SAIMAAN AMMATTIKORKEAKOULU
Tekniikka, Lappeenranta
Tietotekniikka
Viestintätekniikka

Riku Heikkilä

Käsin ladotun aikakausilehtitekstin muuttaminen digitaaliseen muotoon

Opinnäytetyö 2011

TIIVISTELMÄ
Riku Heikkilä
Käsin ladotun aikakausilehtitekstin muuttaminen digitaaliseen muotoon, 49 sivua
Saimaan ammattikorkeakoulu, Lappeenranta
Tekniikka, Tietotekniikan koulutusohjelma
Viestinnän suuntautumisvaihtoehto
Opinnäytetyö 2011
Ohjaajat: Tuntiopettaja Yrjö Utti, Saimaan ammattikorkeakoulu Oy,
Jukka Seppänen, Maasotakoulu

Opinnäytetyönä tehtiin selvitys, kuinka ja millä testintunnistusohjelmalla saada toisen maailmansodan aikaiset, käsin ladotut Sotilaspapin aikakauslehdet digitaaliseen muotoon tutkimuskäyttöön. Selvityksessä käytettiin tekstintunnistusta niin, että siihen voi tehdä sanahakuja. Opinnäytetyön asiakkaana toimi Maasotakoulun sotilaspastori Jukka Seppänen. Aineisto ja tutkimusvälineet kerättiin Internetistä.

Selvitys tehtiin saatavilla olevia ohjelmia käyttäen. Ohjelmia oli paljon ja niiden toimivuus tätä työtä ajatellen vaihteli suuresti. Ohjelmien ilmaisuus oli näennäistä, sillä kun ilmaisia ohjelmia löytyi, niin iso osa oli niin sanottuja maksullisten ohjelmien kokeiluversioita, joissa oli omat rajoituksensa liittyen useimmiten joko ohjelman käyttöikään, tallennettavien sivujen määrään tai molempiin.

Käyttökelpoisimmaksi vaihtoehdoksi tarjolla olevista ohjelmista osoittautui Nuance PDF Converter professional 7 -ohjelmalla tehty vuosikerta yhteen PDF-tiedostoon, johon voi tehdä sanahakuja. Haussa sillä on puutteita, pääasiassa harvaan kirjoitettujen nimien paikantamisessa. Ensimmäisessä tarkastelussa ohjelmasta virheitä ei löytynyt monta, mutta toisen tarkastelun jälkeen virheitä löytyi useampia. Tarjolla olevien ohjelmien huonon laadun vuoksi alkuperäissuunnitelmaan kuulunut tietokantavaihtoehto muuttui tutkimuksen edetessä yhä epätodennäköisemmäksi. Suuri syy tähän oli tekstintunnistuksen heikon laadun tuomat pulmat.

Asiasanat: PDF, tekstintunnistus, käsin ladottu, OCR

ABSTRACT
Riku Heikkilä
Transforming a hand stacked magazine in to digital form, 49 pages
Saimaa University of Applied Sciences, Lappeenranta
Technology, Degree Programme of Information Technology
Communications orientation
Bachelor's thesis, 2011
Instructors: Lecturer Yrjö Utti, Saimaa University of Applied Sciences,
Mr. Jukka Seppänen, Maasotakoulu

This Bachelor's thesis was conducted as an investigation of how to make a digitized version of a hand stacked Army priest magazine using text recognition to research use, so it would be possible to search certain words from it. The client of this thesis was army minister Jukka Seppänen. Materials and research equipment were gathered from the internet.

This thesis was executed with available software, which was plentiful, however the results varied a lot. As free as the programs were, there were a lot of so called trial versions of chargeable software which had their own limitations usually including low period of usage, maximum of saveable documents or both.

The most usable outcome was with Nuance PDF Converter professional 7, which allowed to create a volume to one PDF-file in which to make word searches. It still had flaws, mainly in searching infrequently written words. Because of the flimsy quality of the results from the character recognition, mainly in transforming the file in to text, the database option which was included in the original plan, had eventually to be discarded.

Keywords: PDF, volume, hand stacked, OCR

SISÄLTÖ

TERMIT JA LYHENTEET

BMP	Bittikarttakuva (Bit Map Picture), joka muodostuu pikseleistä eli kuvapisteistä.
Digikamera	Tunnetaan myös nimellä digitaalikamera. Se on kamera, joka tallentaa kuvattavan kohteen digitaalisesti.
Digitalisointi	Dokumentin siirtäminen digitaaliseen muotoon.
DOC	(Document) Tekstitiedostomuoto, joka on käytössä Microsoftin Word–ohjelmassa.
DOS	Disk Operating System eli levynkäyttöjärjestelmä on yhden käyttäjän komentorivipohjainen ei-moniajava käyttöjärjestelmä.
GIF	Graphic Interchange Format on bittikarttagrafiikan tallennusformaatti, joka käyttää häviötöntä pakkausta.
HTML	Hypertext Markup Language eli hypertekstin merkintäkieli on avoimen standardin kuvauskieli, jolla voi kuvata tekstiä, jossa on hyperlinkkejä, eli hypertekstiä.
JPEG	Joint Photographic Experts Group on bittikarttagrafiikan tallennusformaatti, joka käyttää häviöllistä pakkausta.
Krediitti	Credit on nettisivustoilla käytettävää valuuttaa.
MB	Megabyte eli megatavu. Tavu on tiedon tallennuskapasiteetin mittayksikkö tietokoneessa.
MHT	Tiedostotyypistä käytetään myös lyhennettä MHTML (MIME HTML), joka on web–sivun arkistoformaatti. Tätä käytetään yhdistämään kuvia, animaatioita ja äänitiedostoja HTML–koodin kanssa yhdeksi tiedostoksi.
OCR	Optical Character Recognition on teknologia, jonka avulla tunnistetaan koneellisesti tai käsin tuotettua tekstiä sähköiseen muokattavaan muotoon.
PBM/PGM/PPM	Portable bitmap format, Portable greymap format ja Portable pixmap format ovat avoimen lähdekoodin netpbm-projektin tuottamia kuvaformaatteja, jotka luotiin helposti siirreltäväksi eri ohjelmistojen välillä.

PCX	Personal Computer eXchange on kuvaformaatti, joka oli ensimmäinen hyväksytty DOS-kuvastandardi (Disk Operating system), mutta on vähentynyt käytöstä.
PDF	Portable Document Format on Adoben kehittämä PostScript-kieleen pohjautuva ohjelmistoriippumaton, siirrettävä tiedostomuoto, jota käytetään pääasiallisesti sähköiseen julkaisemiseen, tulostamiseen ja painamiseen.
PNG	Portable Network Graphics on bittikarttagrafiikan tallennusformaatti, joka käyttää häviötöntä pakkausta. Luotiin GIF:n (Graphic Interchange Format) korvaajaksi.
PostScript	Tietokoneiden sivunkuvauskieli, jolla kuvataan tulostimille ja muille laitteille tulostettavien dokumenttien ulkoasu.
PPT	Power Point Table on Microsoftin PowerPoint–ohjelman käyttämä tiedostomuoto.
RTF	Rich Text Format on muotoillun tekstin tallennusmuoto joka, toisin kuin tavallinen teksti (.txt), tallentaa tiedostoihin myös fontit, fonttikoot, kursivoinnit, lihavoinnit sekä alleviivaukset.
Selain	Ohjelma, jolla käyttäjä voi katsella ja lähettää kuvia, tekstiä ja muuta, mitä www-sivuilta löytyy.
Skannauskynä	Kynä, jolla voi skannata sanoja rivi kerrallaan.
Skanneri	Laite, joka muuttaa kaksiulotteisen kuvan kuvatiedostoksi.
Tekstin-tunnistus-ohjelma	Ohjelma joka tunnistaa kuvasta merkit ja tulostaa ne määrättyyn tiedostomuotoon.
TIFF	Tagged image file format tai Tag Image File Format, joista voidaan käyttää myös lyhennystä TIF. TIFF on kuvien tallennukseen käytetty tiedostomuoto, joka tukee niin kuvankäsittely- kuin tekstinkäsittelyohjelmia.
TWAIN	Technology Without An Interesting Name on kuvankäsittelyssä standardirajapinta, joka tehtiin varmistamaan, että eri laitteet ja ohjelmistot ovat yhteensopivia. Standardi mahdollistaa skannaamisen suoraan kuvankäsittelyohjelmasta.
TXT	Text file eli tekstitiedosto on tiedostomuoto, joka sisältää vain tekstiä.

Validointi	Prosessi, jolla varmistetaan, täyttääkö jokin asia tietyt vaatimukset.
WORD	Microsoftin kehittämä tekstinkäsittelyohjelma.
WWW	World Wide Web on Internet-verkossa toimiva hypertekstijärjestelmä.
XLS	Excel Spreadsheet on Microsoftin Excel–taulukkolaskentaohjelman käyttämä tiedostomuoto.
XML	eXtensible Markup Language on merkintäkieli, jolla tiedon merkityksen voi kuvata tiedon sekaan. Käytetään tiedonvälitykseen järjestelmien välillä ja formaattina dokumenttien tallentamiseen.

1 Johdanto

Yleensä digitoinnista eli digitalisoinnista puhuttaessa viitataan valokuvien, diojen tai mikrofilmien digitalisointiin. Dokumenttien muuttaminen digitaaliseen muotoon on aivan yhtä lailla digitalisointia, vaikka useammin tekstidokumenttien digitalisoinnista käytetään nimitystä tekstintunnistus tai OCR.

Opinnäytetyön tavoitteena on tutkia keinoja muuttaa toisen maailmansodan aikainen Sotilaspapin aikakauslehti digitaaliseen muotoon asiasanojen hakua varten. Tarkoitus on löytää ilmainen OCR- (Optical character recognition) eli tekstintunnistusohjelma, joka tunnistaa merkit mahdollisimman pienellä virhemarginaalilla. Ohjelmasta tehdään joko alkuperäisen tekstin siirto määritettyyn tietokantaan tai yksinkertaisemmin muutetaan tiedosto PDF:ksi (Portable Document Format), johon voi tehdä sanahakuja. PDF on sähköiseen julkaisemiseen, tulostamiseen ja painamiseen käytetty ohjelmistoriippumaton ja siirrettävä tiedostomuoto. Työn taustalla on aikakausilehden digitaalisen version tarve, koska digitaalisessa muodossa tieto säilyy ja tiedon etsiminen sisällöstä helpottuu.

Ohjelmien etsiminen tapahtuu Googlen avulla. Ohjelmat testataan tekstintunnistuksen jälkeen tutkimalla virheiden määrä, jonka jälkeen virheiden määriä vertaillaan ohjelmien kesken. Sen jälkeen valitaan ohjelma, jolla vuosikerrasta pystytään tekemään digitaalinen versio ja josta löytyy vähiten virheitä.

Tässä työssä käsitellään vain tekstintunnistusohjelmien ilmaisversioita, sillä maksulliset versiot voivat hinnaltaan kivuta useampaan sataan euroon eikä sellaiseen ollut tällä kertaa mahdollisuutta. Ensin tutkitaan selainpohjaisia OCR-ohjelmia, jotka eivät tarvitse lainkaan työasemaan asentamista toimiakseen, ja tämän jälkeen perehdytään asennettaviin ohjelmiin.

2 Tekstintunnistusohjelman valintaprosessi

Testausmateriaalin määritys aloitetaan määrittämällä testausmateriaali eli valitaan kuvatiedosto, joka on sisällöltään monipuolista. Sisällöstä on löydyttävä kaikki lähdemateriaalista löytyvät tekstityypit esimerkiksi eri tavoin painetut kirjaimet, harvaan kirjoitetut sanat ja eri tasoille painetut sanat. Vaihtelevuuden ja kattavuuden haluamiseksi testimateriaaleja voi olla pari lisää.

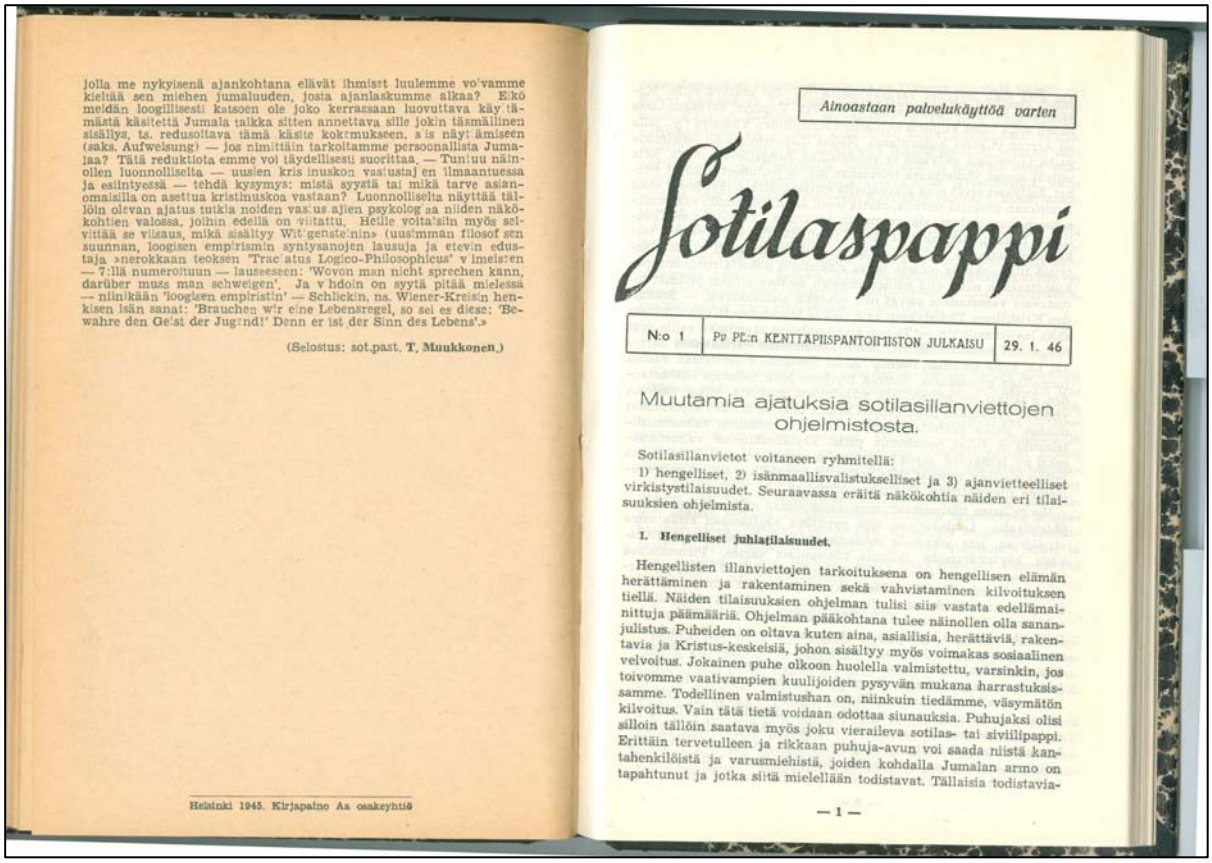
Internetistä etsitään kaikki tekstintunnistukseen viittaavat ohjelmat. Tehdään alustava tutkimus, onko ohjelmia mahdollista testata.

