Joonis 12. Struktuuriline esitus reaalse ametinimetuste näitel.

Töötukassa infosüsteemis hoitakse ametinimetuste andmeid kokku kolmes tabelis. Esimene tabel sisaldab valdkondade klassifikaatoreid, teine tabel on vahetabel, mis seob kokku kõik valdkonnad kõikide ametinimetustega ning kolmas tabel sisaldab kõiki ametinimetuste klassifikaatoreid. (vt. Joonis 13)



Joonis 13. Ametinimetuste andmete struktuuri kirjeldav konseptuaalmudel.

5.3.1.2 Klassifikaatori protsentuaalse sobivuse arvutamine

Klassifikaatorite sobivusprotsendi arvutamiseks kasutab autor eelpool kirjeldatud hierarhiliste klassifikaatorite sobivusprotsendi arvutamise algoritmi.

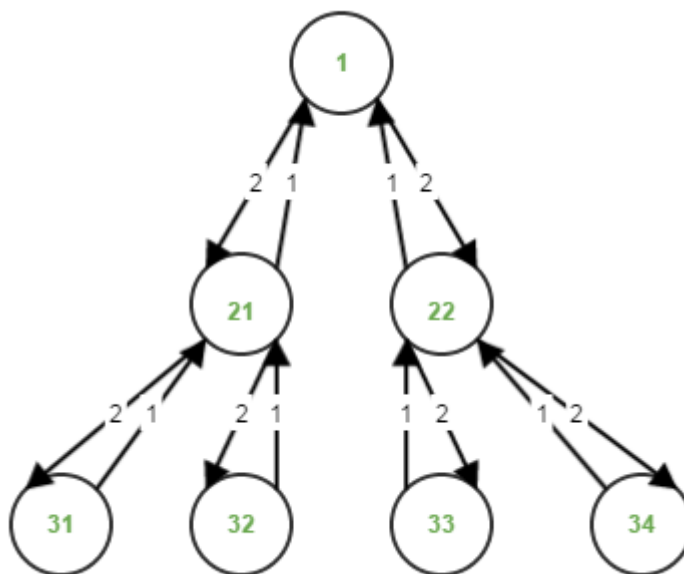
Autor keskendub antud töös, ametinimetuste protsentuaalse sobivuse leidmisel ainult töösoovides sisalduvatele ametinimetustele kuna, nagu ka eespool mainitud, CV ei sisalda töökogemuse näol standardseid ametinimetusi.

Sobivusprotsendi leidmiseks kahe ametinimetuse vahel kasutame ära ülal kirjeldatud hierarhiliste klassifikaatorite sobivusprotsendi arvutamise valemit. Selleks vajavad asendamist järgmised tegurid: k_u on üles sooritatud sammu kaal, n_u - üles sooritatud sammude arv, k_d - alla sooritatud sammu kaal, n_d - alla sooritatud sammude arv, N - tasemete arv hierarhias. k_u ja k_d on valemi konfigureerimiseks ning on olulisteks teguriteks alles rakenduse lõppkasutuses. Seetõttu välja töötamise faasis annab tööautor intuiitiivselt kaaludeks $k_u = 1$; $k_d = 2$. N on tasemete arv hierarias ning tulenevalt ISCO_TK standardist on selleks $N = 5$. n_u ja n_d on vastavalt ülese- ja alla sooritatud sammude arv. Nimetatud tegurite väärtustamine on ühtlasi kõnealuse valemi raskuspunkt. Autor kasutab n_u ja n_d väärtustamiseks „Dijkstra“ algoritmi.

Dijkstra algoritm on Edsger Wybe Dijkstra poolt 1959. aastal avaldatud graafi läbimise algoritm, mis leiab sidusas graafis lühimad teed algtipust kõigisse teistesse tippudesse (Wikipedia - Dijkstra algoritm, 12.04.2015).

Selleks et „Dijkstra“ algoritmi kasutada on vaja praegu kasutusel olev ametinimetuste andmete struktuur töödelda ümber, et saavutada algoritmi kasutamiseks sobiv struktuur. Piltlikult kirjeldab sobivat struktuuri järgnev joonis (vt. Joonis 14). Joonisel on kujutatud hierarhilist graafi, kus iga sõlm, antud kontekstis klassifikaator, sisaldab endas ühte ametinimetust ning iga serva peal olev väärtus kirjeldab kahe klassifikaatori vahelist teekonna pikkust ehk antud kontekstis kaalu. Leidmaks kahe klassifikaatori vahelist kaugust „Dijkstra“

algoritmi kasutades, saame tulemuseks täpselt selle, mida hierarhiliste klassifikaatorite sobivusprotsendi arvutamise valemis nõutud: $k_u n_u + k_d n_d$ ehk ülese sooritatud sammude arv korrutatud ülese sooritamise sammu kaaluga ning liidetud alla sooritatud sammude arv korrutatud alla sooritamise sammu kaaluga.



