

TALLINNA TEHNIKAÜLIKOOL

Infotehnoloogia teaduskond

Informaatikainstituut

Infosüsteemide õppetool

Aljona Trofimova 123979IABB

ARVETE ANDMETE TUVASTAMISE PROTSESSIDE OPTIMEERIMINE

Bakalaurusetöö

Juhendaja: Enn Õunapuu
dotsent

Tallinn 2017

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.

Autor: Aljona Trofimova

22.05.2017

Annotatsioon

Käesolevas töös analüüsitakse paberivaba arvehalduse lahenduse arvete tuvastamise alamosa kvaliteeti. Vaatluse all olevateks protsessideks on arvete import, arvete andmete tuvastamine, valideerimine ja eksport. Töö eesmärgiks on analüüsida arvete andmete tuvastamise protsesside puudused ja võimalused arenduse efektiivsuse optimeerimiseks.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 29 leheküljel, 6 peatükki, 15 joonist, 4 tabelit.

Abstract

Optimization of Invoice Data Recognition Processes

In this bachelor work author is analyzing the quality of invoice data recognition. The main processes are import of invoices, invoice data recognition, recognized invoice data validation and regognized invoice data export. The main purpose of analysis is to find the weaknesses and opportunities of the processes and to find solutions to optimize them.

The thesis is in Estonian and contains 29 pages of text, 6 chapters, 15 figures, 4 tables.

Lühendite ja mõistete sõnastik

MSDAX	Microsoft Dynamics AX 2009
FC	ABBY FlexiCapture 11.0
PHL	Paberivaba arvehalduse lahendus
AAT	Arvete andmete tuvastamine
AK	Arvete kinnitamine
FCL	FlexiCapture Layout
FCL block	FlexiCapture Layout plokk
FCL element	FlexiCapture Layout element
FCL group	FlexiCapture Layout elementide grupp
FCP	FlexiCapture Project

MSDAX on ERP tarkvara ettevõtte andmete hoidmiseks, haldamiseks, töötlemiseks ja anlüüsimiseks.

FC on mitmest komponentidest koosnev tarkvara, mille eesmärk on andmete optiline tuvastamine.

PHL on FC ja MSDAX abil loodud lahendus, mille eesmärk on arvete andmete tuvastamine, importimine MSDAX'sse ja kinnitamine. PHL koosneb kahest loogiliselt omavahel seotud osast - AAT'st ja AK'st.

AAT on PHL'i alamosa arvete andmete tuvastamiseks, loodud FC komponentide abil. AK on PHL'i alamosa arvete andmete MSDAX'sse importimiseks ja kinnitamiseks.

FCL on FC komponent, kus defineeritakse (programmeeritakse ja seadistatakse) arvete tuvastamise loogikat, mille interpreteerivad teised FC komponendid AAT'l andmete välja võtmiseks. FCL'i programmeerib arendaja enam levinud arvete põhjal.

FCL block on FCL objekt, mis representeerib arve rekvisiidi, mis peab olema tuvastatud. Üks FCL block võib koosneda mitu FCL elementidest, mis võivad omavahel seotud olla.

FCL element on FCL objekt, kuhu sisse kirjutakse parameetreid iga otsitava arvete rekvisiidi leidmiseks.

FCL group on FCL'is defineeritud grupp, mis koosneb omavahel seotud FCL elementidest.

FCP on FCL loogikat interpreteeriv FC komponent, kus kasutaja käivitab arvete tuvastamise protsessi, vaatleb arvete tuvastamise tulemusi, vajadusel redigeerib tuvastatud andmeid ja käivitab andmete eksport.

AAT efektiivsus on keskmine iga arve kohta õigesti tuvastatud arvete andmete protsent ja andmete tuvastamise kiirus.

Koondarve on mitu arvetest koosnev ühine arve, millel on üldine päis.

Faktuurarve on transporditavate kaupade ja nende hindade üksikasjalik loetelu, mis on levinud Venemaal ja Europa Liidu riikides.

Kontsernisisene arve ühelt tütarettevõttelt teisele tütarettevõttele esitatud arve.

Sisukord

Sissejuhatus	10
1 Ülevaade	11
1.1 Tellija iseloomustus	11
1.2 Eelanalüüs ja olukord hetkel.....	11
1.2.1 Arvete töötlemise protsess.....	11
1.2.2 Arvete töötlemise protsessi etapid.....	11
1.2.3 PHL'i struktuur.....	12
1.2.4 PHL'i alamosade detailne kirjeldus.....	12
1.2.5 AAT põhilised protsessid	13
1.2.5.1 Arvete import FCP'sse	13
1.2.5.2 Arvete tuvastus	13
1.2.5.3 Tuvastatud andmete kontroll ja valideerimine	13
1.2.5.4 Arvete eksport.....	14
1.3 Töö lühikirjeldus.....	16
1.3.1 Sisend.....	16
1.3.2 Konkretiseeritud töö eesmärgid.....	16
1.3.3 Arenduse analüüsi struktuur	16
1.3.4 Metoodika.....	16
1.3.5 Oodatav tulemus	17
2 AAT arenduse protsesside analüüs enne optimeerimist	18
2.1 Arvete tuvastusamise ja arvete tuvastamise kvaliteet.....	18
2.1.1 Tellija poolt saadetud tagasiside.....	18
2.1.2 Tellija poolt koostatud arvete komplektide ülevaade	18
2.1.3 Tellija poolt koostatud arvete komplektide testimine.....	19
2.1.4 AAT ehituse põhiprintsiibid	19
2.1.5 AAT ehitamisel võimalikud keerukused	22
2.1.6 Võimalikud lahendused suure arvete disainide varieeruvuse puhul...	23
2.1.6.1 Madal disainide varieeruvus	23

2.1.6.2 Suur disainide varieeruvus.....	23
2.1.7 Esmase arvete komplektide analüüs	24
2.1.7.1 Arve number	24
2.1.7.2 Arve summa ja käibemaks.....	24
2.1.7.3 Arve kuupäev ja tähtaeg	24
2.1.7.4 Viitenumber	25
2.1.7.5 Registrikood.....	26
2.1.7.6 Hankija.....	26
2.1.7.7 Arve saaja	26
2.1.7.8 Esialgse arvete rekvisiide ja FCL'i analüüsi tulemused...	26
2.1.7.9 Uude komplekti tuvastamisega seotud probleemid	26
2.2 Arvete tuvastamine ja dünaamiline otsisõnade täiendamine	27
2.2.1 Uue otsisõnade lisamine	27
2.2.2 Mõõdikud.....	27
2.3 Arvete andmete valideerimine	28
2.3.1 Andmete formaadi kontroll	28
2.3.2 Tuvastatud registrikoodi olemasolu.....	28
2.3.3 Mõõdikud.....	28
2.4 Arvete andmete eksport	29
2.4.1 Mõõdikud.....	29
2.5 Tellija nõudmised	29
3 AAT protsesside optimeerimine	30
3.1 Arvete tuvastamine	30
3.1.1 Võimalikud FC ehitusstruktuurid	30
3.1.1.1 FCL'i laiendamine	30
3.1.1.2 Eraldi FCP'd ja FCL'id (1:1).....	30
3.1.1.3 FCL'ide hieraaria põhjaliku FCL'i sees	31
3.1.1.4 Lahenduse valimine	31
3.1.1.5 Mõõdikud.....	31
3.2 Arve otsisõnade kogumite dünaamiiline täiendamise puudus.....	31
3.2.1 Võimalused.....	32
3.2.2 Valitud lahendus	32
3.3 Tuvastatud rekvisiitide valideerimise puudus	32

3.3.1 Arve number	32
3.3.2 Summa ja käibemaks	32
3.3.3 Arve kuupäev ja tähtaeg	32
3.3.4 Viitenumber	32
3.3.5 Hankija.....	33
3.3.6 Registrikood.....	33
3.3.7 Mõõdikud.....	33
3.4 Eksportimine valedesse kaustadesse	33
3.4.1 Mõõdikud.....	33
4 Optimeeritud AAT arenduse ülevaade	34
4.1 Kirjeldus	34
5 AAT arenduse ülevaade.....	35
5.1 Võimalikud alternatiivsed lahendused.....	35
5.2 AAT eelised	35
5.3 AAT puudused.....	35
5.4 FC eelised	35
5.5 FC puudused	35
5.5.1 Muudatused	35
5.5.2 FC stabiilsus	36
5.5.3 FC ja arvete varieeruvus	36
6 Kokkuvõte	39
Kasutatud kirjandus	40

Jooniste loetelu

Joonis 1. Arve töötlemise protsessi struktuur.....	14
Joonis 2. PHL'i üldine struktuur	15
Joonis 3. Kaustade struktuuri näide	16
Joonis 4. Kausta ja CSV faili näide	17
Joonis 5. Eksporditavate failide liikumine	21
Joonis 6. FCL ja FCP komponentide põhifunktsioonid	21
Joonis 7. FCP ja FCL'i seos	21
Joonis 8. Arve rekvisiidid	22
Joonis 9. FCL'i struktuur	22
Joonis 10. FCL'i arendamiskeskonna vaade	23
Joonis 11. FCP keskkonna vaade	23
Joonis 12. FCP keskkonna vaade	24
Joonis 13. Keerukama FCL struktuuri näide.....	24
Joonis 14. AAT ehitus, kus on FCP'd ja FCL'id jagatud mitmeks	25
Joonis 15. AAT näide, kus on kasutatud FCL'ide hieraaria	26

Tabelite loetelu

Tabel 1. Arvete tuvastamise SWOT-tabel.....	29
Tabel 2. Arvete valideerimise SWOT-tabel	30
Tabel 3. Arvete andmete eksportmise SWOT-tabel.....	30
Tabel 4. Mõõdikute võrdlus	31

Sissejuhatus

Autor valis bakalaureusetöö teemaks “Arvete andmete tuvastamise protsesside optimeerimine”. Selle valiku tingis töö autori osalemine majandustarkvara arendajana AAT optimeerimise projektis. Nimetatud arendus oli olnud kasutusel aasta jooksul ning selles olid ilmnunud puudujäägid, mille alusel tellija esitas optimeerimise tellimuse. Lõputöö kirjeldamisele eelnenud aasta jooksul oli selle tarkvara täiustamise protsess töö autori põhitööks. Saadud kogemuse alusel kujunes autoril ülevaade sellest, kuidas toimub arvete töötlemine ja millised probleemid tekivad arendamise ja testimise käigus ning kuidas on võimalik arvete andmete töötlemise protsessi täiustada.