Ohjelmat testataan, kun testattavat ohjelmat on löydetty. Testattujen ohjelmien tuloksia verrataan keskenään. Pienimmän virhemäärän tulostanut ohjelma valitaan. Seuraavaksi materiaali käsitellään valitulla ohjelmalla.

Skannausprosessi ohjelman ollessa tuotannossa: ensin kohde määritetään, jonka jälkeen lähdemateriaali skannataan. Materiaali käsitellään tekstintunnistusohjelmassa, jonka jälkeen tulokset tarkistetaan ja verrataan lähdemateriaaliin. Lopuksi saadut tulokset siirretään Internetiin.

3 Testauskohteen esittely

Projektin kohteena on Sotilaspappi-niminen kenttäpiispan toimiston käsin ladottu aikakausilehti (kuva 1). Käsin ladottu tarkoittaa sitä, että kirjaimet ovat painoa varten käsin asetetut, jonka takia teksti ei saata olla joka rivillä samalla tasolla. Testattavana oli 46. vuosikerta, joka sisältää 114 sivua, jotka skannattiin eli muutettiin kaksiulotteinen kuva kuvatiedostoksi. Skannauksen tuloksena saatiin 59 kuvatiedostoa eli kuva aukeamaa kohti. Lehti sisältää isoja ja paksuja sanoja, harvaan kirjoitettuja sanoja ja erikoismerkkejä kuten heittomerkkejä, sulkumerkkejä ja kaksoispisteitä.



jolla me nykyisenä ajankohtana elävät ihmiset luulemme voivamme kieltää sen miehen jumaluuden, josta ajanlaskumme alkaa? Eikö meidän loogillisesti katsoen ole joko kerrassaan luovuttava käy tämästä käsitettä Jumala talkka sitten annettava sille jokin täsmällinen sisällys, ts. redusoitava tämä käsite kokemukseen, s is näyt ämiseen (saks. Aufweisung) — jos nimittäin tarkoitamme persoonallista Jumalaa? Tätä reduktiota emme voi täydellisesti suorittaa. — Tuntuu näinollen luonnolliselta — uusien kris inuskon vastustaj en ilmaantuessa ja esiintyessä — tehdä kysymys: mistä syystä tai mikä tarve asianomaisilla on asettua kristinuskoa vastaan? Luonnolliselta näyttää tälloin olevan ajatus tutkia noiden vastus ajien psykolog aa niiden näkökohtien valossa, joihin edellä on viitattu. Heille voita'sin myös selvittää se viisaus, mikä sisältyy Wit gen'ste'nin» (uus'mman filosof sen suunnan, loogisen empirismin syntysanojen lausuja ja etevin edustaja »nerokkaan teoksen 'Trac'atus Logico-Philosophicus' v lmeisten — 7:llä numeroituun — lauseeseen: 'Wovon man nicht sprechen kann, darüber muss man schweigen'. Ja v hdoin on syytä pitää mielessä — niinikään 'loogisen empiristin' — Schlickin, ns. Wiener-Kreisin henkisen isän sanat: 'Brauchen w'r eine Lebensregel, so sei es diese: 'Bewahre den Ge'st der Jugend!' Denn er ist der Sinn des Lebens'.»

(Selostus: sot.past. T. Muukkonen.)

Helsinki 1945. Kirjapaino Aa osakeyhtiö

Ainoastaan palvelukäyttöä varten

Sotilaspappi

N:o 1 | Pp PL:n KENTTÄPIISPANTOIMISTON JULKAISU | 29. 1. 46

Muutamia ajatuksia sotilasillanviettojen ohjelmistosta.

Sotilasillanvietot voitaneen ryhmitellä:
1) hengelliset, 2) isänmaallisvalistukselliset ja 3) ajanvietteelliset virkistystilaisuudet. Seuraavassa eräitä näkökohtia näiden eri tilaisuuksien ohjelmista.

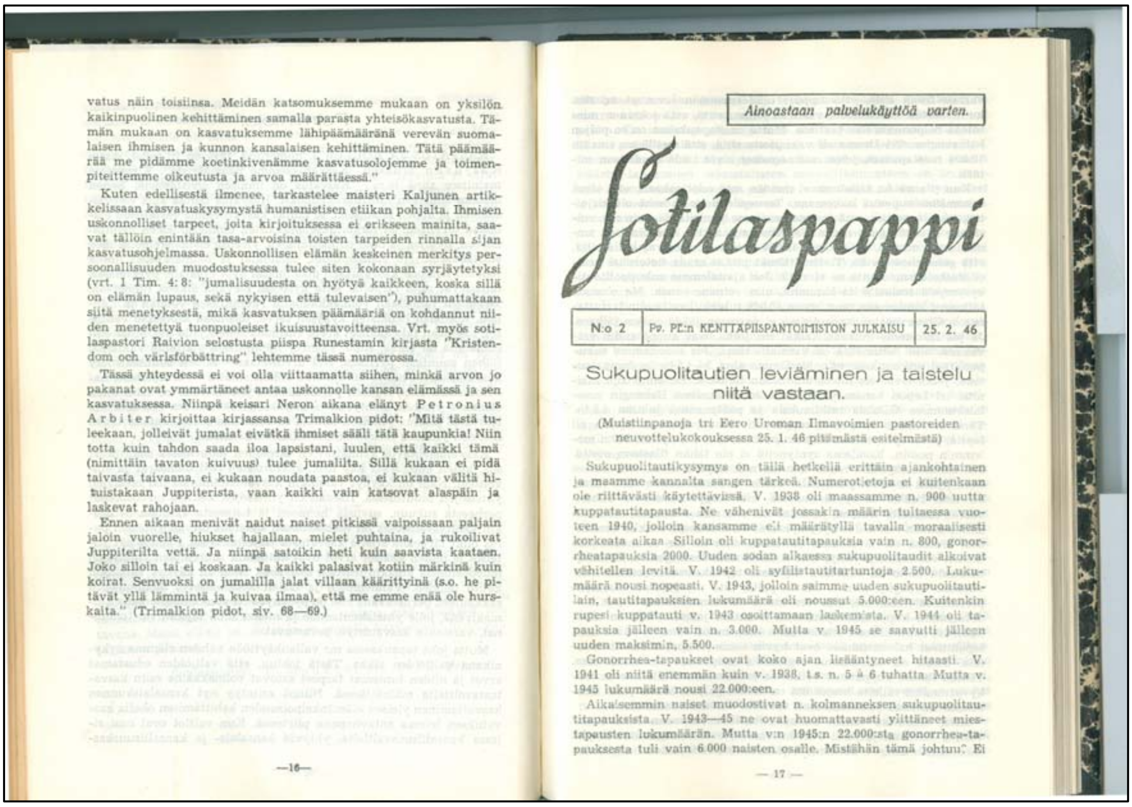
1. Hengelliset juhlatilaisuudet.

Hengellisten illanviettojen tarkoituksena on hengellisen elämän herättäminen ja rakentaminen sekä vahvistaminen kilvoituksen tiellä. Näiden tilaisuuksien ohjelman tulisi siis vastata edellämainittuja päämääriä. Ohjelman pääkohtana tulee näinollen olla sananjulistus. Puheiden on oltava kuten aina, asiallisia, herättäviä, rakentavia ja Kristus-keskeisiä, johon sisältyy myös voimakas sosiaalinen velvoitus. Jokainen puhe olkoon huolella valmistettu, varsinkin, jos toivomme vaativampien kuulijoiden pysyvän mukana harrastuksissamme. Todellinen valmistushan on, niinkuin tiedämme, väsymätön kilvoitus. Vain tätä tietä voidaan odottaa siunauksia. Puhujaksi olisi silloin tällöin saatava myös joku vieraileva sotilas- tai siviilipappi. Erittäin tervetulleen ja rikkaan puhuja-avun voi saada niistä kantahenkilöistä ja varusmiehistä, joiden kohdalla Jumalan armo on tapahtunut ja jotka siitä mielellään todistavat. Tällaisia todistavia-

— 1 —

Kuva 1. Sotilaspappi-aikakausilehden ensimmäinen sivu

Eri sivujen välillä oli ladonnan tasaisuuden vaihteluja, ja se asetti tekstintunnistukselle haasteita. Testauskohteina käytettiin viittä (5) eri aukeamaa, jotta saatiin kattava ja edustava valikoima tekstistä. Testausaineistona 1 oli testatun vuosikerran ensimmäinen aukeama (Kuva 1). Eniten käytössä oli testausaineisto 2 (Kuva 2). Käytössä olivat myös Testausaineistot 3 (Kuva 3), 4 (Kuva 4) ja 5 (kuva 5). Kaikki testiaineistot ovat vuosikerrasta 46.



vatus näin toisiinsa. Meidän katsomuksemme mukaan on yksilön kaikinpuolinen kehittäminen samalla parasta yhteisökasvatusta. Tämän mukaan on kasvatuksemme lähipäämääränä verevän suomalaisen ihmisen ja kunnon kansalaisen kehittäminen. Tätä päämäärää me pidämme koetinkivenämme kasvatusolojemme ja toimenpiteittemme oikeutusta ja arvoa määrättäessä."

Kuten edellisestä ilmenee, tarkastelee maisteri Kaljunen artikkelissaan kasvatuskysymystä humanistisen etiikan pohjalta. Ihmisen uskonnolliset tarpeet, joita kirjoituksessa ei erikseen mainita, saavat tällöin enintään tasa-arvoisina toisten tarpeiden rinnalla sijan kasvatusohjelmassa. Uskonnollisen elämän keskeinen merkitys persoonallisuuden muodostuksessa tulee siten kokonaan syrjäytetyksi (vrt. 1 Tim. 4: 8: "jumalisuudesta on hyötyä kaikkeen, koska sillä on elämän lupaus, sekä nykyisen että tulevaisen"), puhumattakaan siitä menetyksestä, mikä kasvatuksen päämääriä on kohdannut niiden menetettyä tuonpuoleiset ikuisuustavoitteensa. Vrt. myös sotilaspastori Raivion selostusta piispa Runestamin kirjasta "Kristendom och värlsförbättring" lehtemme tässä numerossa.

Tässä yhteydessä ei voi olla viittaamatta siihen, minkä arvon jo pakanat ovat ymmärtäneet antaa uskonnolle kansan elämässä ja sen kasvatuksessa. Niinpä keisari Neron aikana elänyt Petronius Arbiter kirjoittaa kirjassansa Trimalkion pidot: "Mitä tästä tuleekaan, jolleivät jumalat eivätkä ihmiset sääli tätä kaupunkia! Niin totta kuin tahdon saada iloa lapsistani, luulen, että kaikki tämä (nimittäin tavaton kuivuus) tulee jumalilta. Sillä kukaan ei pidä taivasta taivaana, ei kukaan noudata paastoa, ei kukaan välitä hituistakaan Juppiterista, vaan kaikki vain katsovat alaspäin ja laskevat rahojaan.

Ennen aikaan menivät naidut naiset pitkissä vaipoissaan paljain jaloin vuorelle, hiukset hajallaan, mielet puhtaina, ja rukoilivat Juppiterilta vettä. Ja niinpä satoikin heti kuin saavista kaataen. Joko silloin tai ei koskaan. Ja kaikki palasivat kotiin märkinä kuin koirat. Senvuoksi on jumalilla jalat villaan kääriittyinä (s.o. he pitävät yllä lämmintä ja kuivaa ilmaa), että me emme enää ole hurskaita." (Trimalkion pidot, siv. 68—69.)

—16—

Ainoastaan palvelukäyttöä varten.

Sotilaspappi

N:o 2	Pv. PE:n KENTTÄPIISPANTOIMISTON JULKAISU	25. 2. 46

Sukupuolitautien leviäminen ja taistelu niitä vastaan.

(Muistiinpanoja tri Eero Uroman Ilmavoimien pastoreiden neuvottelukokouksessa 25. 1. 46 pitämästä esitelmästä)

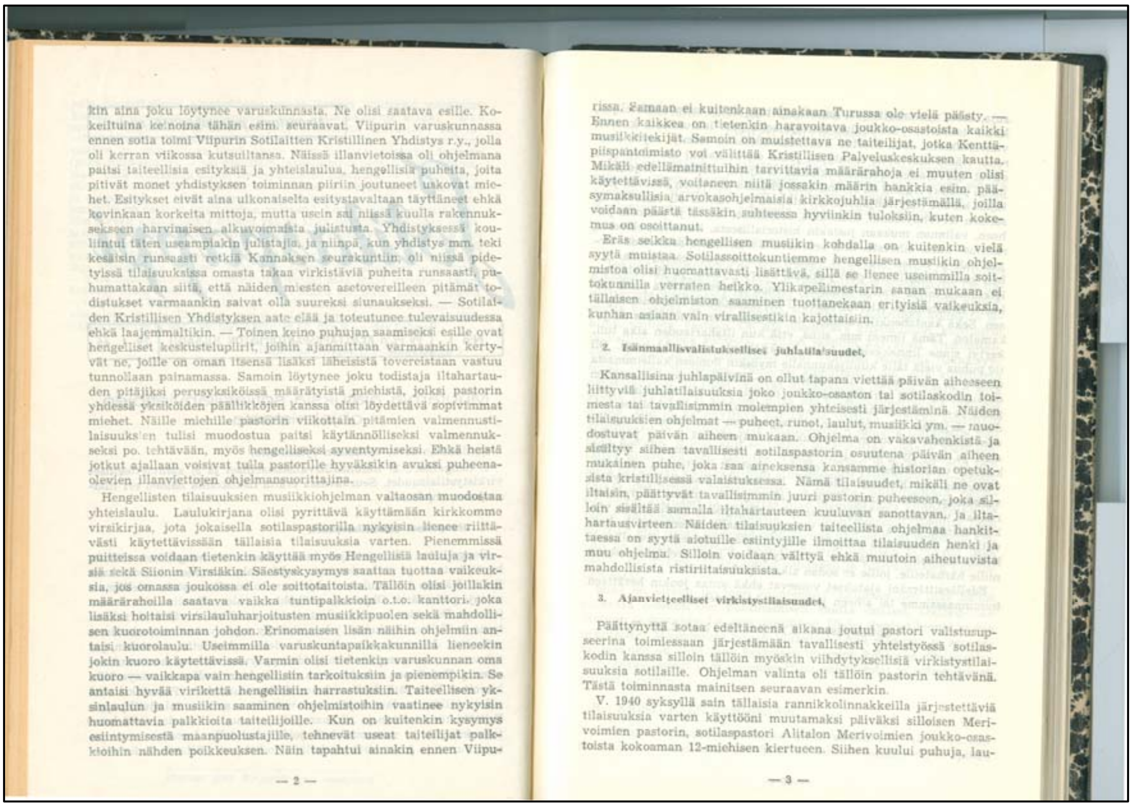
Sukupuolitautikysymys on tällä hetkellä erittäin ajankohtainen ja maamme kannalta sangen tärkeä. Numerotietoja ei kuitenkaan ole riittävästi käytettävissä. V. 1938 oli maassamme n. 900 uutta kuppatautitapausta. Ne vähenivät jossakin määrin tultaessa vuoteen 1940, jolloin kansamme eli määrätyllä tavalla moraalisesti korkeata aikaa. Silloin oli kuppatautitapauksia vain n. 800, gonorrheatapauksia 2000. Uuden sodan alkaessa sukupuolitaudit alkoivat vähitellen levitä. V. 1942 oli syfilistautitartuntoja 2.500. Lukumäärä nousi nopeasti. V. 1943, jolloin saimme uuden sukupuolitautilain, tautitapauksien lukumäärä oli noussut 5.000:een. Kuitenkin rupesi kuppatauti v. 1943 osoittamaan laskemista. V. 1944 oli tapauksia jälleen vain n. 3.000. Mutta v. 1945 se saavutti jälleen uuden maksimin, 5.500.