Joonis 14. Ametinimetuste klassifikaatorite suunatud graaf

Andmete sobiv struktuur andmebaasi tasandil peaks olema selline, mis kirjeldab iga klassifikaatori ühesuunalist seost teiste klassifikaatoritega. Ühesuunalist seetõttu, kuna graafi tingimus näeb ette, et kui kahe klassifikaatori vahel on seos, siis on see alati kahesuunaline ning kõikide suundade maha salvestamine oleks üleliigne ja tekitaks asjatult poole rohkem andmemahtu. Autor valib maha salvestamiseks ülevalt-alla-suunalised seosed ehk joonisel kõik servad, mille kaaluks on märgitud 2 (vt. Joonis 14). Tabeli esitust antud joonise põhjal illustreerib Tabel 3.

Tabel 3 - Hierarhilise struktuuri tabeliesitus Joonis 14 näitel.

Klassifikaator 1	Klassifikaator 2
1	21
1	22
21	31
21	32
22	33
22	34

Kui kõik ametinimetuste andmed on viidud andmebaasi tasandil sobivale struktuurile saame kergesti kasutusele võtta „Dijkstra“ algoritmi, kus algoritmi algandmeteks ongi loodud seoste tabel. Kuna loodud seoste tabel kirjeldab ainult ühesuunalisi seoseid, siis on oluline meele pidada, et andmete ettevalmistamise hetkel tuleb igale seosele juurde luua ka teine vastupidine seos ning seadistada kaalud.

Kui andmed on ette valmistatud, avanebki võimalus algoritmilt küsida kahe klassifikaatori kaugust üksteisest. Joonis 14 näitel peaks algoritm näiteks klassifikaatorite 31 ja 33 vaheliseks kauguseks tagastama $1+1+2+2 = 6$.

Ühtlasi on sellega kõik tegurid valemi jaoks leitud ning kahe klassifikaatori protsentuaalne sobivus on võimalik välja arvutada. Joonis 14 näitel klassifikaatori 31 ja 33 sobivusprotsenti arvutuskäik näeks välja järgmine:

$$P = \left(1 - \frac{k_u n_u + k_d n_d}{(N + 1)(k_u + k_d)}\right) * 100\%$$

kus $k_u = 1$; $k_d = 2$; $N = 5$; $k_u n_u + k_d n_d = 6$. Asendades tegurid valemis saame:

$$P = \left(1 - \frac{6}{(5 + 1)(1 + 2)}\right) * 100\%$$

$$P \approx 67\%$$

Klassifikaatorite 31 ja 33 protsentuaalne sobivus on ligikaudu 67%.

Kui sobivused on välja arvutatud, saame tekkinud tabeli põhjal pärida sobivaid CV-sid, kus on vahetult teada seda, kui hästi töösoov ja töökuulutuse ametinimetus omavahel sobivad.

5.3.2 Töö asukoha protsentuaalne sobivus

Autor leiab, et parima tulemuse saavutamiseks tuleks määrata töösoovis märgitud töö asukoha ja tööpakkumises märgitud töö asukoha sobivusprotsent, võttes arvesse nende vahelist vahemaad kilomeetrites. Kilometraažil põhineva sobivusprotsendi arvutamine välistatakse antud töö raames, seda vajalike andmete puudumise ning keerukuse olulise kasvu tõttu.

Kuna aga aadresside näol on tegemist samuti hierarhiliste klassifikaatoritega oleks nendegi puhul siiski võimalik arvutada klassifikaatorite sobivusprotsent kasutades eelpool kirjeldatud hierarhiliste klassifikaatorite sobivusprotsendi arvutamise meetodit. Küll aga ei pruugi antud

valem toimida ootuspäraselt, sest kui töösoovis sobiv piirkond on märgitud maakonna täpsusega, näiteks Järvamaa, ning tööpakkumisel on samuti sobiv piirkond maakonna täpsusega, milleks on mõni teine maakond, näiteks Saaremaa, siis valemi järgi need sobiks küllaltki hästi kuna nende vahel on samal hierarhilisel tasandil asumise tõttu ainult 2 sammu, kuigi reaalses elus võib nende omavaheline distantis väga suures ulatuses varieeruda. Viimase tõttu otsustatakse antud meetodit mitte-kasutada.