Antud töö eesmärk on välja selgitada, milline potentsiaalne lahendus on kõige mõistlikum ja rakendada seda kasutusoleva AAT optimeerimiseks. Töö tulemusel peaks selguma, kuidas oleks võimalik AAT arendus optimeerida, millist kasu tellija jaoks see toob ja millised AAT’l arvete töötlemise vahendina tugevad ja nõrgad külged on.

Antud töös vaadeldava AAT eesmärk on säästa ettevõtte olulisemaid ressursse - raha ja aega. Selleks tuleb lihtsustada ja vähendada raamatupidaja tööd arvete töötlemisel. AAT peab efektiivselt, täielikult ja kiiresti verifitseerima vajalikud andmed.

1 Ülevaade

Käesolevas alampeatükis annab autor ülevaate tellija firmast, tutvustab analüüsitava AAT arenduse struktuuri enne optimeerimist ja selle optimeerimise ülesande püstitust. Lisaks räägib autor tellija arvete töötlemise protsessist ja selle jaoks loodud tervikliku lahendusest, millele AAT osa kuulub.

1.1 Tellija iseloomustus

Tellijal on suurettevõtte, mis eksisteerib kakskümmend neli aastat ja kasutab kaua aega oma igapäevases töös MSDAX. Kontsernis on 20 tüttarettevõtet, kelle nimele iga päev laekuvad arved kas paberandjal või PDF-formaadis. Kuna kuu jooksul keskmine sissetulevate arvete arv on 800 ja rohkem, nende sisestamine MSDAX'sse muutus pikkaajaliseks ja töömahukaks, ning nõuaks spetsiifilist lahendust. Aastal 2015 oli arvete töötlemise protsessi lihtsustamiseks juba kasutusel AAT'st ja AK'st koosnev PHL, mis oli loodud firma poolt, kus käesoleva lõputöö autor töötab.

1.2 Eelanalüüs ja olukord hetkel

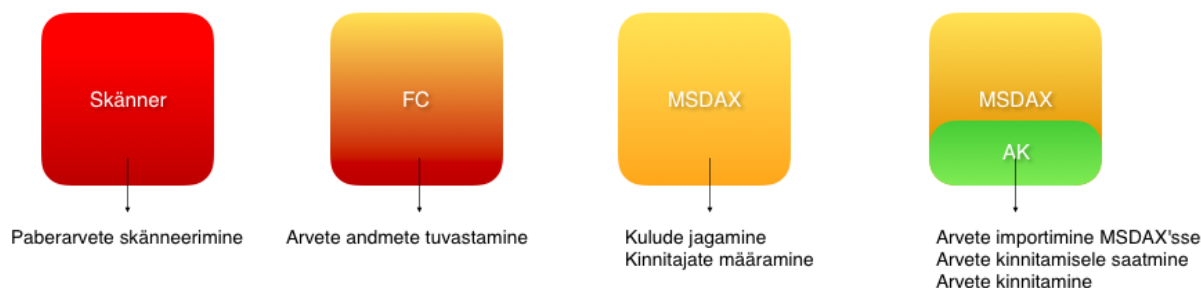
Tellijal on suurettevõtte, mis eksisteerib kakskümmend neli aastat ja kasutab kaua aega oma igapäevases töös MSDAX. Kontsernis on 20 tüttarettevõtet, kelle nimele iga päev laekuvad arved kas paberandjal või PDF-formaadis. Kuna kuu jooksul keskmine sissetulevate arvete arv on 800 ja rohkem, nende sisestamine MSDAX'sse muutus pikkaajaliseks ja töömahukaks, ning nõuaks spetsiifilist lahendust. Aastal 2015 oli arvete töötlemise protsessi lihtsustamiseks juba kasutusel AAT'st ja AK'st koosnev PHL, mis oli loodud firma poolt, kus käesoleva lõputöö autor töötab.

1.2.1 Arvete töötlemise protsess

Arvete töötlemise protsess antud töös analüüsitud ettevõttes koosneb järgmistest etappidest: paberarvete skanneerimine, arvete andmete tuvastamine, andmete sisestamine süsteemi, kulude jagamine, kinnitajate määramine, arvete kinnitamisele saatmine, arvete kinnitamine.

1.2.2 Arve töötlemise protsessi etapid

Järgmine skeem (joonis 1) demonstreerib arvete töötlemise etappe.



Joonis 1. Arve töötlemise protsessi struktuur

1. Skänneri abil tekib arve PDF'i kujul;
2. FC abil toimub PDF'i kujul arvete andmete tuvastamine ja muutmine formaadiks, mille suudab AK moodul vastu võtta;
3. Majandustarkvaras MSDAX jagatakse kulud ja määratakse arvete kinnitajaid;
4. MSDAX AK mooduli abil imporditakse FC's tuvastatud andmeid, arveid saadetakse kinnitamisele ja kinnitatakse;

1.2.3 PHL'i struktuur

Tellijajaoks loodud PHL koosneb kahest loogiliselt omavahel seotud alamosast - AAT ja AK. Esimese osa funktsioon on andmete tuvastamine ja kohandamine selliselt, et tellijal oleks võimalik neid oma majandustarkvara programmi importida. See osa on arendatud FC abil. Teise osa funktsioon on arvete andmete importimine ja kinnitamine. AK on erimoodul tehtud MSDAX majandustarkvara juurde.

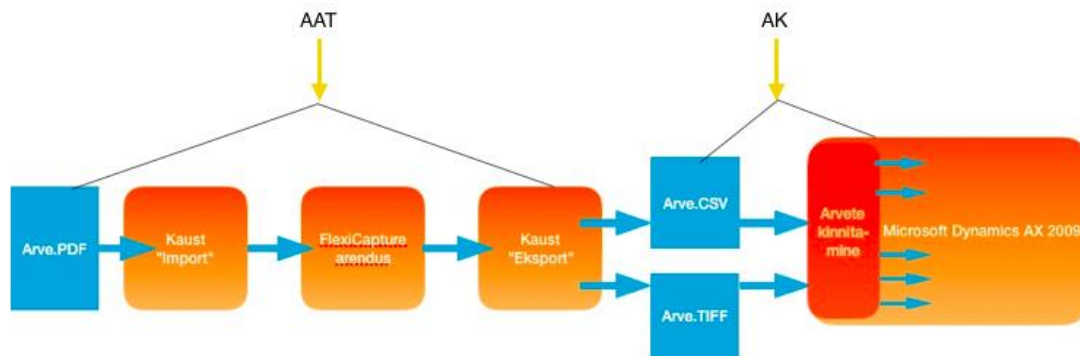
1.2.4 PHL'i alamosade detailne kirjeldus

AAT alamosa kirjeldus:

Kasutaja skännerib paberikandjalt arvet või kasutab olemasoleva PDF-formaadis ja paneb arendaja poolt määratud FCP seadistustes kausta. FCP kontrollib selle kausta iga kindla ajaühiku tagant ja impordib saabunud arved. Tuvastamise protsessi lõpetades kontrollib kasutaja tuvastamisõigust ja ekspordib andmed. Andmed viiakse välja kindlas formaadis teise kokku lepitud kausta, kust seejärel MSDAX suudab importimisel need andmed vastu võtta.

AK alamosa kirjeldus:

Kasutaja avab AK mooduli MSDAX'is ja impordib FCP'st eksporditud failid sisse. Pärast importimist toimub arvete kinnitamine ja finantskannete tekitamine. Antud pildil on PHL'i üldine skeem (joonis 2).



Joonis 2. PHL'i üldine struktuur

1.2.5 AAT põhilised protsessid

- Arvete import FCP'sse;
- Arvete tuvastus;
- Tuvastatud andmete kontroll;
- Andmete eksport;

Järgmisena kirjeldatakse arvete andmete tuvastamise protsesse enne optimeerimist ja vaadeldakse nende puuduseid.

1.2.5.1 Arvete import FCP'sse

Kasutaja paneb PDF-kujul arveid määratud kausta. FCP automaatselt võtab arveid vastu iga 20 sekundi tagant.

1.2.5.2 Arvete tuvastus

FCP interpreteerib FCL'i ning tuvastab arvete andmeid ja näitab tuvastatud andmeid.

Protsessi puudused enne optimeerimist:

1. Madal tuvastamise kiirus ja madal õigesti tuvastatud andmete protsent. Arveid, mis olid enne PHL'i loomist tellija poolt saadetud näidiste hulgas, tuvastatakse 4 sekundiga ja nende keskmine õigesti tuvastatud andmete protsent on 85%. FCL'i jaoks tundmatu disainiga arveid tuvastab FC 10 sekundiga ära ja keskmine õigesti tuvastatud andmete protsent on 46,6%.
2. Erinevate disainide puhul on ebapaindlik. Iga uue otsisõna lisamine programmi peab toimuma arendaja poolt. Faktuurarveid, koondarveid, kontserniseseid arveid ja arveid mitte eesti või inglise keeles FC enne optimeerimist ei ole võimeline tuvastada.

1.2.5.3 Tuvastatud andmete kontroll ja valideerimine

Kui FCP näitab, et imporditud arve on tuvastatud, saab kasutaja vaadata tuvastatud tulemusi üle ning vajadusel redigeerida või puudumise juhul sisestada.

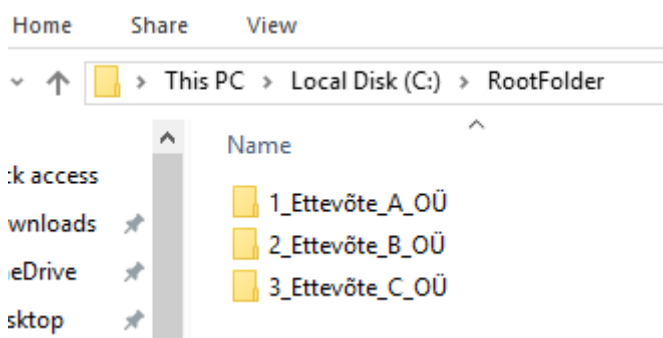
Protsessi puudused enne optimeerimist:

1. Juhul, kui FCL ei oska mõnda teatud arve disaini õigesti tuvastada peab kasutaja seda käsitsi redigeerima või sisestama. Puudub seadistatud andmete esialgne automaatne valideerimine, mis tõstab inimfaktori riski tõenäosust.