Gonorrhea-tapaukset ovat koko ajan lisääntyneet hitaasti. V. 1941 oli niitä enemmän kuin v. 1938, t.s. n. 5 à 6 tuhatta. Mutta v. 1945 lukumäärä nousi 22.000:een.

Aikaisemmin naiset muodostivat n. kolmanneksen sukupuolitautitapauksista. V. 1943—45 ne ovat huomattavasti ylittäneet miestapausten lukumäärän. Mutta v:n 1945:n 22.000:sta gonorrhea-tapauksesta tuli vain 6.000 naisten osalle. Mistähän tämä johtuu? Ei

— 17 —

Kuva 2. Testausaineisto 2:n esittely (sivut 16 ja 17, jolla on luvun kaksi alku)



kin aina joku löytynee varuskunnasta. Ne olisi saatava esille. Kokeiltuina keinoina tähän esim. seuraavat. Viipurin varuskunnassa ennen sotia toimi Viipurin Sotilaitten Kristillinen Yhdistys r.y., jolla oli kerran viikossa kutsuiltansa. Näissä illanvietoissa oli ohjelmana paitsi taiteellisia esityksiä ja yhteislaulua, hengellisiä puheita, joita pitivät monet yhdistyksen toiminnan piiriin joutuneet uskovat miehet. Esitykset eivät aina ulkonaiselta esitystavaltaan täyttäneet ehkä kovinkaan korkeita mittoja, mutta usein sai niissä kuulla rakennukselliseen harvinaisen alkuvoimaista julistusta. Yhdistyksessä koulittui täten useampiakin julistajia, ja niinpä, kun yhdistys mm. teki kesäisin runsaasti retkiä Kannaksen seurakuntiin, oli niissä pidetyissä tilaisuuksissa omasta takaa virkistäviä puheita runsaasti, puhumattakaan siitä, että näiden miesten asetovereilleen pitämät todistukset varmaankin saivat olla suureksi siunaukseksi. — Sotilaiden Kristillisen Yhdistyksen aate elää ja toteutunee tulevaisuudessa ehkä laajemmaltikin. — Toinen keino puhujan saamiseksi esille ovat hengelliset keskustelupiirit, joihin ajanmittaan varmaankin kertyvät ne, joille on oman itsensä lisäksi läheisistä tovereistaan vastuu tunnollaan painamassa. Samoin löytynee joku todistaja iltahartauden pitäjiksi perusyksiköissä määrätyistä miehistä, joiksi pastorin yhdessä yksiköiden päällikköjen kanssa olisi löydettävä sopivimmat miehet. Näille miehille pastorin viikottain pitämien valmennustilaisuuksien tulisi muodostua paitsi käytännölliseksi valmennukseksi po. tehtävään, myös hengelliseksi syventymiseksi. Ehkä heistä jotkut ajallaan voisivat tulla pastorille hyväksikin avuksi puheena-olevien illanviettojen ohjelmansuorittajina.

Hengellisten tilaisuuksien musiikkiohjelman valtaosan muodostaa yhteislaulu. Laulukirjana olisi pyrittävä käyttämään kirkkomme virsikirjaa, jota jokaisella sotilaspastorilla nykyisin lienee riittävästi käytettävissään tällaisia tilaisuuksia varten. Pienemmissä puitteissa voidaan tietenkin käyttää myös Hengellisiä lauluja ja virsiä sekä Siionin Virsiäkin. Säestyskysymys saattaa tuottaa vaikeuksia, jos omassa joukossa ei ole soittotaitoista. Tällöin olisi joillakin määrärahoilla saatava vaikka tuntipalkkioin o.t.o. kanttori, joka lisäksi hoitaisi virsilauluharjoitusten musiikkipuolen sekä mahdollisen kuorotoiminnan johdon. Erinomaisen lisän näihin ohjelmiin antaisi kuorolaulu. Useimmilla varuskuntapaikkakunnilla lieneekin jokin kuoro käytettävissä. Varmin olisi tietenkin varuskunnan oma kuoro — vaikkapa vain hengellisiin tarkoituksiin ja pienempikin. Se antaisi hyvää virikettä hengellisiin harrastuksiin. Taiteellisen yksinlaulun ja musiikin saaminen ohjelmistoihin vaatinee nykyisin huomattavia palkkioita taiteilijoille. Kun on kuitenkin kysymys esiintymisestä maanpuolustajille, tehnevät useat taiteilijat palkkioihin nähden poikkeuksen. Näin tapahtui ainakin ennen Viipu-

— 2 —

rissa. Samaan ei kuitenkaan ainakaan Turussa ole vielä päästy. — Ennen kaikkea on tietenkin haravoitava joukko-osastoista kaikki musiikkitekijät. Samoin on muistettava ne taiteilijat, jotka Kenttäpiispantoimisto voi välittää Kristillisen Palveluskeskuksen kautta. Mikäli edellämainittuihin tarvittavia määrärahoja ei muuten olisi käytettävissä, voitaneen niitä jossakin määrin hankkia esim. pääsymaksullisia arvokasohjelmaisia kirkkojuhlia järjestämällä, joilla voidaan päästä tässäkin suhteessa hyviinkin tuloksiin, kuten kokemus on osoittanut.

Eräs seikka hengellisen musiikin kohdalla on kuitenkin vielä syytä muistaa. Sotilassoittokuntiemme hengellisen musiikin ohjelmistoa olisi huomattavasti lisättävä, sillä se lienee useimmilla soittokunnilla verraten heikko. Ylikapellimestarin sanan mukaan ei tällaisen ohjelmiston saaminen tuottanekaan erityisiä vaikeuksia, kunhan asiaan vain virallisestikin kajottaisiin.

2. Isänmaallisvalistukselliset juhlatilaisuudet,

Kansallisina juhlapäivinä on ollut tapana viettää päivän aiheeseen liittyviä juhlatilaisuuksia joko joukko-osaston tai sotilaskodin toimesta tai tavallisimmin molempien yhteisesti järjestäminä. Näiden tilaisuuksien ohjelmat — puheet, runot, laulut, musiikki ym. — muodostuvat päivän aiheen mukaan. Ohjelma on vakavahenkistä ja sisältyy siihen tavallisesti sotilaspastorin osuutena päivän aiheen mukainen puhe, joka saa aineksensa kansamme historian opetuksista kristillisessä valaistuksessa. Nämä tilaisuudet, mikäli ne ovat iltaisin, päättyvät tavallisimmin juuri pastorin puheeseen, joka silloin sisältää samalla iltahartauteen kuuluvan sanottavan, ja iltahartausvirteen. Näiden tilaisuuksien taiteellista ohjelmaa hankittaessa on syytä aiotuille esiintyjille ilmoittaa tilaisuuden henki ja muu ohjelma. Silloin voidaan välttyä ehkä muutoin aiheutuvista mahdollisista ristiriitaisuuksista.

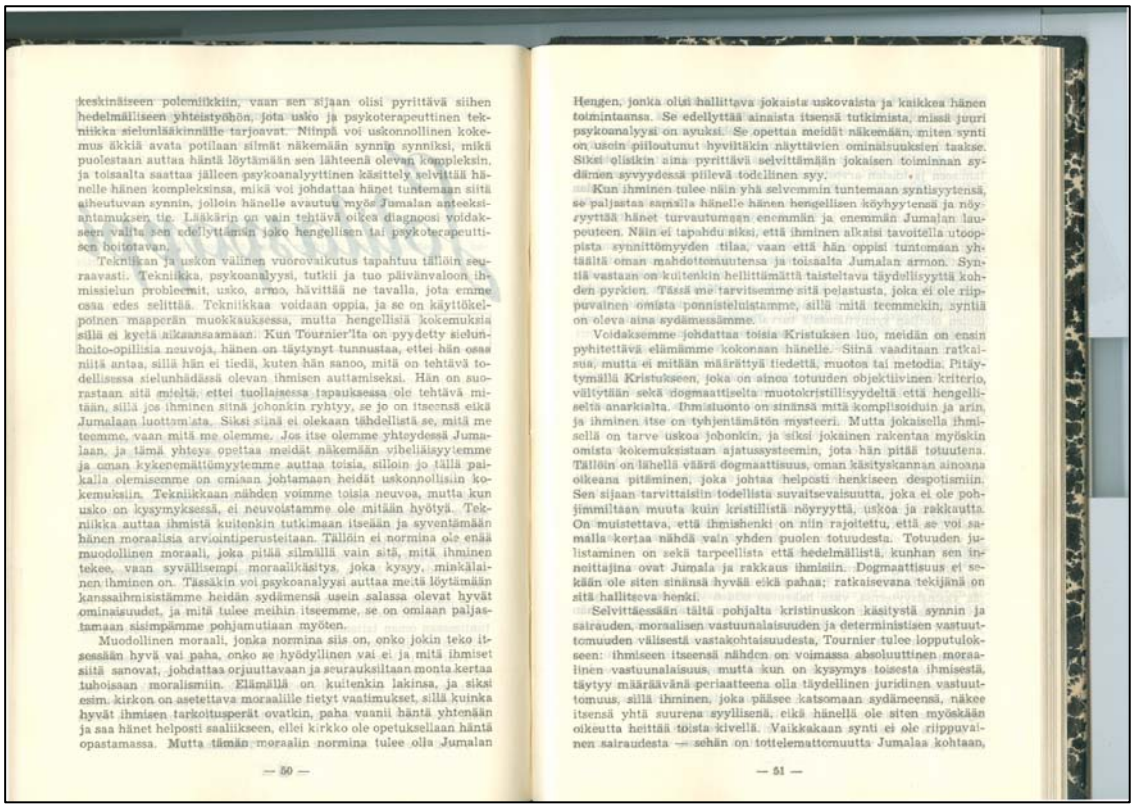
3. Ajanvietteelliset virkistystilaisuudet,

Päättynyttä sotaa edeltäneenä aikana joutui pastori valistusupseerina toimiessaan järjestämään tavallisesti yhteistyössä sotilaskodin kanssa silloin tällöin myöskin viihdytyksellisiä virkistystilaisuuksia sotilaille. Ohjelman valinta oli tällöin pastorin tehtävänä. Tästä toiminnasta mainitsen seuraavan esimerkin.

V. 1940 syksyllä sain tällaisia rannikkolinnakkeilla järjestettäviä tilaisuuksia varten käyttööni muutamaksi päiväksi silloisen Merivoimien pastorin, sotilaspastori Alitalon Merivoimien joukko-osastoista kokoaman 12-miehisen kiertueen. Siihen kuului puhuja, lau-

— 3 —

Kuva 3. Testauskohteen 3 esittely (sivut 2 ja 3 luvusta 1)



keskinäiseen polemiikkiin, vaan sen sijaan olisi pyrittävä siihen hedelmälliseen yhteistyöhön, jota usko ja psykoterapeuttinen tekniikka sielunlääkinnälle tarjoavat. Niinpä voi uskonnollinen kokemus äkkiä avata potilaan silmät näkemään synnin synniksi, mikä puolestaan auttaa häntä löytämään sen lähteenä olevan kompleksin, ja toisaalta saattaa jälleen psykoanalyyttinen käsittely selvittää hänelle hänen kompleksinsa, mikä voi johdattaa hänet tuntemaan siitä aiheutuvan synnin, jolloin hänelle avautuu myös Jumalan anteeksiahtamuksen tie. Lääkärin on vain tehtävä oikea diagnoosi voidakseen valita sen edellyttämän joko hengellisen tai psykoterapeuttisen hoitotavan.

Tekniikan ja uskon välinen vuorovaikutus tapahtuu tälloin seuraavasti. Tekniikka, psykoanalyysi, tutkii ja tuo päivänvaloon ihmissielun probleemit, usko, armo, hävittää ne tavalla, jota emme osaa edes selittää. Tekniikkaa voidaan oppia, ja se on käyttökelpoinen maaperän muokkauksessa, mutta hengellisiä kokemuksia sillä ei kyetä aikaansaamaan. Kun Tournier'lta on pyydetty sielunhoito-opilisia neuvoja, hänen on täytynyt tunnustaa, ettei hän osaa niitä antaa, sillä hän ei tiedä, kuten hän sanoo, mitä on tehtävä todellisessa sielunhädässä olevan ihmisen auttamiseksi. Hän on suorastaan sitä mieltä, ettei tuollaisessa tapauksessa ole tehtävä mitään, sillä jos ihminen siinä johonkin ryhtyy, se jo on itsensä eikä Jumalaan luottamista. Siksi siinä ei olekaan tähdellistä se, mitä me teemme, vaan mitä me olemme. Jos itse olemme yhteydessä Jumalaan, ja tämä yhteys opettaa meidät näkemään vihelläisyytemme ja oman kykenemättömyytemme auttaa toisia, silloin jo tällä paikalla olemisemme on omiaan johtamaan heidät uskonnollisiin kokemuksiin. Tekniikkaan nähden voimme toisia neuvoa, mutta kun usko on kysymyksessä, ei neuvoistamme ole mitään hyötyä. Tekniikka auttaa ihmistä kuitenkin tutkimaan itseään ja syventämään hänen moraalisia arviointiperusteitaan. Tällöin ei normina ole enää muodollinen moraali, joka pitää silmällä vain sitä, mitä ihminen tekee, vaan syvällisempi moraalikäsitys, joka kysyy, minkälainen ihminen on. Tässäkin voi psykoanalyysi auttaa meitä löytämään kanssaihmisistämme heidän sydämensä usein salassa olevat hyvät ominaisuudet, ja mitä tulee meihin itseemme, se on omiaan paljastamaan sisimpämme pohjamutiaan myöten.