Sobiva lahenduse välja töötamisel lähtub autor sellest, kui täpselt on võimalik töösoovis sobivat töö asukohta kirjeldada. Piirkondadeks saab olla:

- kogu Eesti,
- maakond,
- maakond ja linn/vald,
- välismaa.

Antud töös keskendub autor esimesele kolmele ehk jätab käsitlemata välismaa, kuna töövahendus välismaale on Töötukassal küllaltki madala prioriteediga ning sellest tulenevalt toetavad ka Töötukassa infosüsteemid töövahendust välismaale küllaltki vähese funktsionaalsusega.

Sobivus on 100-protsendiline kui tööpakkumisel märgitud töö asukoht on sama või täpsem kui töösoovil. Vastasel korral on sobivusprotsent 100-st protsendist väiksem. Millised need protsendid olla võiks on konfigureerimise küsimus ja on olulised alles lõppkasutuses. Hetkel määrab autor need intuitiivselt. Täpsem protsentide jagunemine on välja toodud Tabel 4.

Tabel 4 - Töö asukoha sobivusprotsentide jaotus

Töösoov \ Tööpakkumine	Eesti	Maakond	Maakond - linn/vald
Eesti	100%	100%	100%
Maakond	70%	100%	100%
Maakond - linn/vald	50%	70%	100%

Tulemuste saamiseks tuleb esmalt välja selgitada töö asukoha täpsusgrupp. Täpsusgruppide all mõtleb autor, kas väärtuseid „Eesti“, „maakond“ või „maakond – linn/vald“. Ehk kui tööpakkumisel on töö asukohaks märgitud kogu Eesti, on tööpakkumise täpsusgrupp „Eesti“.

Kui töö asukohaks on märgitud mõni maakond, on täpsusgrupiks „maakond“ ning kui töö asukohaks on märgitud maakond koos linna või vallaga, on täpsusgrupiks „maakond – linn/vald“.

Kui töö asukoha täpsusgrupp on teada, tuleb piltlikult Tabel 4 valida täpsusgrupile vastav veerg. Näiteks kui tööasukoha täpsusgrupiks on „maakond“, siis keskendume veerule „maakond“. Seejärel tuleb teha piltlikult kolm päringut töösoovide vastu:

1. Kõik töösoovid, mille soovitud töö asukohaks on märgitud kogu Eesti, sobivad 100%.
2. Kõik töösoovid, mille soovitud töö asukoht on märgitud maakonna täpsusega ning nõutud maakond klapib soovitud maakonnaga, sobivad 100%.
3. Kõik töösoovid, mille soovitud töö asukoht on märgitud linna või valla täpsusega ning nõutud maakond klapib soovitud maakonnaga, sobivad tabeli järgi 70%.

Kõik ülejäänud töösoovid, millele sobivusprotsent määramata jäi saavad külge vaikimisi sätestatud sobivusprotsendi. Täpne väärtus on jällegi oluline lõppkasutuses ning on seega konfigureeritav. Autor määrab antud töö raames selleks 10%.

5.3.3 Töökogemus aastates, protsentuaalne sobivus

Kui kandidaadi töökogemus nõutud valdkonnas on sama pikk või pikem kui tööpakkumises märgitud, on kandidaadi töökogemuse protsentuaalne sobivus 100%.

Kandidaadi töökogemuse protsentuaalse sobivuse leidmiseks tuleb kokku liita kõik CV-s märgitud töökogemused aastates, tööpakkumises nõutud valdkonna piires.

Sobivusprotsendi leidmisel kasutab autor valemit:

$$P = \frac{k}{n} * 100\%$$

kus k on kandidaadi töökogemus aastates, nõutud valdkonnas ning n – tööpakkumises nõutud miinimumtöökogemus aastates antud valdkonnas.

Kui tulemus on üle 100% tähendab, et kandidaadil on antud valdkonnas rohkem töökogemust kui tööpakkumises minimaalselt nõutud. Sääraste kandidaatide sobivusprotsent alandatakse kandidaadi terviksobivuse leidmise huvides siiski 100-le protsendile.

5.3.4 Kutsetunnistuse protsentuaalne sobivus

Hetkel lubab süsteem iga tööpakkumisega siduda maksimaalselt ühe kutsetunnistuse. Säärane piirang lihtsustab oluliselt kandidaatide kutsetunnistuse sobivusprotsendi määramist.

Nõutud kutsetunnistus saab kandidaadil, kas olla või mitte. Kui kutsetunnistus ei ole nõutud, arvestatakse kutsetunnistuse protsentuaalset sobivust kõigile kandidaatidele 100%. Kui kutsetunnistus on nõutud, siis nende kandidaatide kutsetunnistuse sobivusprotsent on 100 kelle CV sisaldab nõutud kutsetunnistust. Kui nõutud kutsetunnistus puudub on sobivusprotsendiks 0%.