1.2.5.4 Arvete eksport

Kui andmed on kontrollitud, käivitab kasutaja andmete ekspordi. Arve andmed lähevad välja CSV-formaadis ja faili nimi sisaldab alamkriipsudega eraldatud arve numbrit, hankija nime ja kuupäeva.

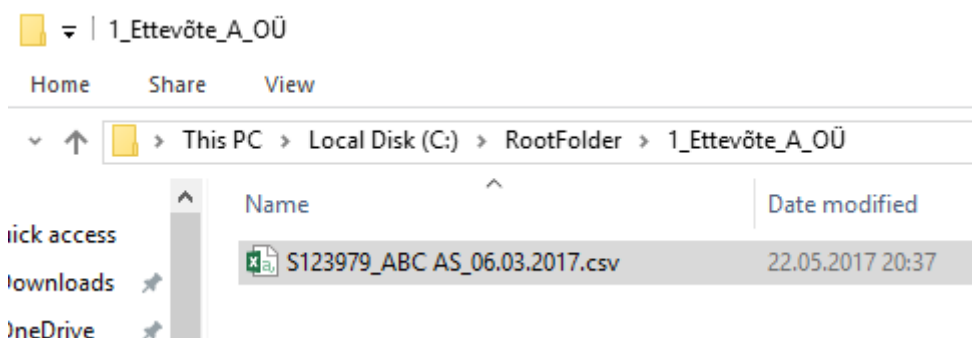
Andmetega failide jaoks tellija lõi spetsiaalse kaustade struktuuri, kus iga tüarettevõtte jaoks on loodud personaalne kaust, mille nimi koosneb unikaalsest numbrist, alamkriipsust ja lühendatud tüarettevõtte ärinimest.



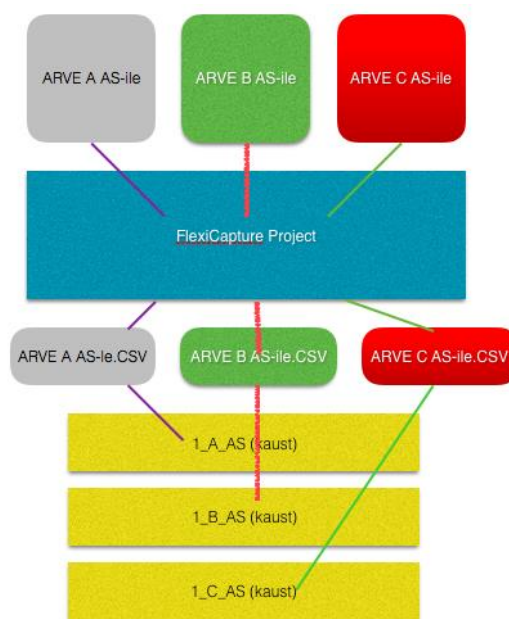
Joonis 3. Kaustade struktuuri näide

Enne faili ekspordimist otsustab FC tuvastatud arve saaja järgi millisse kausta tuvastatud andmetega fail peab minema. FCP programmkoodis on sissekirjutatud algoritm, mis asendab võimalikud tuvastatud arve saaja väärtuseid kaustade väärtustele.

Eksporditava andmete jaoks direktoriumi valiku algoritm on teostatud eksporditava faili nime moodustamisel. Näiteks nii: c:\\RootFolder\\ + „arve saaja väärtus asendatud kausta nime väärtusega“ + ”arve number_hankija nimi_kuupäev”+”.csv”.



Joonis 4. Kausta ja CSV faili näide



Joonis 5. Eksporditavate failide liikumine

Protsessi puudused enne optimeerimist:

1. Kui minimaalselt üks rekvisiit, mida kasutakse eksporditava CSV faili nime koostamiseks, puudub või on tuvastatud ebakorrektselt, toob see negatiivseid tagajärgesid. Vale nimi võib segada kasutajat faili otsimisel või muuta CSV faili kasutamiseks kõlbmatuks. FC versioon enne optimeerimist lubab ekspordi igal juhul, kontrollimata kas on kõik nime moodustamiseks vajalikud väljad on tuvastatud või mitte.
2. Juhul, kui FC valesti tuvastas arve saajat, tekivad uued direktoriumid, kuhu lähevad arvete andmed. FC lubab CSV ekspordi juhul, kui arve saaja väärtus ei ole võrdne ühe olemasoleva kaustadest nimega.

Kõik toodud puudused nõuavad arenduse uurimist, ümbertegemist, täiendamist, kontrolli lisamist.

1.3 Töö lühikirjeldus

1.3.1 Sisend

Töö sisendiks on soov analüüsi kaudu välja selgitada kuidas optimeerida AAT selleks, et vähendada ettevõtte aja- ja rahakulu.

1.3.2 Konkretiseeritud töö eesmärgid

Töö eesmärgiks on AAT analüüsimise tulemusel tuvastada kasutusoleva arenduse puudused ja kitsakohad ning selgitada välja nende potentsiaalsed lahendused ja valida parimat.

1.3.3 Arenduse analüüsi struktuur

AAT arenduse analüüsi esimeses osas vaadeldakse AAT arenduse kohta tagasisidet ning tõstakse esile selle olemasolevad puudused. Arenduse ehitusstruktuuri kirjeldakse sellisel kujul nagu see eksisteeris enne optimeerimist.

AAT analüüsi teises osas kirjeldatakse AAT jaoks uued võimalused kasutusoleva arenduse optimeerimiseks. Lisaks iga võimaluse analüüsimisel selgitakse välja, kas arenduse efektiivsuse optimeerimine selles ulatuses on põhjendatud ja õigustab ennast.

1.3.4 Metoodika

1. Analüüsida tagasisidet AAT kohta ja puuduseid: autor vaatab üle tellija poolt saadetud tagasisidet.
2. Uurida tellija sagedamini laekuvaid arveid esitatud tellija poolt enne ja pärast optimeerimist: autor põhjalikult vaatab esialgse arvete näidiste komplekte, mis oli saadetud enne PHL'i loomist. See aitab välja selgitada FC ehituse printsiibide eripärasid.
3. Testida esialgse arenduse täpsust ja kiirust mõõtes iga arve puhul õigesti tuvastatud andmete protsenti ja arvutada keskmiselt tuvastamiseks kulunud aega.
4. Uurida tellija poolt esitatud sagedamini laekuvate arvete teise komplekti.
5. Analüüsida kasutusoleva AAT arenduse ehitust: autor selgitab välja mis struktuur FC arendusel on.
6. Välja selgitada AAT olemasoleva puuduste põhjuseid.
7. Koostada olemasoleva arenduse jaoks SWOT-tabel.
8. Analüüsi tulemuste järgi välja tuua võimalused arenduse optimeerimiseks.
9. Välja selgitada parim võimalus optimeerimiseks.
10. Optimeerida käigusolevat versiooni.
11. Testida optimeeritud arenduse versiooni.
12. Mõõta tuvastamisele kulunud aega ja õigesti tuvastatud arvete andmete protsenti uues arenduses.
13. Võrrelda mõõtmise tulemusi enne ja pärast optimeerimist, analüüsida parandatud arenduse versiooni.
14. Välja selgitada, kas AAT osa loomiseks kasutatud tarkvara vastab püstitatud nõuetele.

1.3.5 Oodatav tulemus

Töö tulemuseks on parandatud AAT arenduse kirjeldus, optimeeritud AAT versiooni testimise käigus saadud kvaliteedi mõõdude tulemused. Arenduse optimeerimise abil väheneb andmete tuvastamiseks keskmine kuuluv aeg, tõstab keskmine õigesti tuvastatud andmete protsent, väheneb vajadus kasutaja poolt kontrollimises ja andmete redigeerimises.

2 AAT arenduse protsesside analüüs enne optimeerimist

Käesolevas punktis vaadeldakse AAT protsessid enne optimeerimist, nende puudused ja puuduste põhjuseid.

2.1 Arvete tuvastamise ja arvete tuvastamise kvaliteet

2.1.1 Tellija poolt saadetud tagasiside

Tellija kasutas AAT alamosa enne optimeerimist ning leidis seal olulised puudujäägid:

1. FC ei tuvasta arveid, mille rekvisiidid ei ole eesti ega inglise keeles;
2. FC tuvastab valesti kontsernisisesid arveid;
3. FC tuvastab valesti koondarveid;
4. FC tuvastab valesti faktuurarveid;
5. FC tuvastab valesti notari poolt saadetud arveid;
6. FC tuvastab valesti FIE poolt saadetud arveid;

Üldiselt probleemi võiks sõnastada selliselt, et FC komponentide töö ei ole korrektne ning ei võta vastu mõne määratud disainidega arveid, mis toob kasutajale ebamugavust. Kuna FCL'i arvete tuvastamise loogika defineeritakse konkreetse arvete näidiste põhjal, oletatavasti probleemid on seotud sellega, et tellija poolt esmaselt koostatud tüüpiliste arvete komplektis enim mainitud arvete disainid puuduvad.

Kuna tellija saatis arvete näidiseid, mille kohta tekkisid AAT'l probleemid, töö autoril on võimalus võrrelda AAT efektiivsust kahe arvete näidiste komplektide puhul:

1. Esmane komplekt: arved, mis olid saadetud enne AAT loomist, mille alusel oli arvete rekvisiitide otsimisloogika defineeritud;
2. Teine komplekt: arved, mis olid saadetud pärast AAT testimist;

2.1.2 Tellija poolt koostatud arvete komplektide ülevaade

Esmases vaadeldud komplektis on 46 arvet, mis nende päritolu alusel võib jagada kaheks grupiks: 41 arvet (89%) on eesti keeles ning väljastatud Eestis, ülejäänud 5(11%) inglisekeelsed ja väljastatud välisriikides. Arvete disainid näidistel on peaaegu sarnased.

Teises komplektis on 18 uut arvet, mida FC versioon enne optimeerimist ei saanud korrektselt tuvastada. Nende hulgas on kontsernisisesed arved, faktuurarved, koondarved ja arved läti, inglise, soome ja saksa keeles, milliseid esialgsetes arvete komplektis ei olnud.