Muodollinen moraali, jonka normina siis on, onko jokin teko itsessään hyvä vai paha, onko se hyödyllinen vai ei ja mitä ihmiset siitä sanovat, johdattaa orjuuttavaan ja seurauksiltaan monta kertaa tuhoisaan moralismiin. Elämällä on kuitenkin lakinsa, ja siksi esim. kirkon on asetettava moraalille tietyt vaatimukset, sillä kuinka hyvät ihmisen tarkoitusperät ovatkin, paha vaanii häntä yhtenään ja saa hänet helposti saaliikseen, ellei kirkko ole opetuksellaan häntä opastamassa. Mutta tämän moraalin normina tulee olla Jumalan

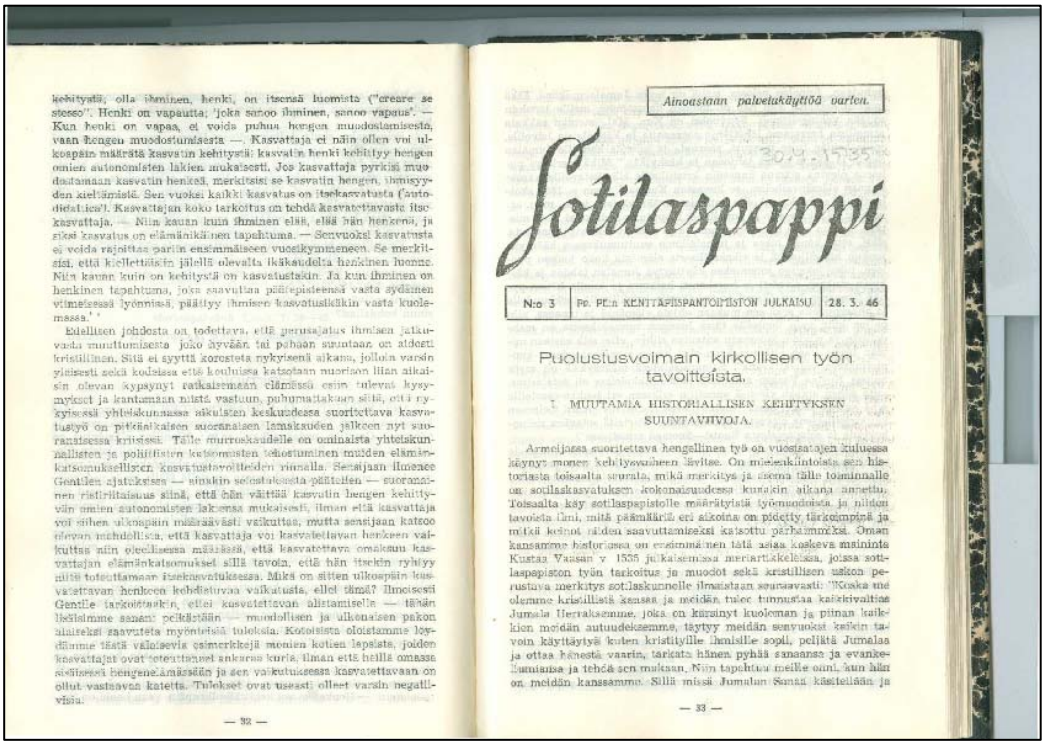
Hengen, jonka olisi hallittava jokaista uskovaista ja kaikkea hänen toimintaansa. Se edellyttää ainaista itsensä tutkimista, missä juuri psykoanalyysi on ayuksi. Se opettaa meidät näkemään, miten synti on usein piiloutunut hyviltäkin näyttävien ominaisuuksien taakse. Siksi olisikin aina pyrittävä selvittämään jokaisen toiminnan sydämen syvyydessä piilevä todellinen syy.

Kun ihminen tulee näin yhä selvemmin tuntemaan syntisyytensä, se paljastaa samalla hänelle hänen hengellisen köyhyytensä ja nöyryyttää hänet turvautumaan enemmän ja enemmän Jumalan lauppuuteen. Näin ei tapahdu siksi, että ihminen alkaisi tavoitella utoopista synnittömyyden tilaa, vaan että hän oppisi tuntomaan yhtäältä oman mahdottomuutensa ja toisaalta Jumalan armon. Syntiä vastaan on kuitenkin hellittämättä taisteltava täydellisyyttä kohden pyrkien. Tässä me tarvitsemme sitä pelastusta, joka ei ole riippuvainen omista ponnistelnistamme, sillä mitä teemmekin, syntiä on oleva aina sydämessämme.

Voidaksemme johdattaa toisia Kristuksen luo, meidän on ensin pyhitettävä elämämme kokonaan hänelle. Siinä vaaditaan ratkaisua, mutta ei mitään määrättyä tiedettä, muotoa tai metodia. Pitäytymällä Kristukseen, joka on ainoa totuuden objektiivinen kriterio, välttyään sekä dogmaattiselta muotokristillisyydeltä että hengelliseltä anarkialta. Ihmistuonto on sinänsä mitä kompliscoiduin ja arin, ja ihminen itse on tyhjentämätön mysteeri. Mutta jokaisella ihmisellä on tarve uskoa johonkin, ja siksi jokainen rakentaa myöskin omista kokemuksistaan ajatussysteemin, jota hän pitää totuutena. Tälloin on lähellä väärä dogmaattisuus, oman käsityskannan ainoana oikeana pitäminen, joka johtaa helposti henkiseen despotismiin. Sen sijaan tarvittaisiin todellista suvaitsevaisuutta, joka ei ole pohjimmiltaan muuta kuin kristillistä nöyryyttä, uskoa ja rakkautta. On muistettava, että ihmishenki on niin rajoitettu, että se voi samalla kertaa nähdä vain yhden puolen totuudesta. Totuuden julistaminen on sekä tarpeellista että hedelmällistä, kunhan sen inneittajina ovat Jumala ja rakkaus ihmisiin. Dogmaattisuus ei sekään ole siten sinänsä hyvää eikä pahaa; ratkaisevana tekijänä on sitä hallitseva henki.

Selvittäessään tältä pohjalta kristinuskon käsitystä synnin ja sairauden, moraalisen vastuunalaisuuden ja deterministisen vastuuttomuuden välisestä vastakohtaisuudesta, Tournier tulee lopputulokseen: ihmiseen itseensä nähden on voimassa absoluuttinen moraalinen vastuunalaisuus, mutta kun on kysymys toisesta ihmisestä, täytyy määräävänä periaatteena olla täydellinen juridinen vastuuttomuus, sillä ihminen, joka pääsee katsomaan sydämeensä, näkee itsensä yhtä suurena syyllisenä, eikä hänellä ole siten myöskään oikeutta heittää toista kivellä. Vaikkakaan synti ei ole riippuvainen sairaudesta — sehän on tottelemattomuutta Jumalaa kohtaan,

Kuva 4. Testiaineisto 4:n esittely (sivut 50 ja 51 luvusta 4-5)



kehitystä, olla ihminen, henki, on itsensä luomista ("creare se stesso"). Henki on vapautta; 'joka sanoo ihminen, sanoo vapaus'. — Kun henki on vapaa, ei voida puhua hengen muodostumisesta, vaan hengen muodostumisesta. — Kasvattaja ei näin ollen voi ulkopuolelta määrätä kasvatin kehitystä: kasvatin henki kehittyy hengen omien autonomisten lakien mukaisesti. Jos kasvattaja pyrkisi muodostamaan kasvatin henkeä, merkitsisi se kasvatin hengen, ihmisyyden kieltämistä. Sen vuoksi kaikki kasvatus on itsekasvatusta ('autodidaktica'). Kasvattajan koko tarkoitus on tehdä kasvatettavasta itsekasvattaja. — Niin kauan kuin ihminen elää, elää hän henkenä, ja siksi kasvatus on elämänikäinen tapahtuma. — Semvuoksi kasvatusta ei voida rajoittaa parin ensimmäiseen vuosikymmeneen. Se merkitsisi, että kielletttäisiin jälellä olevalta ikäkaudelta henkinen luonne. Niin kauan kuin on kehitystä on kasvatustakin. Ja kun ihminen on henkinen tapahtuma, joka saavuttaa päätepisteensä vasta sydämen viimeisessä lyönnissä, päättyy ihmisen kasvatustikkin vasta kuolemassa.'

Edellisen johdosta on todettava, että perusajatus ihmisen jatkuvasta muuttumisesta joko äyvään tai pahaan suuntaan on aidosti kristillinen. Sitä ei syyttä korosteta nykyisenä aikana, jolloin varsin yleisesti sekä kodeissa että kouluissa katsotaan nuorison liian aikaisin olevan kypsynyt ratkaisemaan elämässä esim. tulevat kysymykset ja kantamaan niistä vastuun, puhumattakaan siitä, että nykyisessä yhteiskunnassa aikuisten keskuudessa suoritettava kasvatustyö on pitkäaikaisen suoranaisen lamaskauden jälkeen nyt suoranaisessa kriisissä. Tälle murroskaudelle on ominaista yhteiskunnallisten ja poliittisten katsomusten tehostuminen muiden elämänkatsomuksellisten kasvatustavoitteiden rinnalla. Senaijaan Ilmenee Gentilen ajatuksissa — sinäkin selostuksesta päätellen — suoranainen ristiriitaisuus siinä, että hän väittää kasvatin hengen kehittyvän omien autonomisten lakiensa mukaisesti, ilman että kasvattaja voi siihen ulkopuolelta määräävästi vaikuttaa, mutta seasijaan katsoo olevan mahdollista, että kasvattaja voi kasvatettavan henkeen vaikuttaa niin objektiivisessa määrässä, että kasvatettava omaksuu kasvattajan elämänkatsomukset sillä tavoin, että hän itsekin ryhtyy niitä toteuttamaan itsekasvatuksessa. Mikä on sitten ulkopuolta kasvatettavan henkeen kohdistuvaa vaikutusta, ellei tämä? Ilmeisesti Gentile tarkoittaakin, ettei kasvatettavan alistamiselle — tämän lieisimme sanani pelkistään — muodollisen ja ulkonaisen pakon alaiseksi saavuteta myöntäisiä tuloksia. Kotoista oloistamme löydämme tästä vakuuttavia esimerkkejä monien kotien lapsista, joiden kasvattajat ovat 'ostautuneet ankaras kurissa, ilman että heillä omassa sisäisessä hengenelämässään ja sen vaikutuksessa kasvatettavaan on ollut vastaavan katetta. Tulokset ovat useasti olleet varsin negatiivisia.

Ainoastaan palvelukäyttöä varten.

Sotilaspappi

N:o 3 | Pp. PL:n KENTTÄPIISPANTOIMISTON JULKAISU | 28. 3. 46

Puolustusvoimain kirkollisen työn tavoitteista.

I. MUUTAMIA HISTORIALLISEN KEHITYKSEN SUUNTAVIIVOJA.

Armeijassa suoritettava hengellinen työ on vuosisatojen kuluessa käynyt monen kehitysvaiheen läpitse. On mielenkiintoista seurata historiasta toisaalta sitä, mikä merkitys ja osuus tällä toiminnalle on sotilaskasvatuksen kokonaisuudessa kussakin aikana annettu. Toisaalta käy sotilaspapistolle määrätyistä työnuodoista, ja niiden tavoista ilmi, mitä päämääriä eri aikoina on pidetty tärkeimpinä ja mitkä kehot näiden saavuttamiseksi katsottu parhaimmiksi. Oman kansamme historiassa on ensimmäinen tätä asiaa koskeva maininta Kustaa Vaasan v. 1535 julkaisemassa mertiartikkelikirja, jolssa sotilaspapiston työn tarkoitus ja muodot sekä kristillisen uskon perustava merkitys sotilaskunnolle ilmaistaan seuraavasti: 'Koska me olemme kristillistä kansaa ja meidän tulee turvautua kaikkivaltiaan Jumala Herraksemme, joka on kärsinyt kuoleman ja piinan kaikkien meidän autuudeksemme, täytyy meidän sentuoksi kaikein tavoin käyttäytyä kuten kristityille ihmisille sopii, pelätä Jumalaa ja ottaa hänestä vaarin, tärkeitä hänen pyhää sanaansa ja evankeliumiansa ja tehdä sen mukaan. Niin tapahtuu meille onni, kun hän on meidän kanssamme. Sillä missä Jumalan Sanaa käsitellään ja

Kuva 5. Testiaineisto 5:n esittely (sivut 32 ja 33, jolla on luvun kolme alku)

Tasaisten lopputulosten saavuttamiseksi testiaineisto 2 oli eniten testattu kohde, mutta vaihtelevuuden ja kattavuuden vuoksi testiaineistot 1, 3, 4 ja 5 olivat

myös käytössä. Kuten kuvista on nähtävissä, oman haasteen tekee myös sivujen kääntöpuolen tekstin näkyminen paperin läpi.

4 Digitointityön suoritus

Digitointi- eli digitalisointityö suoritettiin skannaamalla nidottu lehti, jonka jälkeen skannatut kuvat käsiteltiin testattavilla tekstintunnistusohjelmilla. Ohjelmia käytettiin kahdenlaisia. Ensin testattiin selainpohjaisia ohjelmia, ja niiden jälkeen työasemalle asennettavia ohjelmia.

4.1 Asiakirjojen kuvaustapa

Asiakirjojen kuvaamiseen löytyy kaksi vaihtoehtoista tapaa: skanneri ja digikamera, eli kamera, joka tallentaa kuvattavan kohteen digitaaliseksi. Tavallisesti skannerit eivät välttämättä sovellu paksujen ja nidottujen teosten skannaamiseen, koska se voi vahingoittaa aineistoa. Sivut voivat irrota tai teksti suttaantua. Digikameralla kuvattuna puolestaan mittasuhteet ja linjat voivat vääristyä. Skannauskynää ei kokeiltu, koska aineistoa oli satoja sivuja ja skannauskynällä skannataan vain rivi kerrallaan.

”Kamera on skanneria sopivampi laite digitointiin silloin, kun esimerkiksi kirjaa ei voida avata levälleen tai digitoitava materiaali on haurasta ja vaarassa vahingoittua. Kameraa käytettäessä ei laitteen ja kopioitavan materiaalin välillä ole fyysistä kontaktia, joka saattaisi vahingoittaa alkuperäistä aineistoa. ”

”Valokuvaaminen digitaalikameralla saattaa olla ainoa mahdollisuus digitoida materiaalia, jota ei voida muilla keinoin käsitellä esimerkiksi aineiston vahingoittumattomuuden varmistamiseksi. Sen käytössä on omat riskinsä. Esimerkiksi kameran objektiivi saattaa vääristää kohteen mittasuhteita ja linjoja, tai tiedoston resoluutio ja tiedostomuoto saattavat poiketa siitä, mikä on organisaation standardeissa.

Digitaalikameran käyttökelpoisuus digitalisoinnissa onkin harkittava huolella.”

/1/

4.2 Skannaus

Skannaukseen valittiin HP Scanjet G4010 -valokuvaskanneri. Asetuksissa täytyi muistaa asettaa TWAIN (Technology Without An Interesting Name) päälle ohjeiden mukaisesti. TWAIN on standardi, joka varmistaa eri laitteiden ja ohjelmistojen yhteensopivuuden sekä mahdollistaa skannaamisen kuvankäsittelyohjelmasta. Skannaus tehtiin väriskannauksella.