Kutsetunnistuse parameetritele on andmete struktuuri poolest väga sarnane parameeter veel „Isiklik auto“. Sellest tulenevalt on ka nimetatud sobivusprotsendi arvutamine sarnane. Seetõttu jätab autor „Isiklik auto“ parameetri antud võtmes eraldi käsitlemata.

5.3.5 Ametitunnistuse protsentuaalne sobivus

Autor kirjeldab antud alapeatükis ära ametitunnistuste protsentuaalse sobivuse arvutamismeetodi. Ametitunnistusele, oma andmete struktuuri poolest, on väga sarnased veel: hariduse valdkond, keeleoskus, juhiloa ning arvutioskus. Sellest tulenevalt on ka nende sobivusprotsendi arvutamismeetodid väga sarnased, mistõttu loetletud parameetreid, antud võtmes, eraldi ei käsitleta.

Ametitunnistuse puhul on tegemist on paljude jah/ei väärtuste komplektiga. Sellises olukorras ei ole otstarbekas üksikut jah/ei väärtust sobitada teise üksiku jah/ei väärtusega, vaid vaadelda tervet komplekti tervikuna. Kahe väärtuste komplekti protsentuaalset sobivust näitab see, kui suur osa ühe komplekti väärtustest kuulub ka teise komplekti.

Parameetrite komplektide H_1 ja H_2 protsentuaalset sobivust saab väljendada valemitega:

$$P = \frac{N(H_1 \cap H_2)}{N(H_1)} * 100\%$$

kus $N(H)$ tähistab elementide arvu hulgas H .

Struktuuriliselt hoitakse kandidaadi kõiki ametitunnistuste parameetreid ühes tabelis, millel on üks-mitmele seos CV tabeliga.

Tulenevalt andmebaasi struktuurist on parameetrite ühisosa leidmine küllaltki mugav. Ühisosa leidmiseks tuleb teha päring ametitunnistuste tabelist, kus tingimuseks on seatud parameetri eksisteerimine nõutud ametitunnistuste parameetrite hulgas. Seejärel tuleb saadud tulemus grupeerida CV-de järgi ning kokku lugeda igas grupis sisalduvate parameetrite arv. Grupi lõikes saadud tulemus ongi iga kandidaadi ametitunnistuste parameetrite ühisosa tööpakkumises nõutud parameetritega.

Päring Oracle SQL näitel näeb välja järgmine:

```
SELECT CV_AT.CV_ID,  
       Count(1) AS YHISOSA  
FROM   CV_AMETITUNNISTUSED CV_AT  
WHERE  CV_AT.AMETITUNNISTUS_ID IN  
       ( :NOUTUD_ID_1, :NOUTUD_ID_2, ..., :NOUTUD_ID_N )  
GROUP BY CV_AT.CV_ID;
```

Kui ühisosa on leitud, on võimalik välja arvutada iga kandidaadi ametitunnistuste protsentuaalne sobivus tööpakkumises nõutuga kasutades ülal kirjeldatud valemit.

5.4 Terviksobivuse arvutamine

Nüüd, kus on välja töötatud meetodid, kuidas välja arvutada iga parameetrite grupi protsentuaalne sobivus eraldi, on aeg kokku arvutada terviksobivused ning välja selgitada parimad kandidaadid konkreetsetele tööpakkumistele. Terviksobivuse leidmisel kasutab autor varasemalt kirjeldatud terviksobivuse leidmise meetodit, tuginedes juba leitud parameetrite sobivusprotsentidele.

Eelpool selgitas autor parima tulemuse saavutamise ideed, kus kirjeldas iga kandidaadi iga parameetri protsentuaalse sobivuse välja arvutamist. Samas viitas autor ka teatavatele ressursiprobleemidele. Kuna tööpakkuja peab tööpakkumise juures nägema sobivate kandidaatide nimekirja suhteliselt vahetult, seab see teatavad piirangud. Kuna andmeid on palju, on selge, et sellel hetkel kui tööpakkuja soovib nimekirja näha, on arvutusmootori käivitamiseks juba liiga hilja, sest parameetreid on palju ja see võtaks liiga kaua aega. Sellest tulenevalt selgitab autor kahte võimaliku lahendust: sobivusprotsentide ette arvutamine ja parameetrite prioritseerimine.