30% eksemplaaridel arvete disain on esitatud tabelina, teisel 30% on uued sõnad, ülejäänud 40% käibemaks ja summa korduvad iga osturea peal. Välismaalsete eksemplaaride hulgas arve kuupäev on kirjutatud kasutades kaldkriipse või kriipse.

2.1.3 Tellija poolt koostatud arvete komplektide testimine

Esimese komplekti testimise tulemused olid piisavalt head. Keskmise kiirus iga arve tuvastamise kohta oli 4 sekundit. Õigesti tuvastatud andmete protsent on 95%.

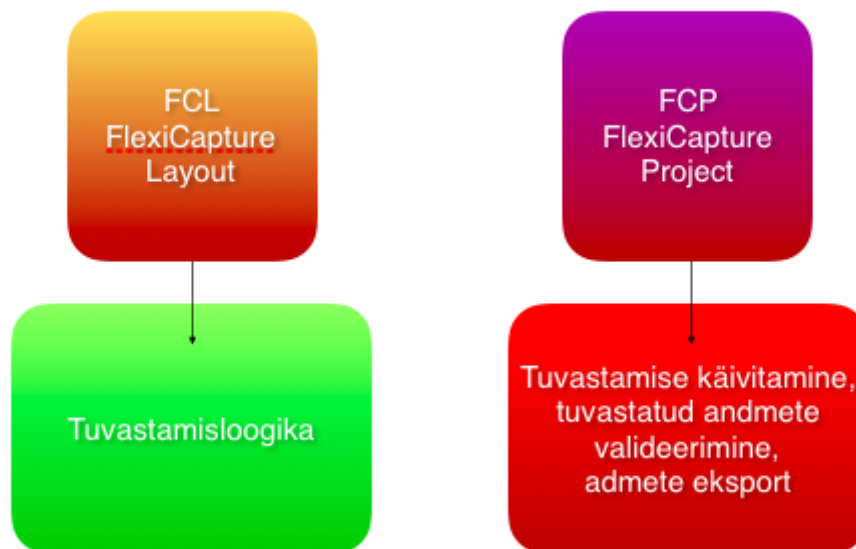
Teise komplekti testimise tulemused olid rahuldavad. Keskmine kiirus iga arve tuvastamise kohta oli 10 sekundit ja keskmine õigesti tuvastatud andmete protsent oli 46,6%. Faktuurarveid ja kontsernisisesed tuvastati 55% täpsusega, uue keelega arved jäid tuvastamata, ehk 0%.

Keskmine vana ja uue arvete tuvastamise protsent on 50%.

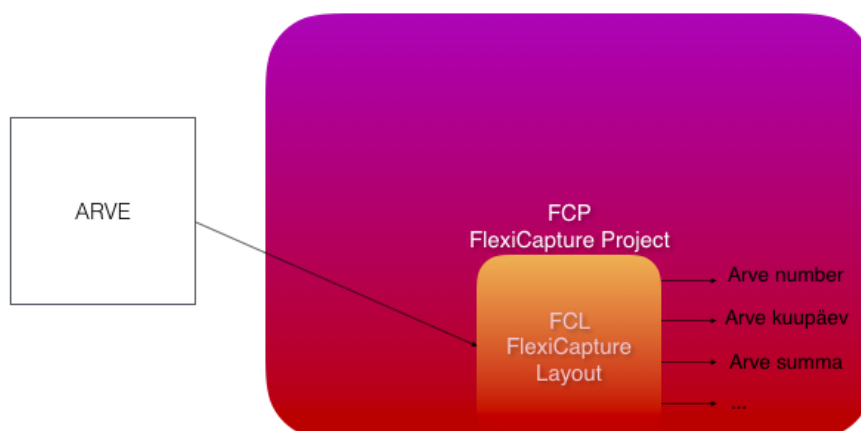
2.1.4 AAT ehituse põhiprintsiibid

Täpsema analüüsi jaoks esmaseks tutvustab autor lühidalt üldiste AAT FC komponentide ehitamise põhiprintsiibidega.

AAT arenduse loomiseks luuakse ja seadistatakse miinimaalselt kaks põhjalikku omavahel seotud osa. Esimene on FCL, mille loob arendaja tellija poolt saadetud tihti laekuvate arvete alusel. FCL sisaldab tuvastavate arvete rekvisiitide tuvastavat loogikat. Järgmine skeem näitab FC komponentide funktsioone.



Joonis 6. FCL ja FCP komponentide põhifunktsioonid



Joonis 7. FCP ja FCL'i seos

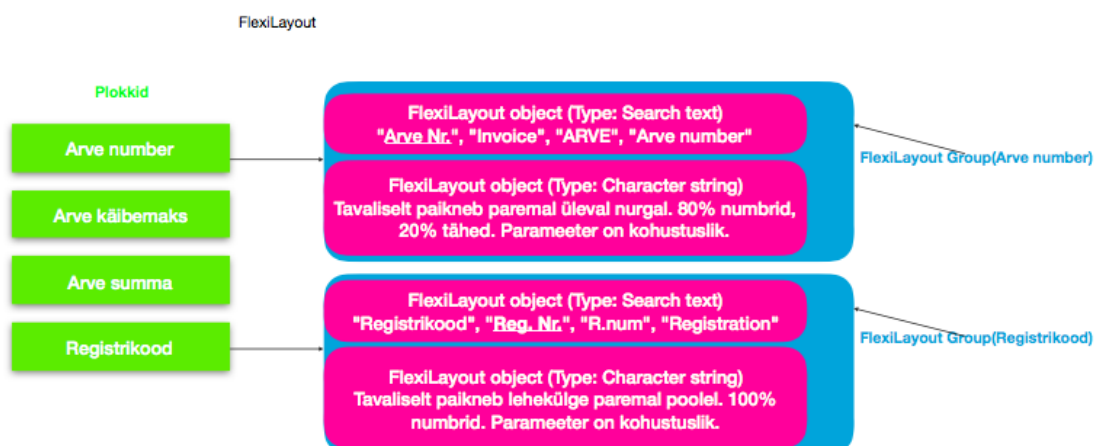
FCL koosneb FCL block'idest, millel on oma FCL group'id, mis omakorda sisaldavad FCL elemente. FCL'i vähim ühik on FCL element, kus ühe arve rekvisiidi jaoks defineeritakse vajalikke andmete tüüpilised paiknemiskohad, otsisõnade kogumid, otsisõnade korrelatsioon otsitavate andmetega. Iga otsitava parameetri jaoks peab olema loodud üks FCL block.

Saaja:	AS TARTU TARTUMAA	Arve Arve Nr. 15103 Kuupäev: 19.05.2015 Maksetähtaeg: 26.05.2015
Reg. Nr.		oü
KMKR Nr.		Reg. Nr. KMKR Nr. EE
PÕHJAMETSA kinnistu; Põhjametsa, Kärsa küla, Ahja vald, Põlvamaa Katastritunnus: 11701:005:0098		Swedbank AS EE

Kirjeldus	Kogus	Hind	Summa (EUR)
Vahendustasu vastavalt pakkumisele 10.05.2015	1 tm	4 000,00	4 000,00
		Summa käibemaksuta 20% (15663348):	4 000,00
		Käibemaks 20%:	800,00
Arve kokku (EUR):			4 800,00

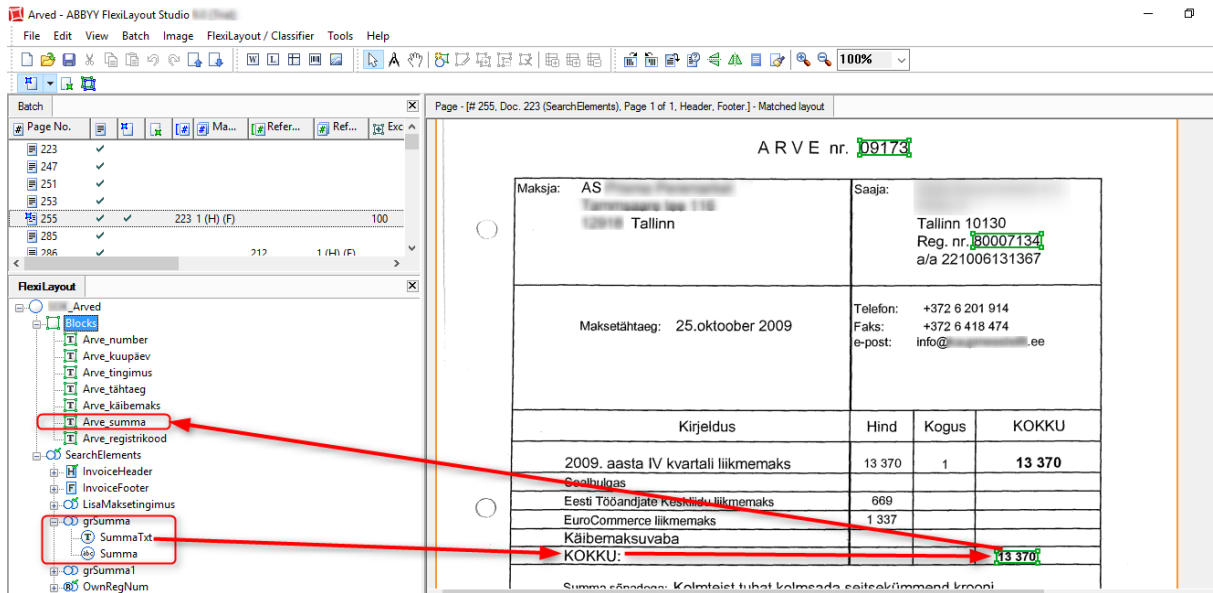
Joonis 8. Arve rekvisiidid

Pildi näide puhul peaks FCL'is looma vastavaid neli FCL block'i. Iga FCL block'i jaoks peab olema loodud vähemalt üks FCL group, mis omakorda tavaliselt koosneb hakest FCL element'idest. Esimeses kirjeldakse nõutud parameetri ledmiseks vajaik staatiliste väärtuste või otsisõnade kogum, teises nõutud parameetri leidmiseks vajalik loogika. Järgmisel skeemil on näha kuidas defineeritakse otsimisloogikat arve numbriga ja registrikoodi leidmiseks.