Yksi aukeama skannattiin kerrallaan. Skannatut tiedostot tallennettiin tiedostomuotoon TIFF (Tagged image file format), joka tunnetaan myös lyhenteellä TIF. TIFF on kuvien tallennukseen käytetty tiedostomuoto. Tiedostot nimettiin sivunumeroidensa mukaan kolmella numerolla, esimerkiksi 001, 014 ja 127, etteivät sivut tulostu kuvia yhdistettäessä väärässä järjestyksessä. Tiedoston nimeen merkittiin myös vuosikerta sekä luku, jotta tiedostot eivät olisi menneet sekaisin. Kun tiedosto oli käsitelty OCR-ohjelmalla, käsitellyn kuvan nimeen lisättiin ohjelman nimi, jotta testatut tiedostot eivät menisi sekaisin eri ohjelmien välillä. 144-sivuisesta vuosikerrasta 46 kuvatiedostoja tulostui yhteensä 58 kappaletta

5 Testatut ohjelmat

Työssä testattiin kahdenlaisia ohjelmia: niitä, jotka toimivat Internet-selaimesta käsin, ja niitä, jotka vaativat asennuksen työasemalle. Tarkoituksena oli muuttaa skannattu kuvatiedosto joko DOC-tiedostoksi (Document) - joka on Microsoftin käyttämä tekstitiedostomuoto - tai PDF-tiedostoksi tai molemmiksi ja katsoa, miltä lopputulos näyttää. Muutoksen valinta riippui ohjelmasta ja siitä, mitä mahdollisuuksia ohjelma antaa tunnistusformaatin suhteen.

5.1 Selaimessa toimivat OCR-ohjelmat

Ensimmäisenä testattiin, kuinka Internet-selaimesta käsin toimivat tekstintunnistusohjelmat toimivat. Selain on ohjelma, jolla käyttäjä voi katsella ja lähettää kuvia, tekstiä ja muuta, mitä WWW-sivuilta (World Wide Web eli maailmanlaajuinen verkko) löytyy. Testattavia ohjelmia kokeiltiin kahdeksan (8) kappaletta ja tutkittiin, kuinka hyvä lopputulos niillä saadaan. Ohjelmia oli kahdenlaisia: niitä, jotka toimivat rajoituksetta, ja niitä, jotka muuttuivat 2 – 20 käsitellyn tiedoston jälkeen maksullisiksi. Ohjelmissa oli eroavaisuuksia tuettujen kielten suhteen (Taulukko 1). Jotkut ohjelmat tukivat monia kieliä, ja jotkut tunnistivat vain englanninkielestä löytyviä merkkejä ja sanoja.

Ohjelmat erosivat siinä, mitä kuvaformaattia ne tukevat (Taulukko 3) ja mihin tiedostomuotoon käsitelty tiedosto tulostuu (Taulukko 4). Tuettuja kuvaformaatteja oli runsaasti: PDF, GIF (Graphic Interchange Format), TIFF, JPEG (Photographic Experts Group) ja PNG (Portable Network Graphics), jotka ovat bittikarttagrafiikan tallennusformaatteja. BMP (Bit Map Picture) on bittikarttakuva. PGM (Portable Greymap Format) ja PPM (Portable Pixmap Format) ovat avoimen lähdekoodin kuvaformaatteja. PCX (Personal Computer eXchange) on kuvaformaatti, joka oli ensimmäinen hyväksytty DOS-kuvastandardi (Disc Operating System). DOS on yhden käyttäjän komentorivipohjainen ei-moniajava käyttöjärjestelmä.

Tiedostomuotoja, jotka voidaan valita käsittelyn jälkeiseksi formaatiksi, oli muutamia: DOC, PDF, TXT (Text file) eli tekstitiedosto, RTF (Rich Text Format), joka on muotoillun tekstin tallennusmuoto sekä XML (eXtensible Markup Lan-

guage), joka on dokumenttien tallennusformaatti ja jota käytetään järjestelmien väliseen tiedonvälitykseen. Lopuksi on XLS (Excel Spreadsheet), joka on Microsoftin Excel-taulukkolaskentaohjelman käyttämä tiedostomuoto.

Ohjelma	suomi	englanti	ruotsi	muut kielet yht.
Free OCR	x	x	x	26
Free Online OCR		x		
i2OCR	x	x	x	30
New ocr	x	x	x	26
OCR Convert		x		4
OCR Now!	x	x	x	13
Ocr online	x	x	x	150
OCR terminal	x	x	x	150

Taulukko 1. Selainpohjaisten tekstintunnistusohjelmien tunnistamat kielet

Käsittelyissä käytettiin neljää (4) eri testiaineistoa (Taulukko 2), jotta saataisiin erilaisia näkökulmia ja vaihtelevuutta. Kielten tukeminen oli ohjelmissa vaihtelevaa ja OCR Online – ohjelmassa joutui asetuksista lisäämään suomen kielilistaan, koska listassa on aluksi vain 15 kieltä, joita pystyy asetuksista valitsemalla helposti muokkaamaan. Yhteensä kielivaihtoehtoja on 153.

Ohjelma	Testiaineisto 1	Testiaineisto 2	Testiaineisto 3	Testiaineisto 4
Free OCR	x			
Free Online OCR		x		
i2OCR		x		
New ocr	x	x		
OCR Convert		x		
OCR Now!		x		
Ocr online			x	
OCR terminal				x

Taulukko 2. Ohjelmissa käytetyt testiaineistot

Ohjelmat	PDF	GIF	TIFF	JPEG	BMP	PNG	PGM	PPM	PCX
Free OCR	x	x	x	x	x				
Free Online OCR	x	x	x	x	x	x			
i2OCR		x	x	x	x	x	x	x	
New ocr		x	x	x	x	x			
OCR Convert	x	x		x	x				
OCR Now!	x	x	x	x		x			
Ocr online	x	x	x	x		x			
OCR terminal		x	x	x	x	x			x

Taulukko 3. Selainpohjaisten tekstintunnistusohjelmien käsiteltävät kuvaformaatit

Ohjelma	DOC	PDF	TXT	RTF	XML	XLS
Free OCR	x	x	x	x		
Free Online OCR	x	x	x	x		
i2OCR	x	x		x	x	
New ocr			x			
OCR Convert			x			
OCR Now!	x				x	x
Ocr online	x	x	x	x		
OCR terminal	x	x	x	x		x

Taulukko 4. Selainpohjaisten tekstintunnistusohjelmien tulostettavat tiedostomuodot

Ohjelmien rajoitukset (Taulukko 5) vaihtelivat ilmaisuuden ja hintatason välillä.

Ohjelma	Ei rajoituksia	Hinta	Kokeilu-aika	Tiedostorajoitus	Ladattavan tiedoston koko (MB)
Free OCR					2
Free Online OCR	x				
i2OCR	x				
New OCR					5
OCR Convert	x				
OCR Now!		1,99 puntaa		2	
Ocr online		7,99 dollaria		10/viikko	
OCR terminal		5 dollaria		20/kuukausi	

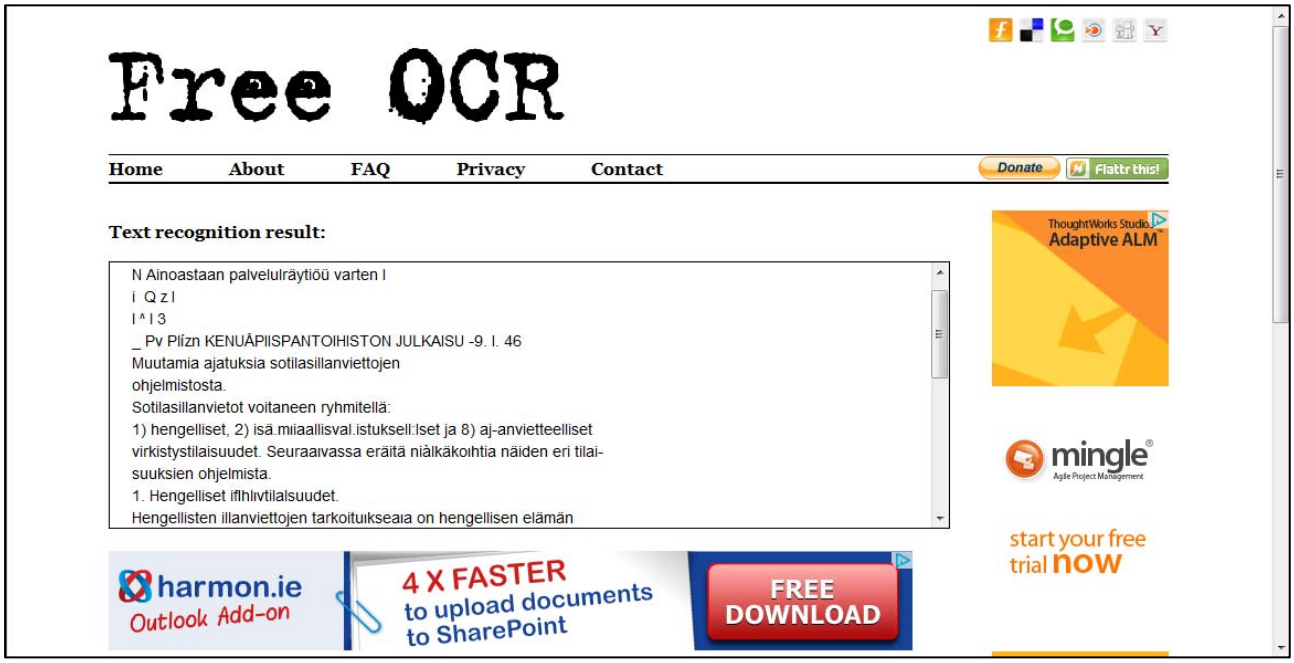
Taulukko 5. Selainpohjaisten tekstintunnistusohjelmien rajoitukset

Hinta on minimihinta, joka täytyy maksaa, kun tiedostorajoitus täyttyy. Minimihinta tarkoittaa sitä hintaa, jolla saa ostettua krediittejä (Credit), eli sivustoilla käytettävää valuuttaa. Ostettavien krediittien tai sivujen määrä vaihteli. OCR Now! -ohjelmassa 1,99 punnalla sai 10 krediittiä, OCR Online -ohjelmassa 7,99 dollarilla sai 200 sivua, joka tarkoittaa 200:aa kuvatiedostoa. OCR Terminal -ohjelmassa 50 sivua maksoi 5 dollaria.

5.1.1 Free OCR -ohjelma

Tuetut kuvaformaatit näkyvät Taulukossa 3. Kieleksi voi valita suomen, englannin ja ruotsin lisäksi 26 kieltä, joten myös ä- ja ö-kirjaimet löytyvät. Käytön validoimiseksi, eli sen varmistamiseksi, täyttääkö jokin asia tietyt vaatimukset, täytyy vain kirjoittaa ruudulla näkyvä merkkijono.

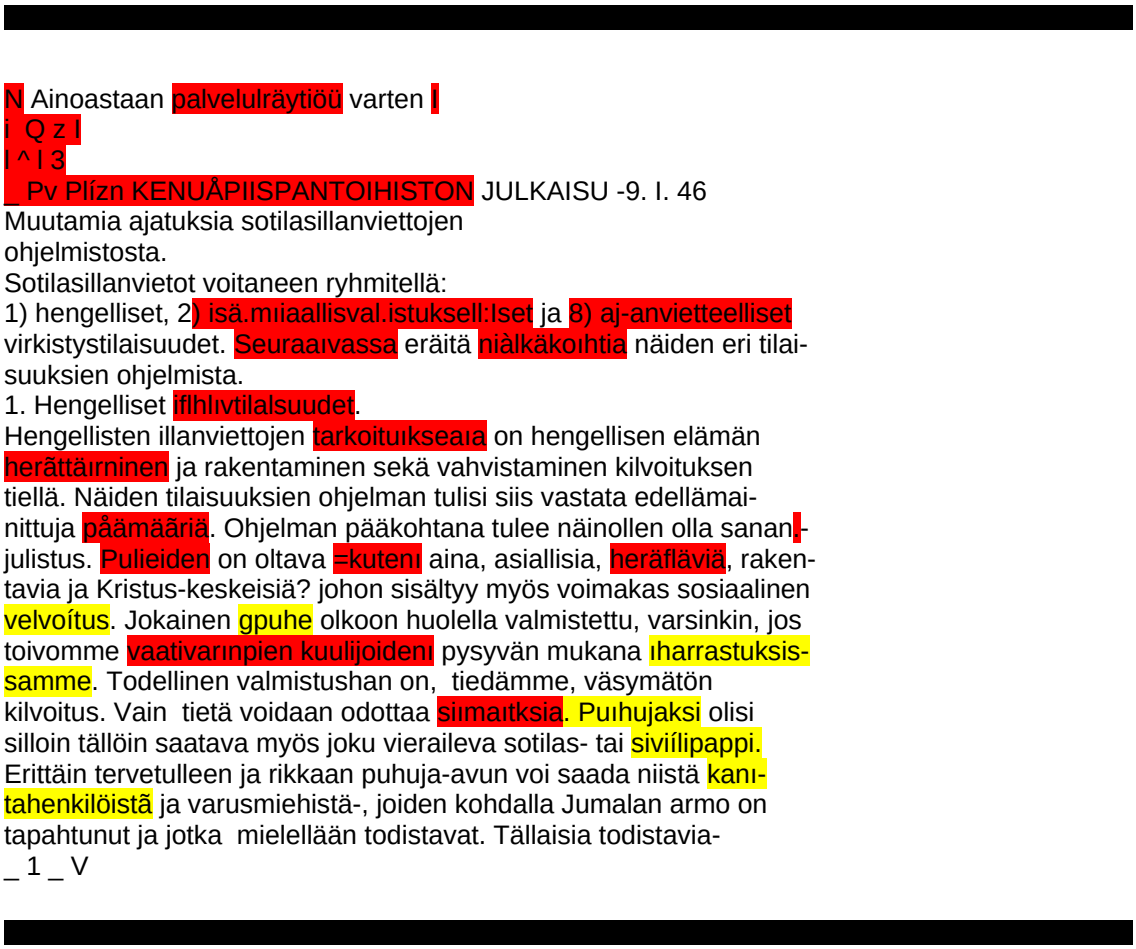
Toisin kuin moni muu ohjelma, Free OCR /2/ ei siirrä tekstiä kuvineen Wordiin (Microsoftin kehittämä tekstinkäsittelyohjelma), vaan jättää tekstin ilman kuvia selaimeen vapaaseen käyttöön (Kuva 6), josta ne voi siirtää haluamaansa ohjelmaan.



Kuva 6. Free OCR -ohjelman tekstintunnistusnäyte

Seuraavanlaisesti tulostui käännös (Kuva 7) testiaineistosta 1 (Taulukko 2); Punaisella ovat virheet, joissa ohjelma on jättänyt sanasta tunnistamatta tai

tunnistanut väärin enemmän kuin yhden merkin. Keltaisella ovat sanat, joissa ohjelma on tunnistanut sanasta tai tunnistanut väärin vain yhden merkin. Punaisella merkittyjä sanoja löytyi 20 kappaletta. Keltaisella merkittyjä sanoja löytyi 6 kappaletta.