5.4.1 Sobivusprotsentide ette arvutamine

Sobivusprotsentide ette arvutamise idee on taustatööna teatavad arvutused ette ära arvutada ning tulemused maha salvestada. Teoreetiliselt on võimalik ka kõik arvutused ette ära teha. Idee seisneb selles, et taustatöö arvutab välja kõikidele tööpakkumistele kõikide kandidaatide terviksobivusprotsendid ning salvestab need maha. Selle tulemusel saaks lõppkasutaja kõige täpsemad tulemused ja soovitud hetkel kõige kiiremini. Lõppkasutaja peaks tulemuste saamiseks sisuliselt tegema ühe päringu välja arvutatud andmete vastu.

Autor aga viitab, et tegemist on teoreetilist laadi lahendusega. Seda seetõttu, et tegemist on väga ressursirohke lahendusega. Tunnetuse ressursi rohkusest annab see, kui välja arvutada ligikaudne tulemuste arv, mida lõpptulemus peaks sisaldama. Ehk struktuuriliselt, kui lõpptulemus oleks tabel, milles on väljad: tööpakkumise_id, cv_id ning terviksobivusprotsent, siis ridade arvu selles tabelis näitab järgnev arvutuskäik: kui süsteemis on ligikaudu 1500 tööpakkumist ning ligikaudu 26 000 aktiivset CV-d, millega seotud vähemalt üks töösoov, siis ligikaudsest ridade arvust, kõnealuses tabelis, annab aimu see, kui need kaks välja toodud numbrit omavahel korrutada, ehk: $1500 * 26\ 000 = 37\ 500\ 000$. Tulemus on antud süsteemi mõistes väga suur, mis nõuab andmebaasis hoiustamiseks ning kiireks protsessimiseks juba erilahendusi. Lisaks lõpptulemuse saavutamise nõuab ka eelnevalt, üksikparameetrite protsentuaalse sobivuse välja arvutamise juures, süsteemilt väga suurt ressursi. Viimane ja kõige suurem probleem on see, et parameetrite näol on tegemist dünaamiliste muutujatega, mis tervikuna ajas väga kiiresti muutuvad. Ka kõige väiksem muudatus ühes tööpakkumises, CV-s või töösoovis tähendab sobivustabeli andmete aegumist. Et andmed oleks ajakohased, tähendaks iga muudatus terve sobivustabeli uuesti välja arvutamist. Kuna see aga nõuab kõik liialt palju ressursi - nii ajalist, andmemahulist, arvutus-võimsuslikku - siis autor selle variandi välistab.

Sobivusprotsendi ette arvutamine on mõistlik säärase sisendandmete puhul, mille väärtused muutuvad ajas harva, mille tulemusi kasutatakse tihti ning mille andmemahud pole lõpptulemusena ebamõistlikult suured. Autori hinnangul vastab antud töö raames kirjeldatud kriteeriumitele ametinimetuste klassifikaatorid. Ametinimetuste klassifikaatorid muutuvad ajas harva. Uues lahenduses on tulemused väga olulised ja tihedalt kasutuses ning tulemused andmemahu poolest mõistlikud. Erinevaid ametinimetusi on hetkel süsteemis ligi 1 000, millest tulenevalt kõikide ametinimetuste omavaheliste sobivusprotsentide hulk oleks

$1000 \cdot 1000 = 1\,000\,000$. Tulemust analüüsid leiab autor, et seda on võimalik veelgi optimeerida. Kuidas optimeerida, selleks süveneb autor konteksti.

Autor leiab, et kahe valdkonna vahelise ametinimetuste sobivusprotsendi arvutamine on otstarbetu. Ühest otsast on kahe valdkonnavahelise ameti sobivusprotsent juba niigi liialt madal. Teisalt kui valemit vaadata, siis kaks eri valdkonnas asuvat klassifikaatorit, mis asuvad vahetult pärast valdkonna tasandit järgmisel tasandil, arvutub nende sobivusprotsendiks sama, mis valdkonna sees oleva kahe sügavamal asuva klassifikaatori sobivusprotsent, mis aga pole ootuspärane.

Sellest tulenevalt otsustab autor kaotada valdkonnavahelised ühendused ning keskenduda sobivusprotsentide välja arvutamisel valdkonnasiseselt. Süsteemis on hetkel 51 ametivaldkonda ja igas valdkonnas keskmiselt 20 ametinimetust. Sellest tulenevat saame uueks tulemuseks $20 \cdot 20 \cdot 51 = 20\,400$ kirjet. Autor hindab seda tulemust heaks ning leiab selle baasilt, et ametinimetuste protsentuaalne sobivus on mõistlik ette arvutada. Autor nimetab seda tulemust ametinimetuste sobivustabeliks. Nüüd, kui piltlikult on vaja välja selgitada kandidaadi soovitud ametinimetuse sobivust tööpakkumise ametinimetusega, siis päring vastu ametinimetuste sobivustabelit annab vastuse üsnagi vahetult. Kui ametinimetuste paar tabelis puudub, tähendab see seda, et ametinimetused on kahest eri valdkonnast ning nende sobivusprotsent on väga madal. Autor leiab, et lõppkasutuses tuleks säärasele vähesobivatele ametinimetustele konfigureerida vaikumisi madal sobivusprotsent. Selleks määratakse 10%.