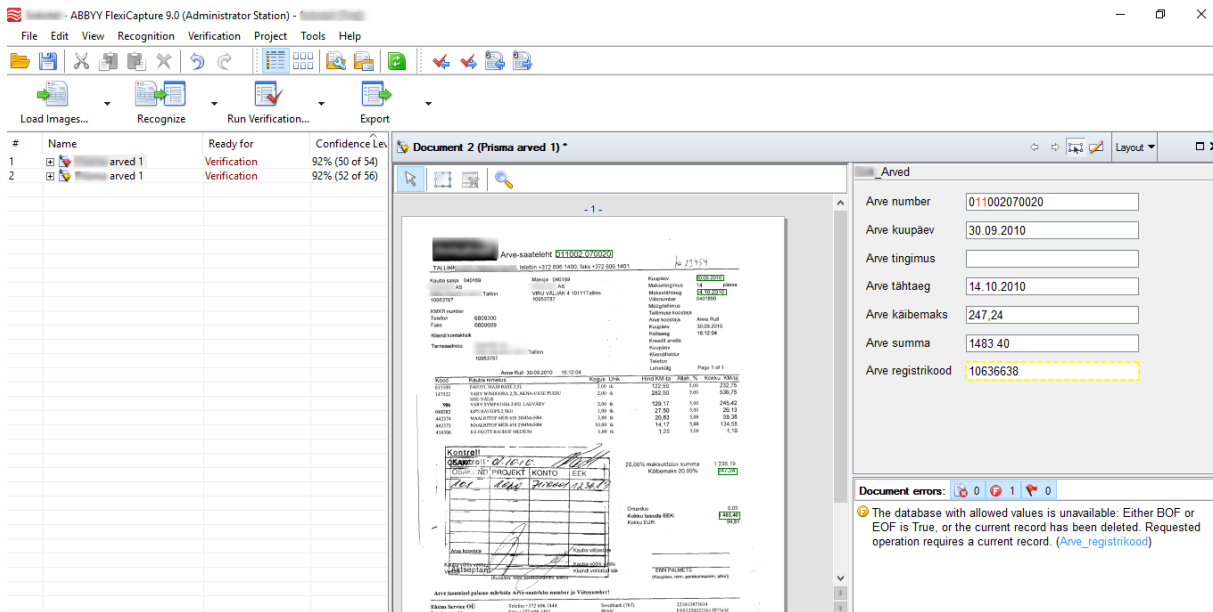
Joonis 9. FCL'i struktuur

Järgmisel pildil on toodud pilt FCL arendamiskeskonnas, kus on näidatud arve summa FCL block'i ja FCL elementide seos.



Joonis 10. FCL'i arendamiskeskonna vaade

Teine osa on FCP, kuhu laaditakse üks või mitu FCL'i. FCP'is arendaja poolt seadistatakse arvete automaatse või manuaalse impordimist, tuvastatud andmete valideerimist, arvete andmete esialgse automaatse redigeerimist, andmete automaatse või manuaalse eksportimise reegleid, kohtasid ja eksporditavate andmete formate.



Joonis 11. FCP keskkonna vaade

#	A...	Name	Ready for	Confidence Level	Processing Notes
5		Arvete_tuvastamine_2016	Verification	89% (67 of 75)	
6		<Unprocessed>	Recognition		
7		<Unprocessed>	Recognition		
8		<Unprocessed>	Recognition		
9		<Unprocessed>	Recognition		
1		Arvete_tuvastamine_2016	Export	100% (38 of 38)	
2		Arvete_tuvastamine_2016	Verification	96% (51 of 53)	
3		Arvete_tuvastamine_2016	Verification	98% (51 of 52)	
4		Arvete_tuvastamine_2016	Verification	91% (48 of 53)	

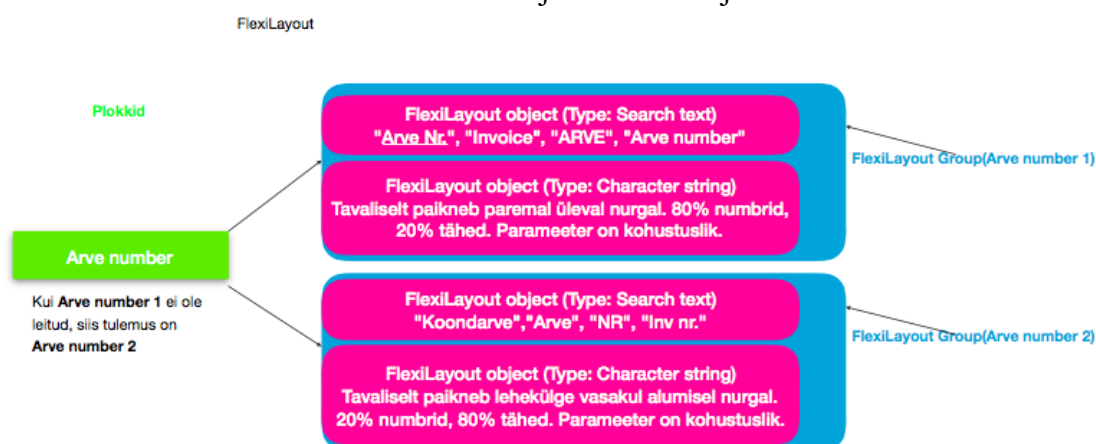
Joonis 12. FCP keskkonna vaade

Kasutaja puutub kokku ainult teise osaga ehk FCP'ga, kus saab arveid importida ja eksportida, jälgida iga arvete kohta tuvastatud andmete protsenti, käsitsi redigeerida valesti tuvastatud andmeid kas hiirega sobivate andmetega kohtade määramisega või klaviatuuri abil trükkimisega ja käivitada eksporti.

2.1.5 AAT ehitamisel võimalikud keerukused

Sellest, kuidas on FCL ehitatud oluliselt sõltub kogu AAT arenduse kvaliteet. Liiga suur FCL elementide hulk (rohkem, kui 3) aeglustab tuvastamise kiirust, liiga üldine otsimisloogika tõstab valesti tuvastatud andmete riiski. AAT arenduse optimeerimine on tihedalt seotud FCL analüüsiga.

Tihti juhtub, et erinevatel nõutud parameetrite kuju, tüüpiline paiknemiskoht või mingi muu parameeter kahel arvete näidistel varieerub oluliselt. Näiteks, paikneb arvel "A" arve number vasakul ülemises nurgas ja koosneb ainult arvudest, samas kui arvel "B" on samad andmed teisel kujul ja alumises nurgas. Selleks, et arve numbril kirjeldus ei muutuks liiga üldiseks ja otsitava parameetri kriteeriumi kirjeldused ei kaotaks täpsust ega käiks üksteisele vastu, lahendusena on võimalik arve "A" ja arve "B" arve numbrite jaoks kirjeldada otsimisloogikat kahel eraldi FCL'group'idel, ning seadistada plokkis if-then-else tüüpi kontsruksiooni arve numbril objekti leidmise jaoks.



Joonis 13. Keerukama FCL struktuuri näide

2.1.6 Võimalikud lahendused suure arvete disainide varieeruvuse puhul

Juhul, kui ühes FCL kirjeldada liiga palju keerukamaid FCL element'e ja group'e, muutub AAT arenduse andmete tuvastamise protsess liiga aeglaseks ning see keerukus ei garanteeri head tulemust. Arvete erinevate varieeruvuse taseme järgi võib valida erinevaid FC ehitamise struktuure.

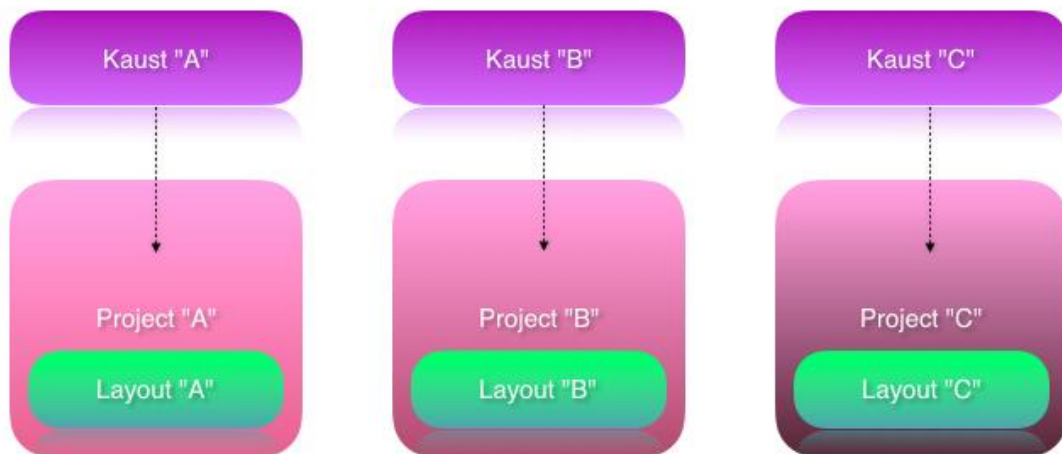
2.1.6.1 Madal disainide varieeruvus

Kõik arved on sama tüübi. Arvete elemendid vajavad 1-3 FCL group'e. Võimalik FC ehituse struktuur on üks FCP, mis interpreteerib ühe üldist FCL'i.

2.1.6.2 Suur disainide varieeruvus

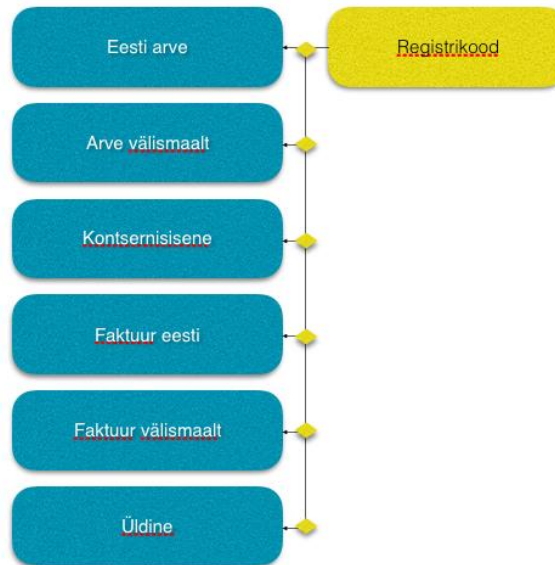
Arvede otsisõnade paiknemiskohad on erinevad, andmete formaadid erinevad oluliselt, mõned disainid on esitatud tabelina, kirjeldatud FCL'is tingimused võivad üksteisele vastu minna.