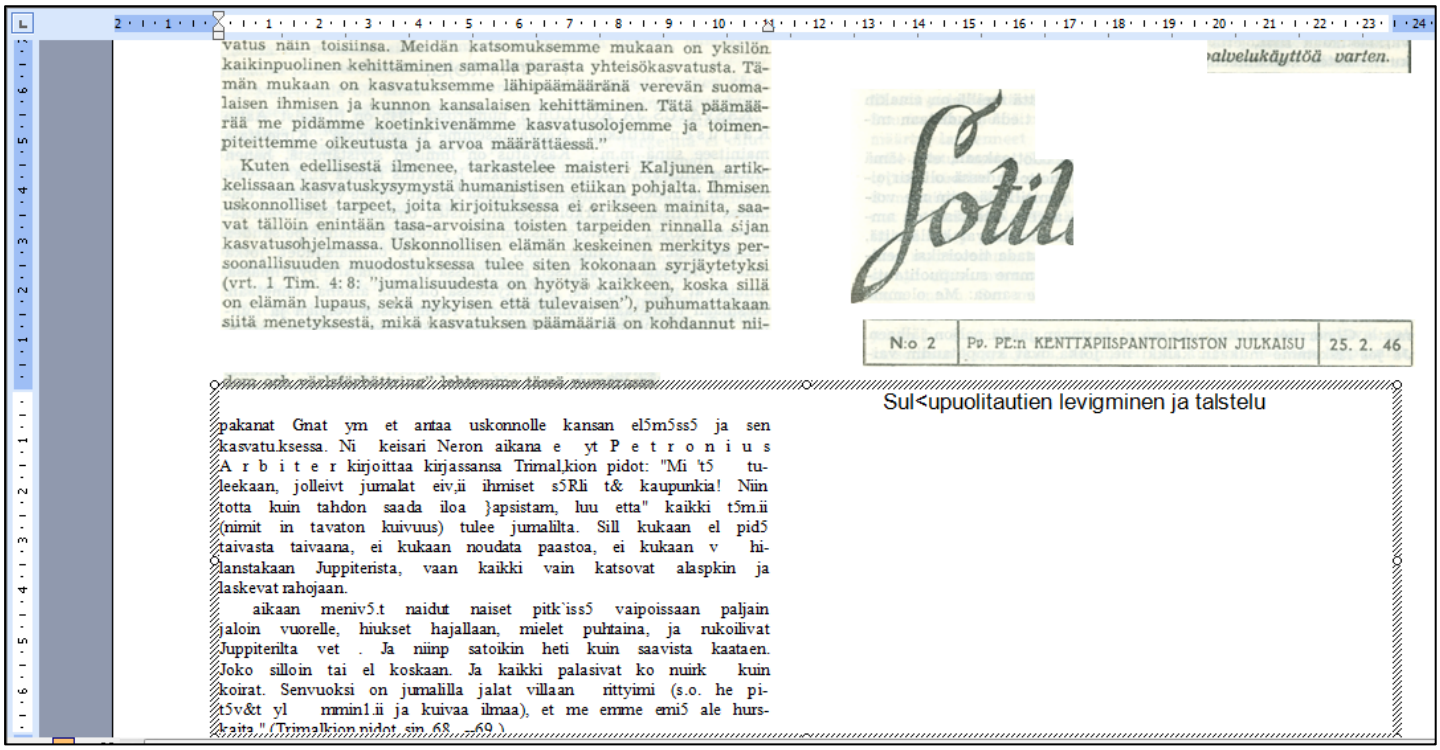

N Ainoastaan palvelulräytiöü varten I
i Q z I
I ^ I 3
_ Pv Plízn KENUÅPIISPANTOIHISTON JULKAISU -9. I. 46
Muutamia ajatuksia sotilasillanviettojen
ohjelmistosta.
Sotilasillanvietot voitaneen ryhmitellä:
1) hengelliset, 2) isä.mıiaallisval.istuksell:Iset ja 8) aj-anvietteelliset
virkistystilaisuudet. Seuraaıvassa eräitä niàlkäkoıhtia näiden eri tilaisuuksien ohjelmista.
1. Hengelliset iflhlıvtilalsuudet.
Hengellisten illanviettojen tarkoituıkseaıa on hengellisen elämän
herãttäırninen ja rakentaminen sekä vahvistaminen kilvoituksen
tiellä. Näiden tilaisuuksien ohjelman tulisi siis vastata edellämainittuja päämääriä. Ohjelman pääkohtana tulee näinollen olla sanan-julistus. Pulieiden on oltava =kutenı aina, asiallisia, heräfläviä, rakentavia ja Kristus-keskeisiä? johon sisältyy myös voimakas sosiaalinen velvoítus. Jokainen gpuhe olkoon huolella valmistettu, varsinkin, jos toivomme vaativarınpien kuulijoidenı pysyvän mukana ıharrastuksissamme. Todellinen valmistushan on, tiedämme, väsymätön
kilvoitus. Vain tietä voidaan odottaa siimaıtksia. Puıhujaksi olisi
silloin tällöin saatava myös joku vierailevа sotilas- tai siviílipappi.
Erittäin tervetulleen ja rikkaan puhuja-avun voi saada niistä kanıtahenkilöistã ja varusmiehistä-, joiden kohdalla Jumalan armo on
tapahtunut ja jotka mielellään todistavat. Tällaisia todistavia-
_ 1 _ V

Kuva 7. Free OCR -ohjelman tekstintunnistusnäyte testiaineistosta 1.

Lopputuloksessa oli virheitä eli tunnistamattomia merkkejä niin paljon, että ohjelman käyttäminen projektissa olisi ollut turhaa, joten virheitä ei korjattu suuren työmäärän takia.

5.1.2 Free Online OCR -ohjelma

Free online OCR /3/ on yksinkertainen suoraan selaimesta toimiva tekstintunnistusohjelma. TIFF-kuvan muuntamisen jälkeen huomattiin, että suurin osa pienistä L-kirjaimista on muuttunut f-kirjaimiksi. PDF:n muutoksessa osa tekstistä tunnistettiin, mutta virheitä löytyi, kun taas suurin osa tulostui kuvana. Käsiteltäessä Jpeg-kuva Word-tiedostoksi ohjelma tunnisti osan (kuva 8) tekstistä, mutta lähes joka toinen sana oli epäonnistunut, joten sillä ei ole mitään merkitystä, ettei ohjelma tunnistanut ä- ja ö-kirjaimia.


vatus näin toisiinsa. Meidän katsomuksemme mukaan on yksilön kaikinpuolinen kehittäminen samalla parasta yhteisökasvatusta. Tämän mukaan on kasvatuksemme lähipäämääränä verevän suomalaisen ihmisen ja kunnon kansalaisen kehittäminen. Tätä päämäärää me pidämme koetinkivenämme kasvatusolojemme ja toimenpiteittemme oikeutusta ja arvoa määrättäessä."

Kuten edellisestä ilmenee, tarkastelee maisteri Kaljunen artikkelissaan kasvatuskysymystä humanistisen etiikan pohjalta. Ihmisen uskonnolliset tarpeet, joita kirjoituksessa ei erikseen mainita, saavat tällöin enintään tasa-arvoisina toisten tarpeiden rinnalla sijan kasvatusohjelmassa. Uskonnollisen elämän keskeinen merkitys persoonallisuuden muodostuksessa tulee siten kokonaan syrjäytetyksi (vrt. 1 Tim. 4: 8: "jumalisuudesta on hyötyä kaikkeen, koska sillä on elämän lupaus, sekä nykyisen että tulevaisen"), puhumattakaan siitä menetyksestä, mikä kasvatuksen päämääriä on kohdannut nii-

N:o 2 | Pp. PE:n KENTTÄPIISPANTOIMISTON JULKAISU | 25. 2. 46

Sul<upuolitautien levigminen ja talstelu

pakanat Gnat ym et antaa uskonnolle kansan el5m5ss5 ja sen kasvatu.ksessa. Ni keisari Neron aikana e yt P e t r o n i u s A r b i t e r kirjoittaa kirjassansa Trimal,kion pidot: "Mi 't5 tuleekaan, jolleivt jumalat eiv,ii ihmiset s5Rli t& kaupunkia! Niin totta kuin tahdon saada iloa }apsistam, luu etta" kaikki t5m.ii (nimit in tavaton kuivuus) tulee jumalilta. Sill kukaan el pid5 taivasta taivaana, ei kukaan noudata paastoa, ei kukaan v hilanstakaan Juppiterista, vaan kaikki vain katsovat alaspkin ja laskevat rahojaan.

aikaan meniv5.t naidut naiset pitk`iss5 vaipoissaan paljain jaloin vuorelle, hiukset hajallaan, mielet puhtaina, ja rukoilivat Juppiterilta vet . Ja niinp satoikin heti kuin saavista kaataen. Joko silloin tai el koskaan. Ja kaikki palasivat ko nuirk kuin koirat. Senvuoksi on jumalilla jalat villaan rittyimi (s.o. he pit5v&t yl mmin1.ii ja kuivaa ilmaa), et me emme emi5 ale hurskaita." (Trimalkion pidot, sin. 68 ...69.)

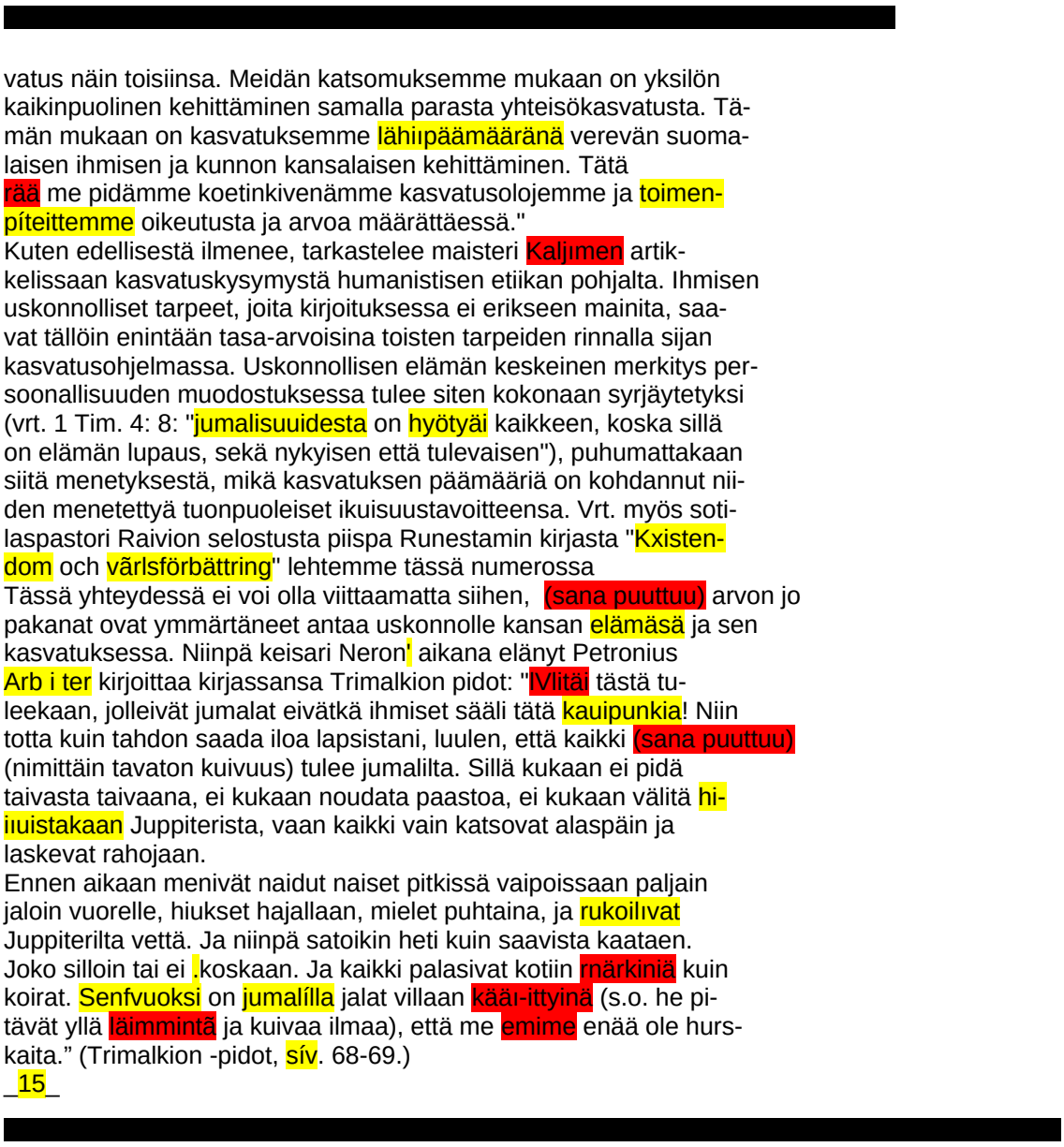
Kuva 8. Free online ocr -ohjelman tekstintunnistusnäyte

Free Online OCR -ohjelman testissä käytettiin testiaineistoa 2 (Taulukko 2, s. 17). Tulos oli tekstintunnistuksen kannalta heikko, sillä virheitä oli paljon. Niiden korjaaminen testissä olisi vaatinut lähes koko tekstin kirjoittamisen uudelleen.

5.1.3 i2OCR -ohjelma

Englannin ja suomen lisäksi i2OCR-ohjelmasta löytyy 31 vaihtoehtoista kieltä. Ohjelma /4/ tukee useita kuvaformaatteja (Taulukko 3, s.17). Lopputuloksessa (Kuva 9) oli yllättävän vähän virheitä, mutta silti liikaa.

Seuraavanlaisesti tulostui käännös testimateriaalista 2 (Taulukko 2, s.17). Punaisella ovat virheet, joissa ohjelma on jättänyt sanasta tunnistamatta tai tunnistanut väärin enemmän kuin yhden merkin. Keltaisella ovat sanat, joissa ohjelma on tunnistanut väärin vain yhden merkin. Punaisella merkittyjä sanoja löytyi 9 kappaletta. Keltaisella merkittyjä sanoja löytyi 17 kappaletta.