5.4.2 Parameetrite prioritseerimine

Et sobivate kandidaatide leidmise kiirust optimeerida selgitatakse ideed, kuidas parameetrite prioritseerimise teel vähendada liialt vähe sobivate kandidaatide arvu veelgi. Parameetrite prioritseerimise all mõeldakse parameetrite järjestamist nende olulisuse järgi terviksobivuse määramisel. Prioriteetide määramiseks määrab autor esmalt igale parameetrile kaalud protsentides nii, et kaalude kogu summa oleks 100%. Seejärel järjestatakse need nõnda, et kõige suurema kaaluga parameeter on prioriteet nr. 1, teisena kõige suurema kaaluga prioriteet nr. 2 jne. Autori poolt määratud kaalud ja prioriteedid on toodud välja Tabel 5.

Tabel 5 - Parameetrite kaalud ja prioriteetid

Prioriteet	Kaal/100%	Parameeter
1	20	Standardne ametinimetus/ametivaldkond
2	14	Töö asukoht
3	10	Töökogemus (aastates)
4	8	Kutsetunnistus
5	7	Hariduse valdkond
6	7	Ametitunnistus
7	6	Haridustase
8	6	Keeleoskus
9	6	Juhiload
10	4	Tööaeg
11	4	Töötasu
12	4	Arvutioskus
13	4	Isiklik auto

Prioritiseerimise idee seisneb selles, et arvutada välja sobivusprotsendid esmalt olulistematele parameetritele suuremas kandidaatide hulgas ning seejärel tuginedes saadud tulemustele vähendada kandidaatide hulka järgnevate, vähem prioriteetsemate, parameetrite sobivusprotsendi arvutamisel.

Esmalt arvutatakse välja kõigi kandidaatide soovitud ametinimete protsentuaalne sobivus antud tööpakkumisega. Kui selle koha peal on tulemuste hulk liialt suur, on võimalus kandidaatide hulka vähendada jättes kõrvale liiga madala sobivusprotsendiga kandidaadid. Seejärel arvutatakse järele jäänud kandidaatide hulgale sobivusprotsendid, prioriteet nr. 2, tööasukoha mõistes, kasutades eelpool kirjeldatud arvutusmeetodit.

Kui kahele esimesele, prioriteetsemale parameetrile, on kandidaatide sobivusprotsendid arvutatud, leiab autor, et on mõistlik kitsendada kandidaatide arv ühele kindlale konfigureeritavale konstantsele arvule, võttes aluseks leitud kahe parameetri sobivusprotsentide terviksobivuse. Terviksobivuse leidmiseks kasutame valemit, mis on ka eelnevalt ära kirjeldatud:

$$S = \sum k_i P_i$$

kus tegurid P_1, P_2, \dots, P_n tuleb asendada, praeguseks hetkeks leitud sobivusprotsentidega ning kus k_i suhtes kehtib seos

$$\sum k_i = 1$$

Näiteks kui tööpakkumise ametinimetuse ning soovitud ametinimetuse sobivusprotsent on 75% ning töö asukohta ja soovitud töö asukohta sobivusprotsent on 90% siis

$$P_1 = 75\%$$

$$P_2 = 90\%$$

k_i – määrame tulenevalt juba määratud kaaludest nõnda, et kahe parameetri kaalude suhe jääks samaks. Arvutuskäik on järgmine:

$$k_1 = \frac{20}{20+14} \approx 0,58$$

$$k_2 = \frac{14}{20+14} \approx 0,42$$

Asendades tegurid terviksobivuse leidmise valemis saame näidistulemuseks: $S = 75 * 0,58 + 90 * 0,42 = 81,3\%$.