Esimene võimalik FC ehituse struktuur: jagada arveid disainide järgi gruppideks ja luua mitu erinevat FCP'e ja FCL'e nii, et üks FCP interpreteeriks ühe FCL'i.



Joonis 14. AAT ehitus, kus on FCP'd ja FCL'id jagatud mitmeks

Teine võimalik FC ehituse struktuur: luua FCL'ide hieraarhiat ja seadistada FC mingi kindla parameetri järgi otsida sobiva FCL'i.



Joonis 15. AAT näide, kus on kasutatud FCL'ide hieraarhia

2.1.7 Esmase arvete komplektide analüüs

Kuna FCL loogikast oluliselt sõltub kogu AAT efektiivsus, on vaja aru saada kuidas esmase arve komplekti põhjal oli defineeritud rekvisiitide otsimisloogika. Selleks uurib autor iga potentsiaalse FCL block'i jaoks olulisemaid andmete otsimise kriteeriume: iga rekvisiidi erinevate sümbolite esinemise protsenti, keskmiselt esinevate tähemärkide arvu, tüüpilist paiknemiskohta, otsisõna korrelatsiooni otsitavate andmetega, arvel esinevate kuid välditavate andmete määretlemine, kohustuslikke andmete tuvastamine. Tellija poolt on iga arve kohta nõutud järgmised rekvisiidid: arve number, arve summa, käibemaks, kuupäev, tähtaeg, viitenumber, registrikood, hankija ja arve saaja.

2.1.7.1 Arve number

Tüüpiline paiknemiskoht: antud arvetel asub arve number paremal üleval nurgas;
 Otsitav pikkus: varieerub kahest kuni üheteistkümne tähemärkini;
 Sümbolite esinemise protsent: 81% arvetel koosneb arvenumber ainult arvudest, ülejäänud 19% sisaldab arve number ka tähti, 12,5% kuni 25%;
 Otsisõna korrelatsioon otsitavate andmetega: arvete numbrid võivad asuda otsisõna kõrval vasaku pool, paremal või all;
 Kohustuslikkus: iga arve puhul on alati kohustuslik;
 Otsitava parameetri korrelatsioon teiste parameetritega: arve number asub teistest parameetritest kõrgemal;
 Selleks, et arve numbril kirjeldused ei oleks käinud üksteisele vastu ega kaotaks täpsust, otsimisloogika on kirjeldatud kasutades 2 FCL group'i, kus iga FCL group omakorda sisaldab 2 FCL element'i;

2.1.7.2 Arve summa ja käibemaks

Tüüpiline paiknemiskoht: Lehekülge paremal alumine pool. Summa on alati käibemaksu all
 Otsitav pikkus: ei saa kindla vahemikku määrata, kuna tegemist on hinnaga. Miinimum väärtus on 0,00;

Sümbolite olemasolu protsent: 1 jagatud tähtede arvuga * 100% on koma, ülejäänud on numbrid;

Otsingusõna korrelatsioon otsitavate andmetega: Arvete andmed võivad asuda otsingusõnast kõrval paremal või all;

Kohustuslikkus: iga arve puhul on alati kohustuslik;

Arve summa peab olema käibemaksu all;

2.1.7.3 Arve kuupäev ja tähtaeg

Tüüpiline paiknemiskoht: lehekülge ülemine parem pool;

Otsitav pikkus: 6 varianti. Esimene: 1.3.2016, Teine: 01.03.2016, Kolmas: 1.märts 2016;

Sümbolite olemasolu protsent: varieerub;

Otsingusõna korrelatsioon otsitavate andmetega: Arvete andmed võivad asuda otsingusõnast kõrval vasakul, paremal, üleval või all;

Kohustuslikkus: arve kuupäev on iga arve puhul on alati kohustuslik, tähtaeg ei ole;

Selleks, et kuupäevade kirjeldused ei oleks käinud üksteisele vastu ega kaotaks täpsust, antud kriteeriumide otsimiseks vajalik loogika on kirjeldatud kasutades 6 FCL group'i, mis kus iga FCL'group omakorda sisaldab 2 FCL element'i.

2.1.7.4 Viitenumber

Tüüpiline paiknemiskoht: Arve header või footer;

Otsitav pikkus: Näidiste komplekti arvetel selle pikkus on tavaliselt 6 kuni 22 märki;

Sümbolite olemasolu protsent: 100% numbrid;

Otsingusõna korrelatsioon otsitavate andmetega: 2 arvetel otsingusõna all, ülejäänud 20 paremal pool;

Kohustuslikkus: ei ole kohustuslik, kuna on määratus ainult 22 arvete eksemplaaridel

Viitenumbri tuvastamise loogika on kirjeldatud kasutades ühe FCL group'i;

2.1.7.5 Registrikood

Tüüpiline paiknemiskoht: Arve header või footer;

Otsitav pikkus: Näidiste komplekti arvetel selle pikkus on tavaliselt 6 kuni 22 märki.

Sümbolite olemasolu protsent: 100% numbrid;

Otsingusõna korrelatsioon otsitavate andmetega: 22 arvetel otsingusõna all, ülejäänud 24 paremal pool;

Kohustuslikkus: iga arve puhul on alati kohustuslik;

Vältimine: Registrikoodi tuvastamise käigus palus tellija vältida tütarettevõtete registrikoodide. Selleks on vaja luua keelatud registrikoodide nimekirja;

Registrikoodi tuvastamiseks on loodud kaks FCL group'i.

2.1.7.6 Hankija

Tüüpiline paiknemiskoht: Arve footeri vasak pool;

Otsitav pikkus: Näidiste komplekti arvetel selle pikkus on tavaliselt 6 kuni 22 märki.

Sümbolite olemasolu protsent: 99% tähed;

Tüüpilised otsingusõnad: ettevõtte juriidilise vormi nimetused ("AS", "OÜ", "MTÜ" jne);

Otsingusõna korrelatsioon otsitavate andmetega: kuna otsingusõnad ettevõtte juriidilise vormi nimetus, siis vasakul või paremal. Enamikul näidistel otsitavad andmed asuvad otsimissõnast paremal poolel;

Kohustuslikkus: iga arve puhul on alati kohustuslik;

Vältimine: Peab vältima pangade nimed ja tütaretevõtete nimed;

Hankija tuvastamiseks on loodud üks FCL group.

2.1.7.7 Arve saaja

Tüüpiline paiknemiskoht: Kindel koht puudub;

Otsitav pikkus: Näidiste komplekti arvetel selle pikkus on tavaliselt 6 kuni 22 märki;

Sümbolite olemasolu protsent: 99% tähed;

Tüüpilised otsingusõnad: kindla otsinguväärtuste nimekiri, mis koosneb tütaretevõtete ärinimedest;

Kohustuslikkus: iga arve puhul on alati kohustuslik;

Vältimine: Peab vältima hankijate ja pangade nimesid;

Arve saaja leidmiseks on loodud üks FCL group;

2.1.7.8 Esialgse arvete rekvisiide ja FCL'i analüüsi tulemused

On näha, et esialgse komplektist arvete rekvisiitide formaatide varieeruvus ei ole liiga suur. Selletõttu on FC arendus oli loodud kasutades sellist struktuuri, kus on üks FCP ja üks FC, kuid kahjuks uude arvete puhul see ei tööta stabiilselt. Uude disaini ei suuda FC lugeda, kuna FCL ei ole nende tuvastamiseks sedistatud.

2.1.7.9 Uude komplekti tuvastamisega seotud probleemid

Kui võrrelda kahte arvete komplekte, siis on näha, et nendel on erinevused, mis põhjustavad järgmiseid probleeme:

1. Tütaretevõtte arvete tuvastamine. Kuna tütaretevõtete registrikoodide ja tütaretevõtete nimekiri oli arendaja poolt pandud otsimise käigus vältimiseks, siis ei oska FC seda tuvastada;
2. Koondarvete tuvastamine. Koondarve sisaldab palju ridu, kus korduvad varem ühe parameetri otsimise jaoks määratud sõnad "Arve nr." ja "Kokku". Tulemusena FC võtab esimese leitud rea;
3. Faktuurarvete tuvastamine. Faktuurarved nagu koondarved sisaldavad palju korduvaid ridu, kus korduvad otsimissõnad "KM" ja "Kokku";
4. Notari poolt saadetud arvete tuvastamine. Notari arve peal registrikoodi asemel kasutakse isikukoodi. Notari nime juures puudub ettevõtte juriidilise vormi nimetust, kuna seadistuse järgi notar on ettevõtja. Selle tõttu "Hankija" jääb leidmata;
5. FIE poolt saadetud arvete tuvastamine. Arvete komplekti eksemplaaril puudub ettevõtte juriidilise vormi nimetus. Selle tõttu "Hankija" jääb leidmata;
6. Tabelina esitatud andmete tuvastamine. FCL on õpitud otsida andmeid ülalpoolt alla. Tabeli puhul osa andmeid on paigaldatud vasakult paremale;

2.2 Arvete tuvastamine ja dünaamiline otsisõnade täiendamine

2.2.1 Uue otsisõnade lisamine on võimalik ainult arendaja kaudu

Tellija poolt oli sõnastatud probleem, et uude otsisõnade lisamiseks on vaja paluda arendaja abi, mis toob ebamugavust.

Tavaliselt hinna tõttu ostab tellija endale piiratud tarkvara litsentsi, mis teeb FC seadistamist vähem paindlikumaks ehk keerukamaks. Kuna tellija ostetud litsents võimaldab kasutada ainult FCP't, Iga uue muudatust saab teha ainult arendaja abil. Praeguses AAT arenduses arve rekvisiitide leidmiseks otsisõnad on kirjutatud nimekirjana FCL elementidel.