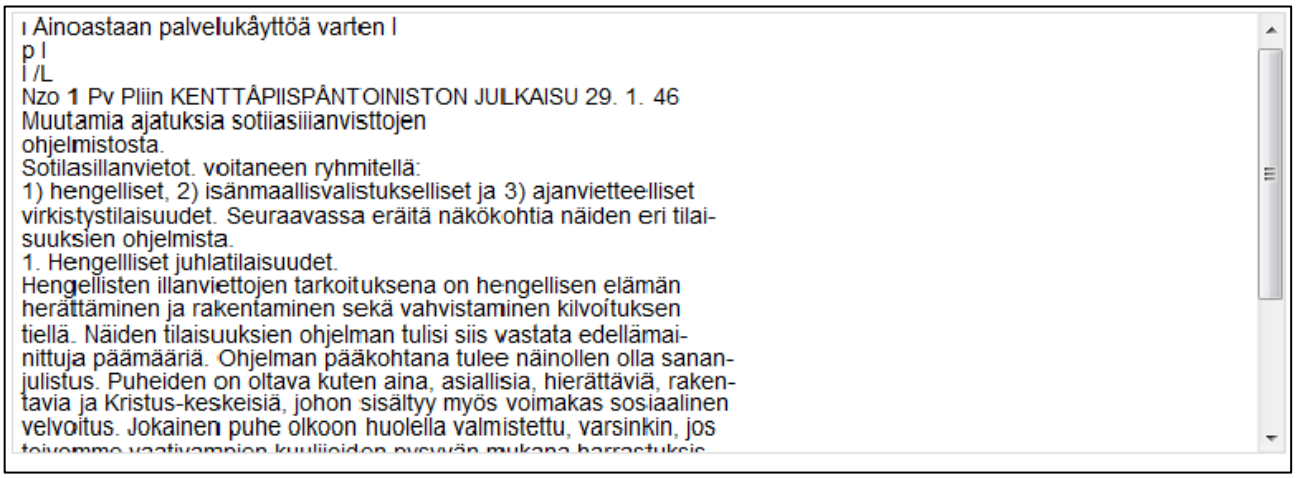

vatus näin toisiinsa. Meidän katsomuksemme mukaan on yksilön kaikinpuolinen kehittäminen samalla parasta yhteisökasvatusta. Tämän mukaan on kasvatuksemme lähiıpäämääränä verevän suomalaisen ihmisen ja kunnon kansalaisen kehittäminen. Tätä rää me pidämme koetinkivenämme kasvatusolojemme ja toimenpíteittemme oikeutusta ja arvoa määrättäessä."
Kuten edellisestä ilmenee, tarkastelee maisteri Kaljımen artikkelissaan kasvatuskysymystä humanistisen etiikan pohjalta. Ihmisen uskonnolliset tarpeet, joita kirjoituksessa ei erikseen mainita, saavat tällöin enintään tasa-arvoisina toisten tarpeiden rinnalla sijan kasvatusohjelmassa. Uskonnollisen elämän keskeinen merkitys persoonallisuuden muodostuksessa tulee siten kokonaan syrjäytetyksi (vrt. 1 Tim. 4: 8: "jumalisuuidesta on hyötyäi kaikkeen, koska sillä on elämän lupaus, sekä nykyisen että tulevaisen"), puhumattakaan siitä menetyksestä, mikä kasvatuksen päämääriä on kohdannut niiden menetettyä tuonpuoleiset ikuisuustavoitteensa. Vrt. myös sotilaspastori Raivion selostusta piispa Runestamin kirjasta "Kxistendom och värlsförbättring" lehtemme tässä numerossa
Tässä yhteydessä ei voi olla viittaamatta siihen, (sana puuttuu) arvon jo pakanat ovat ymmärtäneet antaa uskonnolle kansan elämäsä ja sen kasvatuksessa. Niinpä keisari Neron' aikana elänyt Petronius Arb i ter kirjoittaa kirjassansa Trimalkion pidot: "Ivlitäi tästä tuleekaan, jolleivät jumalat eivätkä ihmiset sääli tätä kauipunkia! Niin totta kuin tahdon saada iloa lapsistani, luulen, että kaikki (sana puuttuu) (nimittäin tavaton kuivuus) tulee jumalilta. Sillä kukaan ei pidä taivasta taivaana, ei kukaan noudata paastoa, ei kukaan välitä hiiuistakaan Juppiterista, vaan kaikki vain katsovat alaspäin ja laskevat rahojaan.
Ennen aikaan menivät naidut naiset pitkissä vaipoissaan paljain jaloin vuorelle, hiukset hajallaan, mielet puhtaina, ja rukoilıvat Juppiterilta vettä. Ja niinpä satoikin heti kuin saavista kaataen. Joko silloin tai ei .koskaan. Ja kaikki palasivat kotiin rnärkiniä kuin koirat. Senfvuoksi on jumalílla jalat villaan kääı-ittyinä (s.o. he pitävät yllä läimmintã ja kuivaa ilmaa), että me emime enää ole hurskaita." (Trimalkion -pidot, sív. 68-69.)
_15_

Kuva 9. i2OCR-ohjelman tekstintunnistusnäyte testiaineistosta 2

Ohjelma pystyy tulostamaan testatun tiedoston vain Word-tiedostoksi tai tekstiksi selaimeen. Virheiden korjaaminen kestää kauan vertaamalla alkuperäistekstiin.

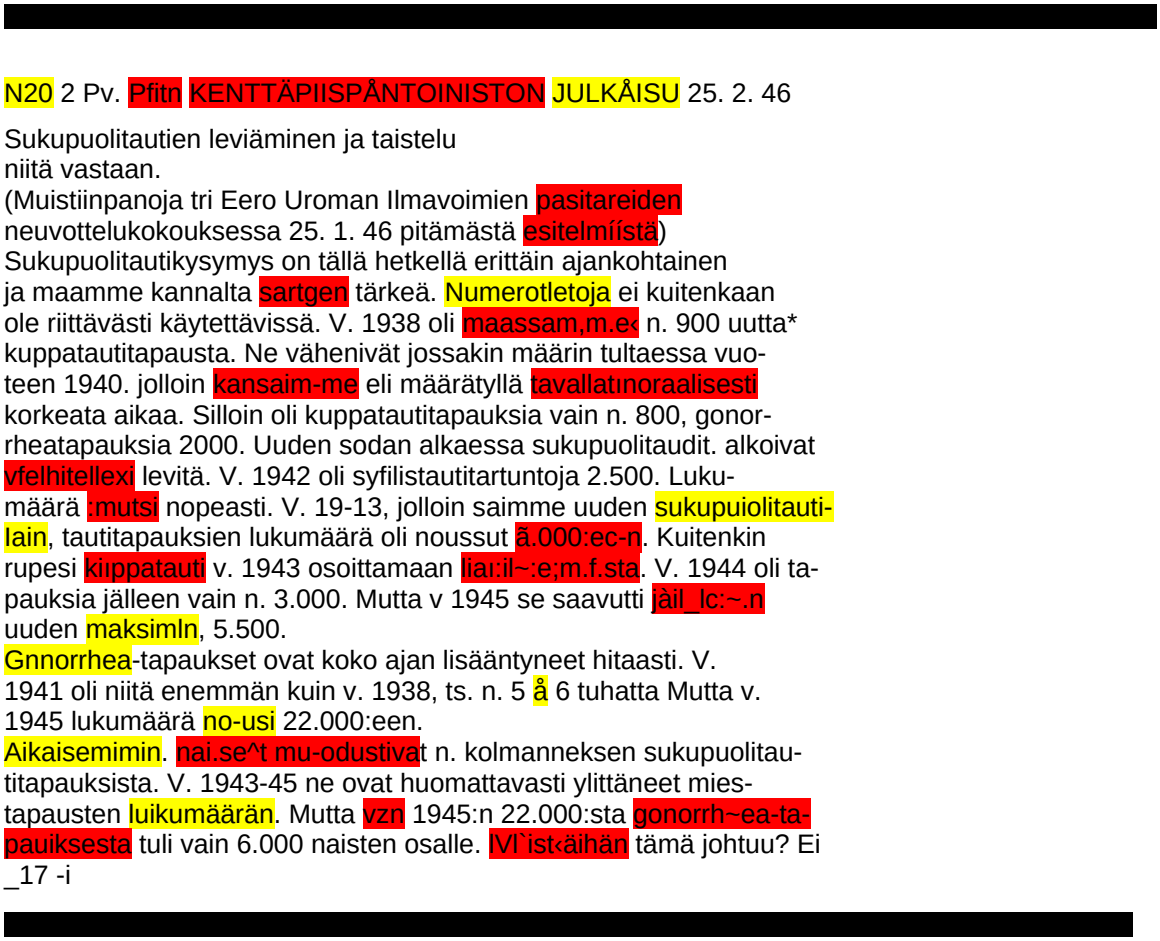
5.1.4 New OCR -ohjelma

New OCR -ohjelmassa kieleksi voi valita suomen, englannin ja ruotsin lisäksi 26 kieltä (Taulukko 1, s.16), mutta eräässä testissä vaikka ohjelmassa /5/ valittiin kieleksi suomi, ä ja ö eivät silti aina tulostuneet, vaan joissain kohdin ne olivat a ja o. Käsiteltävän tiedoston koko saa olla enintään 5 MB (Megabyte) eli megatavua. Tavu on tallennuskapasiteetin mittayksikkö. Testissä (Kuva 10) käytettiin testiaineistoja 1 ja 2 (Taulukko 2, s.17).


ı Ainoastaan palvelukäyttöä varten ı
p ı
ı /L
Nzo 1 Pv Pliin KENTTÅPIISPÅNTOINISTON JULKAISU 29. 1. 46
Muutamia ajatuksia sotiiasiiianvisttojen
ohjelmistosta.
Sotilasillanvietot. voitaneen ryhmitellä:
1) hengelliset, 2) isänmaallisvalistukselliset ja 3) ajanvietteelliset
virkistystilaisuudet. Seuraavassa eräitä näkökohtia näiden eri tilaisuuksien ohjelmista.
1. Hengellliset juhlatilaisuudet.
Hengellisten illanviettojen tarkoituksena on hengellisen elämän
herättäminen ja rakentaminen sekä vahvistaminen kilvoituksen
tiellä. Näiden tilaisuuksien ohjelman tulisi siis vastata edellämainittuja päämääriä. Ohjelman pääkohtana tulee näinollen olla sanan-
julistus. Puheiden on oltava kuten aina, asiallisia, hierättäviä, rakentavia ja Kristus-keskeisiä, johon sisältyy myös voimakas sosiaalinen
velvoitus. Jokainen puhe olkoon huolella valmistettu, varsinkin, jos

Kuva 10. New OCR -ohjelman tekstintunnistusnäyte

Seuraavanlaisesti tulostui käännös (Kuva 11) testiaineistosta 2 (Taulukko 2, s.17); Punaisella ovat virheet, joissa ohjelma on jättänyt sanasta tunnistamatta tai tunnistanut väärin enemmän kuin yhden merkin. Keltaisella ovat sanat, joissa ohjelma on tunnistanut sanasta tai tunnistanut väärin vain yhden merkin. Punaisella merkittyjä sanoja löytyi 19 kappaletta. Keltaisella merkittyjä sanoja löytyi 9 kappaletta.



N20 2 Pv. Pfitn KENTTÄPIISPÅNTOINISTON JULKÅISU 25. 2. 46

Sukupuolitautien leviäminen ja taistelu
niitä vastaan.
(Muistiinpanoja tri Eero Uroman Ilmavoimien pasitareiden
neuvottelukokouksessa 25. 1. 46 pitämästä esitelmíístä)
Sukupuolitautikysymys on tällä hetkellä erittäin ajankohtainen
ja maamme kannalta sartgen tärkeä. Numerotletoja ei kuitenkaan
ole riittävästi käytettävissä. V. 1938 oli maassam,m.e‹ n. 900 uutta*
kuppatautitapausta. Ne vähenivät jossakin määrin tultaessa vuo-
teen 1940. jolloin kansaim-me eli määrätyllä tavallatınoraalisesti
korkeata aikaa. Silloin oli kuppatautitapauksia vain n. 800, gonor-
rheatapauksia 2000. Uuden sodan alkaessa sukupuolitaudit. alkoivat
vfelhitellexi levitä. V. 1942 oli syfilistautitartuntoja 2.500. Luku-
määrä .mutsi nopeasti. V. 19-13, jolloin saimme uuden sukupuiolitauti-
lain, tautitapauksien lukumäärä oli noussut ã.000:ec-n. Kuitenkin
rupesi kiıppatauti v. 1943 osoittamaan liaı:il~:e;m.f.sta. V. 1944 oli ta-
pauksia jälleen vain n. 3.000. Mutta v 1945 se saavutti jàil_lc:~.n
uuden maksimln, 5.500.
Gnnorrhea-tapaukset ovat koko ajan lisääntyneet hitaasti. V.
1941 oli niitä enemmän kuin v. 1938, ts. n. 5 å 6 tuhatta Mutta v.
1945 lukumäärä no-usi 22.000:een.
Aikaisemimin. nai.se^t mu-odustivat n. kolmanneksen sukupuolitau-
titapauksista. V. 1943-45 ne ovat huomattavasti ylittäneet mies-
tapausten luikumäärän. Mutta vzn 1945:n 22.000:sta gonorrh~ea-ta-
pauiksesta tuli vain 6.000 naisten osalle. IVI`ist‹äihän tämä johtuu? Ei
_17 -i

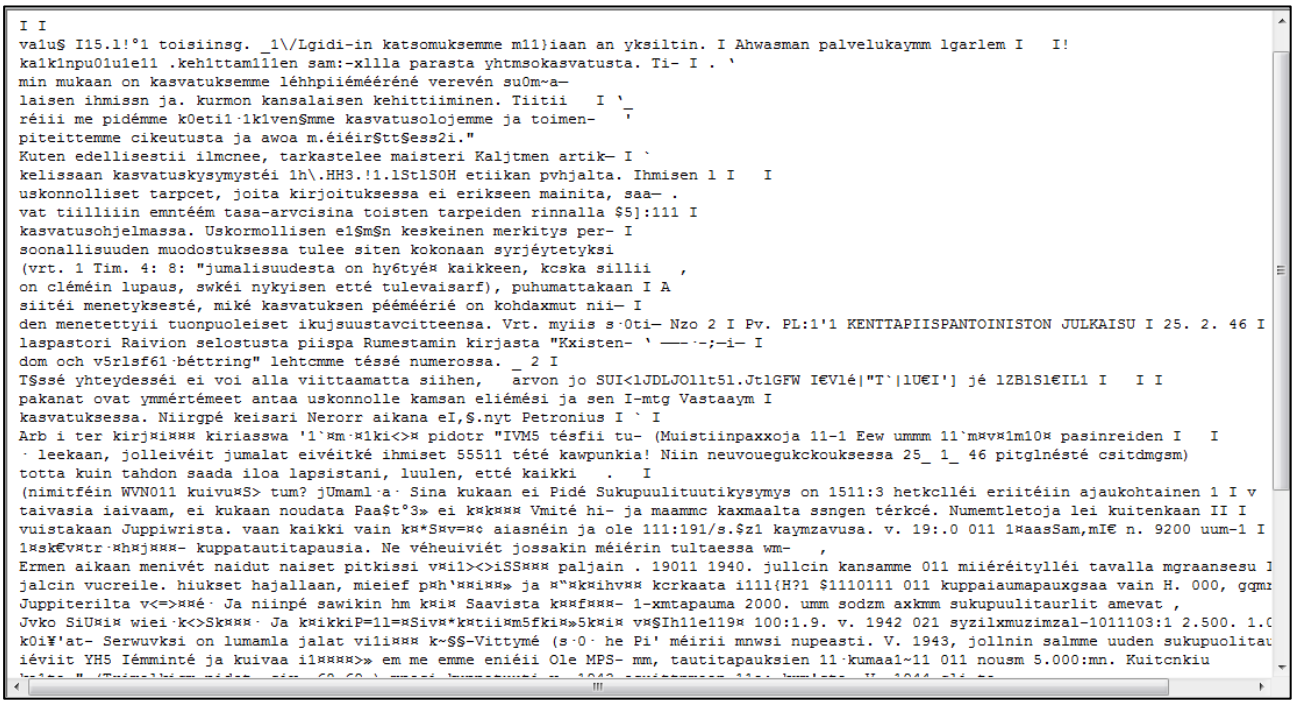
Kuva 11. New OCR -ohjelman tekstintunnistusnäyte testiaineistosta 2

Ohjelman tulostamassa Word-tiedostossa oli erittäin paljon virheitä eli tunnistamattomia merkkejä niin paljon, että virheiden korjaaminen olisi vienyt liian paljon aikaa.