Kirjeldatud võtteid kasutades saame moodustada kandidaatidest pingerea, mis moodustatud kahe olulisema parameetri terviksobivuse alusel. Kui pingerida on moodustatud, valitakse sealt esimesed 100 (konstantne konfigureeritav arv), millest saab piiratud kandidaatide grupp. Seejärel arvutatakse valitud kandidaatide grupi piires ka kõigi teiste parameetrite sobivusprotsendid. Kui needki leitud, rakendatakse terviksobivuse leidmise valemit teist korda - nüüd juba sisendiks kõigi parameetrite sobivusprotsendid. Saadud tulemuste põhjal moodustatakse taas kandidaatidest pingerida, mis on ühtlasi ka autori lõppeesmärk ehk sellega on leitud sobivate kandidaatide järjestatud nimekiri nõutud tööpakkumisele.

Kokkuvõte

Käesoleva töö algseks eesmärgiks oli analüüsida praegu kasutusel oleva Eesti Töötukassa Iseteenindusportaali tarvis välja töötatud soovitusüsteemi lahendust ning selle puudujääke. Teiseks töö eesmärgiks oli analüüsida 2003. aastal koostatud Tööturuameti Selveinfosüsteemis kasutusel olnud töökuulutuse, töösoovide ja CV-de sobitamise arvutusmootorit, millest lähtuvalt sooviti leida ning anda ülevaate, milliseid meetodeid ning kuidas oleks võimalik Selveinfosüsteemi jaoks välja töötatud lahendustest ära kasutada Töötukassa Iseteenindusportaali tarbeks. Kolmandaks ja ühtlasi töö põhieesmärgiks oli pakkuda välja autoripoolne lahendus uue ning täiuslikuma soovitusüsteemi väljaarendamiseks, mis peaks tagama Töötukassale, tööpakkujale kui ka töötajale võidu nii ajalises, finantsilises, kui ka emotsionaalses mõttes.

Töö kõik algselt püstitatud eesmärgid said autori hinnangul täidetud, mis kajastub ka järgnevas.

Töö olulisemate tulemuste näol saab tõdeda, et praegune kasutusel olev soovitusüsteem ei ole tõesti parim lahendus tööpakkumise ning sobivate kandidaatide kokku viimisel, kuna ta ei oma sobivusotsingute teostamisel konfigureeritavaid parameetreid, teda on keeruline täiustada, ta ei võta arvesse suurt osa kandidaadi poolt sisestatud ning samas struktureeritud andmetest ning jääb silma ka kergekäelise, samas reaalses elus potentsiaalselt üsnagi sobivate kandidaatide välistamisega.

Lisaks võib olulise tulemusena ära märkida ka autoripoolse detailselt kirjeldatud uue ning täiuslikuma soovitusüsteemi algoritmi ning kasutatava lähteandmestiku kirjelduse, mis baseerub osalt Selveinfosüsteemis kasutusel olnud töökuulutuste, töösoovide ja CV-de sobitamise arvutusmootoril.

Uue soovitusüsteemi kirjeldamise käigus oleva andmemahtude analüüsi osas selgus aga ka tõsiasi, et ette arvatava sobivusprotsentide näol kasutatav arvutus- ning andmemahtude osa on lahenduse näol üsnagi suur ja ressursimahukas, mille raames tuli teha ka algselt välja mõeldud algoritmis teatud korrekture ning välistamisi. Näiteks selgus, et tööpakkumise ja kandidaadi sobivusprotsendi ettearvutamine on mõistlik vaid ajas harva muutuvate parameetrite korral, mille valikud autor töös ka välja tõi. Teisalt mõeldi andmemahtude ja

arvutusvajaduse optimeerimiseks välja ka võrdlusparameetrite kaalude ja prioritseerimise idee, mida on võimalik rakendada, katkestamaks päringu edasine töö, kui teatud arv sobilikke kandidaate on juba leitud. Viimane välistab ülemäärase tulemuste hulga arvutamiseks muidu tühja kulutatava ressursi.

Antud tööd on võimalik kindlasti edasi arendada ja täiustada. Näiteks tuli Iseteenindusportaali soovitusüsteemi tarvis valitavate lähteandmete selekteerimise käigus välja palju töövahenduse seisukohalt tähtsaid, ent kasutaja sisendi seisukohalt struktureerimata andmehälju, mida efektiivsemaks kandidaatide ja tööpakkumiste kokkusobitamiseks käesoleva töö raames kirjeldatud algoritmis kasutusele võiks võtta. Selleks tuleks esmalt Iseteenindusportaalis lubada vastavate vormide näol sisendiks anda vaid struktureeritud andmeid. Lisaks on autor asunud lahendust ka reaalselt implementeerima, mille käigus on reaalne, et ilmnevad asjaolud, mida ei osatud soovitusüsteemi kirjeldamisel algselt arvestada või ette näha. Arendusprotsess võib aga algoritmi potentsiaalsete paranduste korral omakorda veelgi täiuslikumaks muuta. Lisaks saaks antud töö tulemust lõplikult hinnata alles pärast reaalse kirjelduse implementatsiooni täielikku valmimist ja kasutuselevõtmist ning sellele järgneva Töötukassa poolt vahendatava töövahenduse kohta käiva statistilise andmestiku kui ka tööpakkujate ja töötajate rahuloluuringute põhjal.