2.2.2 Mõõdikud

1. Madal tuvastamise kvaliteet ja aeglane kiirus kulutab töötaja 10% tööajast;
2. Keskmine ühele arvele tuvastamisele kuluv aeg on 7 sekundit;
3. Keskmine õigesti tuvastatud andmete protsent on 50%;
4. Liigne inimkäsitöö redigeerimisel;
5. Kõrgendatud oht inimliku eksimuse esinemiseks;

Tabel 1. Arvete tuvastamise SWOT-tabel

<p>Tugevused:</p> <ul style="list-style-type: none"> • Paljude arvete jaoks osaliselt sobib; • Lihtne ja minimalistlik FC ja FCP ehituse struktuur; 	<p>Nõrkused:</p> <ul style="list-style-type: none"> • Aeglane; • Ebatäpne - mõned andmed valesti tuvastatakse; • Uued disainid ei oska üldse vastu võtta; • Uude otsisõnade sisestamise puudus; • Suur kasutaja töötaja kulu kontrollimiseks; • Stabiilsuse puudus;
<p>Võimalused:</p> <ul style="list-style-type: none"> • Töötaja kulu kokkuhoid; • Tuvastamise protsessi kiirustamine; • Tuvastamise tulemuste muutmise täpsusemaks; • Muutmise otsisõnade kogumite dünaamiliseks; 	<p>Ohud:</p> <ul style="list-style-type: none"> • Muutmise FCL ja FCP struktuuri keerukamaks, mille eest on raske hoolata; • Jälle ilmuvad arvete uued disainid, mille FCL ei oska korrektselt lugeda;

2.3 Arvete andmete valideerimine

2.3.1 Andmete formaadi kontroll

Kuna FC tuvastab andmeid optiliselt, see tähendab, õigesti tuvastatud andmete protsent sõltub mitte ainult FCL ja FC ehitamist, vaid ka arve väljatrüki kvaliteedist. Vaatamata sellele, et FCL võib olla targalt ja täpselt arendatud, väljatrüki kvaliteedi tõttu võivad olla tuvastatud ebavajalikkud tähemärgid Selle vältimiseks peab kasutaja tähelepanelikult tuvastatud andmeid üle vaadata.

2.3.2 Tuvastatud registrikoodi olemasolu

Tihti tekib olukord, Kui arve on kahtlane, kasutaja peab kontrollida arve hankija registri numbri MSDAX'i andmebaasis, mis kulutab aega.

2.3.3 Mõõdikud

1. Kontroll võtab 2% kasutaja tööajast;
2. Liigne inimkäsitöö redigeerimisel;
3. Kõrgendatud oht inimliku eksimuse esinemiseks;

Tabel 2. Arvete valideerimise SWOT-tabel

<p>Tugevused:</p> <ul style="list-style-type: none"> • Minimalistlik ja lihtne arenduse struktuur; 	<p>Nõrkused:</p> <ul style="list-style-type: none"> • Vajadus kasutaja pideva ja tähelepanelikus kontrollis;
<p>Võimalused:</p> <ul style="list-style-type: none"> • Väiksem võimalus inimlikeks vigadeks; 	<p>Ohud:</p> <ul style="list-style-type: none"> • Importida MSDAX'sse andmeid vale formaadis või liigsete sümbolitega; • Importida MSDAX'sse ebakorrektsed ostuarvet hankijalt, kelle registrikood puudub andmebaasis;

2.4 Arvete andmete eksport

Arvete andmete kontrolli pärast käivitab kasutaja kas ühe või kõikide arvete ekspordi. Ekspordi käigus peab FC otsustama mis kausta peab CSV fail arve andmega minema. Otsustamine käib CSV'faili nime moodustamisel. Kogu kaustade struktuuri asukoht on määratud FC seadistuses "root"-direktooriumina, vaid kausta valimine tütarettevõtte jaoks moodustakse tuvastatud arve saaja väärtuse järgi. Sellel põhjusel, et kaustade nimed ei ole sarnased võimalike arve saaja väärtusega, kasutusoleval arendusel on loodud script, mis asendab tuvastatud arve saaja väärtuseid kaustade nimedeks. Kui arve saaja väärtus on valesti tuvastatud, siis script ei suuda suvalise väärtuse asendada õigeks ja tulemusena luuakse uus kaust, kuhu läheb CSV fail.

2.4.1 Mõõdikud

1. Liigne inimkäsitöö redigeerimisel;
2. Kõrgendatud oht inimliku eksimuse esinemiseks;

Tabel 3. Arvete andmete eksportimise SWOT-tabel

<p>Tugevused:</p> <ul style="list-style-type: none"> · Lihtne ja minimalistlik arenduse ehitus 	<p>Nõrkused:</p> <ul style="list-style-type: none"> · Ebastabiilsus ja ebatäpsus eksportimisel
<p>Võimalused:</p> <ul style="list-style-type: none"> · Väiksem võimalus inimlikeks vigadeks; · Töötaja aja kokkuhoid; 	<p>Ohud:</p> <ul style="list-style-type: none"> · Vale direktoriumi moodustamisel võib kaotada CSV-faili andmetega

2.5 Tellija nõudmised

Tellijal esitas nõudmisi, mille optimeerimise käigus üritakse saada.

Keskmine arvete tuvastamise protsent peab olema kõrgem, kui 85% ja tuvastamisele kuluv aeg peab olema maksimaalselt 4 sekundit;

1. FC peaks lubama ekspordi ainult juhul, kui CSV faili nime moodustamiseks vajalikud rekvisiidide väljad on tuvastatud ja täidetud;
2. FC peab lubama CSV ekspordi ainult juhul, kui arve saaja väärtus on asendatud väärtusega, mis kuulub olemasoleva kaustade nimede hulka;

3 AAT protsesside optimeerimine

Protsesside kirjeldamisel ja tellija tagasiside ülevaatamisel selgus, et AAT protsessid ei ole piisavalt efektiivsed.

Selleks, et protsesside efektiivsust tõsta, tuleb muuta FC ehituse struktuuri, lisada tuvastatud rekvisiitide valideerimist, seadistada dünaamilise otsisõnade täiendamise võimalust ning piirata FC eksportimisreegleid.

3.1 Arvete tuvastamine

Arvete tuvastamise kiiruse tõstmiseks ja kvaliteedi parandamiseks tuleb muuta FC struktuuri. Selleks on kolm varianti.

3.1.1 Võimalikud FC ehitusstruktuurid kvaliteedi ja kiiruse optimeerimiseks

3.1.1.1 FCL'i laiendamine

Võib täiendada olemasoleva FCL'i, lisades sinna uued FCL group'id ja FCL element'id. Mõnede arvete tüübidele see sobiks, mõnedele mitte. Kontsernisestest arvete probleemi puhul võiks teha nii, et FCL esialgu leiaks registrikoode, mis ei ole kontsernisisesed ja kui nad ei ole leitud, siis võtaks arve Footerist olemasoleva.

Antud FC ehitus on hea seetõttu, et nagu oli varem, jäävad üks FC ja FCP. Üldine FC ehitus jääb lihtsaks vähe FC komponentide põjusel.

Kahjuks see loogika teeks iga arve tuvastamisele kuuluva aega 10 sekundi võrra pikemaks, kuna kontsernisestest arvete puhul see sundib FCL otsida 2 korda rohkem registrikoode. Kui defineerida sama loogikat koos hankijate nimekirjaga, et FCL ei saa esialgu Hankija nimeks võtta tütarettevõtte ärinime, siis iga arve tuvastamiseks kuuluv aeg muutub veel 10 sekundi võrra kauemaks.

Mis puudutab faktuurarveid ja saateleheid, kus võiks kirjeldada korduvaid objekte (näiteks "Arve nr.", "Käbemaks", "Kokku (EUR)") vältimisloogikat, mis pikendaks keksmist arvete tuvastamise protsessi keskmiselt 5 sekundi võrra veel.

Juhul, kui kirjeldatud täiendusega laekub 40-leheküljeline kontsernisisene faktuurarve, siis FC viskab kõik ressursse selle töötlemiseks. Sellel on kolm suurt ohtu:

1. Teised arved jäävad seisma;
2. FCP hangub tuvastamise käigus ja võib sulgeda vigadega;
3. AAT muutub kohe ebaefektiivsemaks;

3.1.1.2 Eraldi FCP'd ja FCL'id (1:1)

Võib jagada FCP mitmeks. Iga projekti jaoks kirjeldada eraldi FCL . Aga sellel lahendusel on puudused:

- a. Kasutaja on kohustatud arveid enne tuvastamist sorteerida;
- b. Kasutaja võib sorteerimise käigus eksida;
- c. Mõned arveid on raske klassifitseerida;

- d. Kasutaja peab hoidma peas kuidas arveid jagada;
- e. Kasutaja võib mitte teada kuidas klassifitseerida antud arve;
- f. Kõik üleval toodud põhjused ei võimalda protsessi automatiseerida;

3.1.1.3 FCL'ide hieraaria põhjaliku FCL'i sees

Võib luua FCL'i mitu alam FCL'idega erinevate arvete tüüpide jaoks ning valida õiget FCL'i teatud kindla klassifikaatori järgi, näiteks registrikoodi või käibemaksukohuslase numbriga järgi.

Iga alam FCL peab olema lihtne ja minimalistlik, väga kitsalt kirjeldatud, vähe ja täpse kriteeriumidega. Iga FCL block'ile peab vastama ainult üks FCL group.

Antud juhul on järgmised ohud:

- Kui arve peal klassifikaator nimekirjas puudub, siis jääb see arve tuvastamata. Selle vältimiseks peavad FCL'id moodustama hierarhiast, mis tähendab, et juhul, kui arve on jäänud klassifitseerimata, siis läheb ta viimasele kõige üldisemale FCL'ile;
- Keerulise klassifitseerimise puhul alam FCL'i valimiseks läheb aega;
- FCL'ide hulk aeglustab tuvastamisprotsessi. Kuna FCL'e on palju, siis saab neid teha väga lihtsamaks ja minimalistlikumaks. Iga parameetri jaoks luua ainult ühe vastava kriteeriumi, mis muutub tuvastamise protsessi kiiremaks;
- Kuna otsimiskriteeriumid muutuvad kitsamaks, teeb see tuvastamisetulemuseid täpsemaks ja kindlamaks. See võimaldab vältida kontsernisisesega arvetega ja faktuurarvetega tekkinud probleeme, kuna iga selle arve tüübi jaoks määratakse eraldi FCL'e.

3.1.1.4 Lahenduse valimine

Kuna kasutuselolev FC on juba aeglane ja ebatäpne, kõige paremaks valikuks oleks luua FCL'i alam FCL'idega.