Summary

The primary goal of current thesis was to analyze the shortcomings of the current recommendation system in use for the Estonian Unemployment Insurance Fund's self-service portal.

The second goal was to analyze a matching calculation engine that was based on adjusting job offers, job requirements and CV-s and put together in 2003 by Tööturuameti Selveinfosüsteem. The purpose of this analysis was to identify which solutions worked for the Selveinfo system and how these solutions could be applied to improve the Estonian Unemployment Insurance Fund's self-service portal recommendation system that is currently in use.

The third goal, and also the major goal of this thesis, was to give the authors ideas and concepts for new and improved recommendation system, which would assure a better and more convenient experience (both financially and emotionally) that also saves time.

The author of this thesis finds that all the goals mentioned above were fulfilled during the thesis course.

Based on most significant results found out during the thesis, it can be said that the recommendation system that is currently used to bring together the most eligible candidates with available job offers is not the best. This is because it is difficult to expand its logic, it does not have configurable parameters when performing searches for matching, it does not take some of the structured data, entered by the unemployed people, into consideration. This can potentially exclude eligible candidates for the job offers. Additionally, a detailed description of a new and improved recommendation system algorithm and its input data has been successfully described.

While describing this new recommendation system, it was discovered that the pre-calculation and data volume portion, that is going to be used for calculating the percentage of compatibility, will be too big and takes too much resources for the primal algorithm. The result was that corrections and some exclusions were required to improve the algorithm. For example, it was discovered that pre-calculating the match between job offer and eligible

candidates is only reasonable for parameters that change rarely during time. These rarely changing parameters were also pointed out as an input data for the algorithm.

The author also identified examples for optimizing data volume and need for calculation. For that, comparison parameters scales and prioritization idea was thought out. This identification should be able to be applicable for recommending discontinuing status request, if a certain amount of candidates will be found. The last method mentioned should decrease the number of calculated results that are unnecessary and point out that when resources are used for making excess calculations.

There is definitely a possibility to expand and improve current thesis. For example, while selecting initial data, the author discovered many significant data fields in self-service portal, that are important for intermediating jobs but are not structured in terms of user input model. These fields should first be standardized in self-service portal and then also included into algorithm described during current thesis course.

Also, the author has started to implement his recommendation system solution, described by this thesis, but is aware that during the process, circumstances that were unable to see or take into consideration beforehand, while describing the recommendation system, can arise. But these potential circumstances can help to improve the algorithm even more. However, goodness of the algorithm, developed in current thesis, can be finally evaluated once the implementation described is finished in program level and also taken into active use. After that, statistical data about labor mediation and satisfaction research can be respectively gathered and carried out by the Estonian Unemployment Insurance Fund. Previously mentioned data would eventually reflect the success of the application.

Kasutatud kirjandus

1. ILO – ISCO. [WWW] <http://www.ilo.org/public/english/bureau/stat/isco> (12.02.2015)
2. ITP artikkel infosüsteemidest. [WWW] <https://www.tootukassa.ee/uudised/kadri-luhiste-infosusteemid-peavad-tootama-tootukassa-ja-tootute-heaks> (06.02.2015)
3. Joseph Konstan and Michael Ekstrand – Intro. Recommender Systems. [WWW] <https://www.coursera.org/learn/recommender-systems> (03.02.2015)
4. Nortal – EMPIS. [WWW] <http://www.nortal.ee/meist/uudised/nortal-v%C3%B5itis-t%C3%B6tukassa-tarkvara-arendust%C3%B6de-hanke> (14.03.2015)
5. Prem Melville and Vikas Sindhwani - Recommender Systems. [WWW] <http://vikas.sindhwani.org/recommender.pdf> (23.03.2015)
6. Selveinfosüsteem. [WWW] <http://mwx.nlib.ee/et/download/23667> (12.03.2015)
7. Sotsiaalministri määruste muutmine. – RTL 2009, 36, 481
8. Tööturuameti põhimäärus. – RTL 2005, 121, 1894
9. Tööturuteenuste ja –toetuste seadus. – RT I, 04.06.2014, 8
10. Wikipedia - Dijkstra algoritm. [WWW] http://et.wikipedia.org/wiki/Dijkstra_algorithm (12.04.2015)