Selleks võib jagada kõik arveid 9 gruppideks:

- 1) Eesti arved;
- 2) Eesti arved tabelina;
- 3) Arved välismaalt;
- 4) Arved välismaalt tabelina;
- 5) FIE arved;
- 6) Kontsernisisesed;
- 7) Eesti faktuurarved;
- 8) Faktuurarved välismaalt;
- 9) Koondarved;

Selleks tuli luua iga gruppi jaoks 9 alam FCL'i kitsa parameetridega ja ühe üldise parameetriga.

3.1.1.5 Mõõdikud

1. Väheneb vigade tekkimise tõenäosus;
2. FC on laiendatud suurte arvete disainide varieeruvuseks;
3. Väheneb tuvastamisele kuluv aeg;

4. Pareneb arvete tuvastamise kvaliteet;

3.2 Arve otsisõnade kogumite dünaamiline täiendamise puudus

3.2.1 Võimalused

1. Seadistada FCL'i niimodi, et ta võtaks väärtuseid välisallikatest - tekstifailist Sellel lahendusel on puudused: Tekstifailist saab kasutaja andmeid ära kustutada. Aga kui otsing siis majandustarkvara andmebaasis saab tabelis anda kasutamissoigusi, ning määrata sellele vastutavat isikut.

1. Seadistada FCL'i niimodi, et ta võtaks väärtuseid välisallikatest - MSDAX andmebaasist.

Välisallikate seadistamine kasutaja jaoks annab kasutajale järgmised võimalused:

- a. Notari ja FIE nimede otsimiseks määrata nende nimed ettevõtte juriidilise vormi nimetuste nimekirjale;
- b. Luua uued tütarettevõtte kaustad ja seadistada eksporti;
- c. Lisada teistes keeltes otsimissoõnu üldiste disainidega arvete jaoks;
- d. Lisada registrikoode arve jaoks, millel on sarnane disain esitatud arvete komplektist;

Mõõdikud:

1. Väheneb sõltuvus teenindava arendajalt

3.2.2 Valitud lahendus

Kuna tellija igapäevane töö toimub kasutades MSDAX, on võimalik luua tabelit väärtustega, kust FCL on seadistatud otsisõnade kogumit välja võtta.

3.3 Tuvastatud rekvisiitide valideerimise puudus

Vigade vältimiseks on vaja seadistada enamikku tuvastatud rekvisiidi valideerimist.

3.3.1 Arve number

Valideerimisreegel peab automaatselt kustutada tühikuid, punktid ja komad. Ei saa olla tühi, kuna sellest moodustab eksporditava CSV faili nimi.

3.3.2 Summa ja käibemaks

Valideerimisreegel peab jälgida seda et tuvastatud summa ei sisalda tähteid, kriipsud ja tühikuid. Summa ja käibemaks Peavad olema numbrilised

3.3.3 Arve kuupäev ja tähtaeg

Kuupäeva välismaalse formaadi muutmise jaoks peal olema loodud script, mida käivitakse koos tuvastamisega ja manuaalne nupp, mida kasutaja võib oma mugavuseks kasutada.

Arve kuupäev ei saa olla tühi, kuna sellest moodustab eksporditava CSV faili nimi.

3.3.4 Viitenumber

Valideerimisreegel peab jälgida seda, et viitenumber peab olema numbriline.

3.3.5 Hankija

Valideerimisreegel peab kontrollida seda, et hankija väli ei saa olla tühi, kuna sellest moodustab eksporditava CSV faili nimi.

3.3.6 Registrikood

Ühendada FCP MSDAX hankijate andmebaasiga ning lubada eksporti ainult juhul, kui tuvastatud registrikood on baasis olemas.

3.3.7 Mõõdikud

1. Väheneb vajadus tähelepaneliku kasutaja kontrollis. Juhul, kui esineb mingi formaadiviga, FCP annab kasutajale sellest teada.

3.4 Eksportimine valedesse kaustadesse

Selleks, et FC tuvastamise vea tõttu ei tekitaks uusi kaustaid, seadistakse FCP ühendust MSDAX andmebaasiga ning tuvastamise protsessis peab FC asendada tuvastatud arve saajat lubatud kausta väärtusega ning lubada eksporti ainult juhul, kui arve saajaks on määratud MSDAX andmebaasis olemasolev kausta nimi.

3.4.1 Mõõdikud

1. Väheneb failide kaotamise risk
2. Failide eksport muutub täpsemaks ja stabiilsemaks

4 Optimeeritud AAT arenduse ülevaade

4.1. Kirjeldus

On tehtud olulised täiendused, nagu FC struktuuri muutmine, dünaamilise otsisõnade täiendamise võimaluse loomine, andmete valideerimine ja ekspordi piiramine.

Tabel 4. Mõõdikute võrdlus

Mõõdik			Vahe
<i>Töötaja ajakulu</i>	12 %	2%	10%
<i>Kiirus</i>	7 sekundit ühe arve kohta	4 sekundit ühe arve kohta	3 sekundit
<i>Õigesti tuvastatud andmete protsent</i>	50%	85% (Kontserniseseid ja eesti ostuarved 100%-ni)	35%

Kontserniseseid ja tavaliseid eesti tavalised ostuarveid FC arendus tuvastab 100% õigesti mille tõttu olid kliendid nõus lasta arendajale seadistada 100% tuvastatud andme puhul lubada automaatse ekspordi, mis on optimeerimise tulemuslikkuse näidur.

5 AAT arenduse ülevaade

See osa on loodud selleks, et anda ülevaadet AAT arenduse aktuaalsusest ja vajalikkust.

5.1 Võimalikud alternatiivsed lahendused

Ühe lahendusena on otstarbetu kulu vältimiseks kasutatud ettevõtete Omniva ja Fitek digiteerimist, mida firmad saavad teenusena sisse osta, kuid sellisel töökorraldusel on oma miinused. Peamine puudus on arvetel toodud andmete avalikustamine, suhteliselt kallis teenus ja samuti asutuse-spetsiifiliste pädevuste puudumine.

5.2 AAT eelised

Mõistlikult seadistatud AAT aitab vältida inimfaktoriga seotud vigu, raha- ja ajakulu. Võrreldes alternatiivse Fiteki ja Omniva poolt pakutava digiteerimisega annab FC'i kasutamine võimaluse avalikustamata firma finantstegevust.

5.3 AAT puudused

Vaatamata sellele, et tänapäeval seaduse järgi võib paber kandjal arvet esitada, on tõenäoline, et aja jooksul minnakse täielikult üle elektroonformaadis arvetele ning vajadus tuvastamises kaob. Praeguseks momendiks ei ole täpselt teada millal see juhtub.

5.4 FC eelised

FC on väga paindlik, sest seda on võimalik seadistada erinevatel viisidel. Näiteks saab seda ühendada andmebaasiga, valida erinevaid eksportimisformaate, seadistada failide nimetamist ja andmete valideerimist kasutades programmikoodi ja määrates seadistusi, mis võimaldavad integreerida erinevate majandustarkvaradega.

Juhul, kui FC arendus on loodud keskmise või madala tuvastamise kvaliteediga, on see tarkvara kasutaja jaoks mugavam kui traditsioonilisel arvete töötlemisel. Esiteks annab see kasutajal järgmised võimalused ja eelised:

- Teostada samaaegselt mitmeid erinevaid tööülesandeid;
- Iga parameetri andmeid (näiteks arve number) võib käsitsi trükkimise asemel valida hiirega sobiva piirkonnast;

Lisaks FC võimaldab seadistada automaatse eksporti arvetel, kus on tuvastatud rohkem, kui seadistuses määratud protsent andmeid ja valideerimine õnnestus. See tähendab, et targalt planeeritud ja arendatud FC arenduse puhul on potentsiaalne võimalus arvete tuvastamise protsessi osaliselt või täielikult automatiseerida.

5.5 FC puudused

5.5.1 Muudatused

Tavaliselt hinna tõttu ostab tellija endale piiratud tarkvara litsentsi, mis teeb FC projekti seadistamist vähem paindlikumaks ehk keerukamaks. Iga uue muudatuse loomine sõltub teenuse pakkuva firma hinnakijast ja pädevusest.

5.5.2 FC stabiilsus

FlexiCapture on optilise tuvastamise tehnoloogiat kasutatav tarkvara, mis tähendab, et arve tuvastamise protsess ei pruugi alati olla stabiilselt täpne, sest vigade tõenäosus sõltub ka skanneeritud arve pildi kvaliteetidst.

5.5.3 FC ja arvete varieeruvus

FC ilmus riigis, kus selle loomise hetkel olid kõik arved ja blanketid unifikseeritud ehk pidid vastama üldisele riigi standardile. Sellest tekib järeldus, et see tarkvara on pigem loodud suure sarnast blanketide hulga tuvastamiseks, kui suure arvete disainide varieeruvuse töötlemiseks.

6 Kokkuvõte

Käesoleva töö “Paberivaba arvehalduse lahenduse arvete andmete tuvastamise protsessi optimeerimine ” eesmärgiks oli analüüsida AAT arenduse protsesse, uurida võimalusi seda optimeerida ja luua optimeeritud arendust.

Autor uuris töös kuidas oli esialgne AAT versioon loodud ja välja selgitas põhjuseid miks tekisid probleemid, vaatas võimalusi optimeerimiseks üle, valis nendest sobivama ning rakendas seda.

AAT põhilisteks puudusteks olid AAT arvete andmete tuvastamise madala kiiruses ja madala õigesti tuvastatud andmete protsendis, dünaamiliste seadistamise võimaluste puudumises ning valideerimise ja failide eksportimise piiramise puudumises. Nimetatud puudused suurendasid töötaja ajakulu, tõstsid valesti tuvastatud arve rekvesiidi tõenäosust ja eksporditava CSV faili kaotamise võimalust.

Optimeeritud AAT näitas head tulemusi ning perspektiivis hoiab rahakulu kasutaja tööajale kokku, aitab hoida arvete informatsiooni avalikustamata ja kiiresti ja lihtsalt manageerida suure hulka saabuvate arvetega.

Kasutatud kirjandus

[1] ABBY FlexiCapture koduleht [Online]

<https://www.abby.com>

[2] ABBY FlexiCapture 11.0 Admin Guide [Online]

<https://www.abby.com/media/4399/flexicapture-11-admin-guide.pdf>