5.1.5 OCR Convert -ohjelma

OCR Convert -ohjelman /6/ kielivaihtoehtoina löytyvät espanja, englanti, saksa, italia ja hollanti (Taulukko 1, s.16), eli suomen merkkejä ei tästä tuotteesta löydy.

Tulokset aukeavat tekstinä uuteen välilehteen. Ensimmäisessä testitiedostossa oli edes muutama sana joista sai selvää (kuva 12), mutta toisella kerralla tuloksena oli vain muutama rivi, joista ei saa selvää.



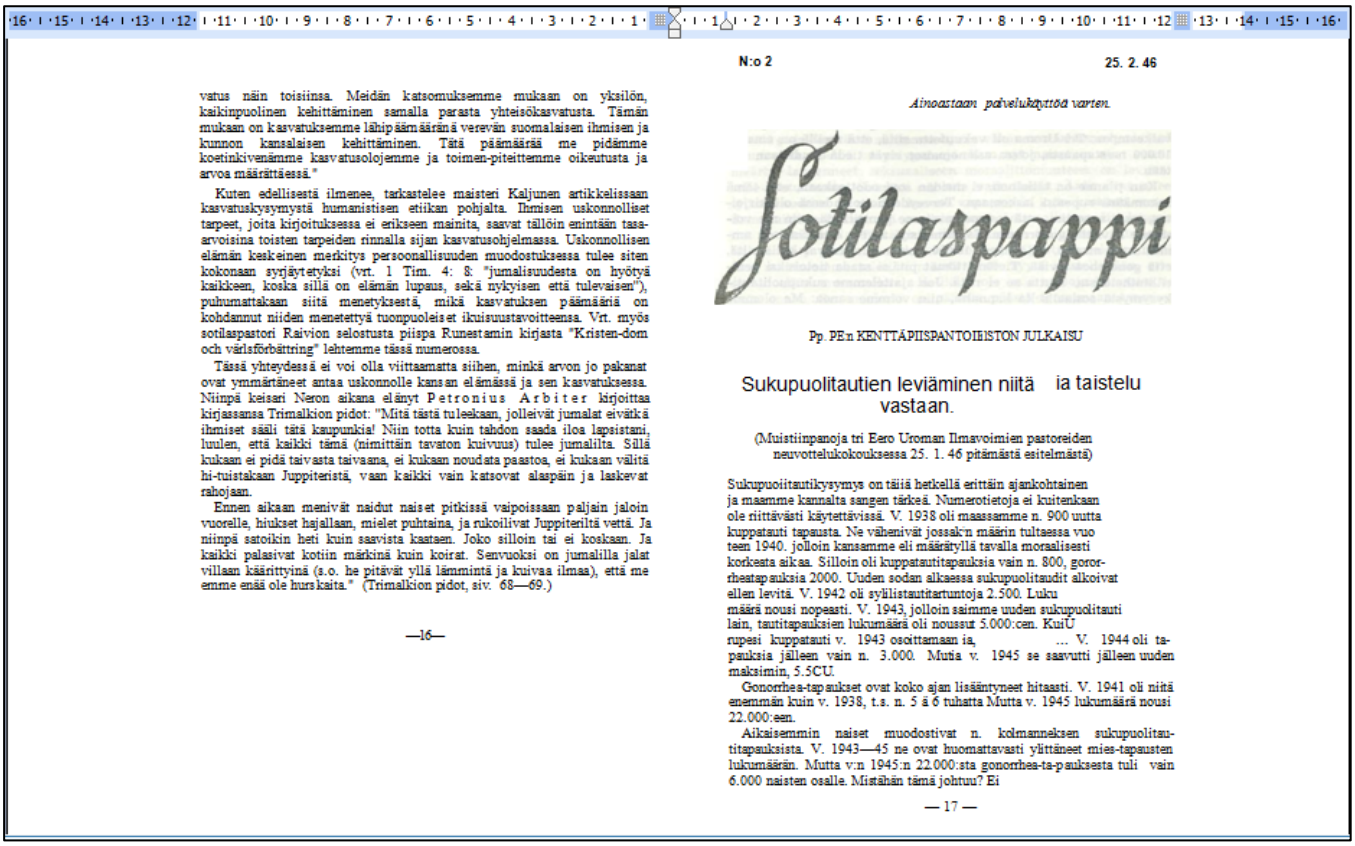
Kuva 12. OCR Convert -ohjelman tekstintunnistusnäyte

OCR Convert -ohjelman testissä käytettiin testiaineistoa 2 (Taulukko 2, s.17). Tulos oli käyttökelpoisuudeltaan heikko. Lähes jokainen sana oli tulostunut väärin, eikä tämän työn tarkoituksena ole kirjoittaa koko tekstiä uudelleen.

5.1.6 OCR Now! -ohjelma

OCR Now! -sivusto vaatii rekisteröitymisen /7/, jotta kuvia pääsee käsittelemään. Kieliä löytyy englannin ja suomen lisäksi 14.

Lopputulokset Wordilla olivat tyydyttävät (Kuva 13). Monia virheitä silti löytyi. Tiedostojen käsittely vaatii krediittejä, joita alussa on vain kaksi, joten kovin ilmaiseksi tätä sivustoa ei voi kutsua. Testissä käytettiin testiaineistoa 2 (Taulukko 2, s.17).



vatus näin toisiinsa. Meidän katsomuksemme mukaan on yksilön, kaikinpuolinen kehittäminen samalla parasta yhteisökasvatusta. Tämän mukaan on kasvatuksemme lähipäämääränä verevän suomalaisen ihmisen ja kunnon kansalaisen kehittäminen. Tätä päämäärää me pidämme koetinkivenämme kasvatusolojemme ja toimen-piteittemme oikeutusta ja arvoa määrättäessä."

Kuten edellisestä ilmenee, tarkastelee maisteri Kaljunen artikkelissaan kasvatuskysymystä humanistisen etiikan pohjalta. Ihmisen uskonnolliset tarpeet, joita kirjoituksessa ei erikseen mainita, saavat tällöin enintään tasa-arvoisina toisten tarpeiden rinnalla sijan kasvatusohjelmassa. Uskonnollisen elämän keskeinen merkitys persoonallisuuden muodostuksessa tulee siten kokonaan syrjäytetyksi (vrt. 1 Tim. 4: 8: "jumalisuudesta on hyötyä kaikkeen, koska sillä on elämän lupaus, sekä nykyisen että tulevaisen"), puhumattakaan siitä menetyksestä, mikä kasvatuksen päämääriä on kohdannut niiden menetettyä tuonpuoleiset ikuisuustavoitteensa. Vrt. myös sotilaspastori Raivion selostusta piispa Runestamin kirjasta "Kristen-dom och världsförbättring" lehtemme tässä numerossa.

Tässä yhteydessä ei voi olla viittaamatta siihen, minkä arvon jo pakanat ovat ymmärtäneet antaa uskonnolle kansan elämässä ja sen kasvatuksessa. Niinpä keisari Neron aikana elänyt Petronius Arbiter kirjoittaa kirjassansa Trimalkion pidot: "Mitä tästä tuleekaan, jolleivät jumalat eivätkä ihmiset säälі tätä kaupunkia! Niin totta kuin tahdon saada iloa lapsistani, luulen, että kaikki tämä (nimittäin tavaton kuivuus) tulee jumalilta. Sillä kukaan ei pidä taivasta taivaana, ei kukaan noudata paastoa, ei kukaan välitä hi-tuistakaan Juppiteristä, vaan kaikki vain katsovat alaspäin ja laskevat rahojaan.

Ennen aikaan menivät naidut naiset pitkissä vaipoissaan paljain jaloin vuorelle, hiukset hajallaan, mielet puhtaina, ja rukoilivat Juppiteriltä vettä. Ja niinpä satoikin heti kuin saavista kaatаen. Joko silloin tai ei koskaan. Ja kaikki palasivat kotiin märkinä kuin koirat. Senvuoksi on jumalilla jalat villaan käärittyinä (s.o. he pitävät yllä lämmintä ja kuivaa ilmaa), että me emme enää ole hurskaita." (Trimalkion pidot, siv. 68—69.)

—16—

N:o 2 25. 2. 46

Ainoastaan palveluksyttöä varten.

Sotilaspappi

Pp. PE:n KENTTÄPIISPANTOIHISTON JULKAISU

Sukupuolitautien leviäminen niitä ia taistelu vastaan.

(Muistiinpanoja tri Eero Uroman Ilmavoimien pastoreiden neuvottelukokouksessa 25. 1. 46 pitämästä esitelmästä)

Sukupuoiitautikysymys on täiiä hetkellä erittäin ajankohtainen ja maamme kannalta sangen tärkeä. Numerotietoja ei kuitenkaan ole riittävästi käytettävissä. V. 1938 oli maassamme n. 900 uutta kuppatauti tapausta. Ne vähenivät jossak'n määrin tultaessa vuo teen 1940. jolloin kansamme eli määrätyllä tavalla morsalisesti korkeata aikaa. Silloin oli kuppatautitapauksia vain n. 800, gonor-rheatapauksia 2000. Uuden sodan alkaessa sukupuolitaudit alkoivat ellen levitä. V. 1942 oli syiilistautitartuntoja 2.500. Luku määrä nousi nopeasti. V. 1943, jolloin saimme uuden sukupuolitauti lain, tautitapauksien lukumäärä oli noussut 5.000:cen. KuiU rupesi kuppatauti v. 1943 osoittamaan ia, ... V. 1944 oli tapauksia jälleen vain n. 3.000. Mutia v. 1945 se saavutti jälleen uuden maksimin, 5.5CU.

Gonorrhea-tapaukset ovat koko ajan lisääntyneet hitaasti. V. 1941 oli niitä enemmän kuin v. 1938, t.s. n. 5 ä 6 tuhatta Mutta v. 1945 lukumäärä nousi 22.000:een.

Aikaisemmin naiset muodostivat n. kolmanneksen sukupuolitautitapauksista. V. 1943—45 ne ovat huomattavasti ylittäneet mies-tapausten lukumäärän. Mutta v:n 1945:n 22.000:sta gonorrhea-ta-pauksesta tuli vain 6.000 naisten osalle. Mistähän tämä johtuu? Ei

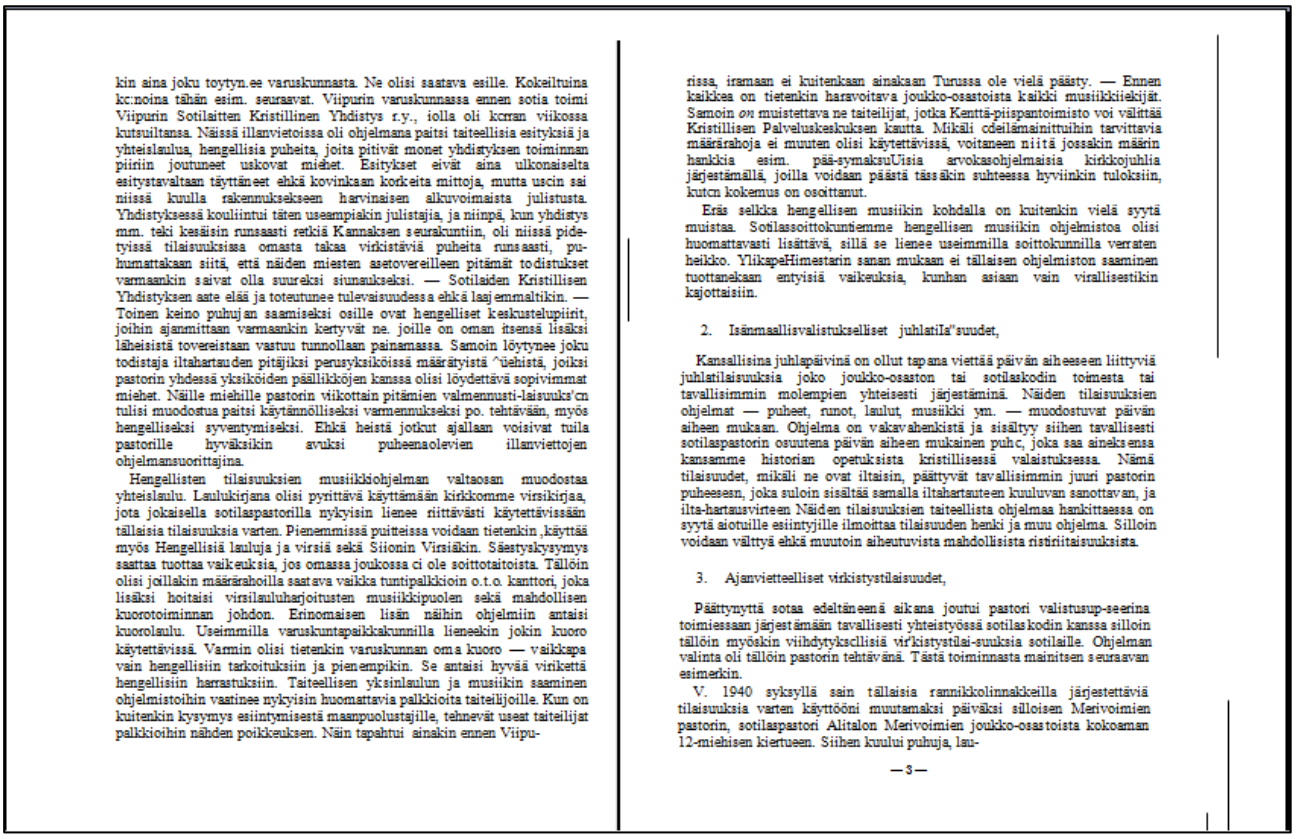
— 17 —

Kuva 13. OCR Now! -ohjelman tekstintunnistusnäyte

Krediitit maksavat (Taulukko 5, s.18) 20 kpl 1.99 puntaa ja 500 krediittiä (Credit) taas on 31,99 puntaa. Krediitit ovat Internet-sivustoilla käytettävää valuuttaa. Virheiden korjaus veisi niiden pienen määrän huomioon ottaen vain vähän aikaa.

5.1.7 OCR online -ohjelma

OCR online -ohjelma vaatii rekisteröitymisen /8/. Ilmaiseksi kuvatiedostoja voi käsitellä vain 10 kappaletta viikossa, jonka jälkeen ohjelma muuttuu maksulliseksi. Kielivaihtoehtoina (Taulukko 1, s.16) listassa ovat englanti, tšekki, ranska, saksa, italia, latina, latvia, puola, portugali, romania, venäjä, espanja ja turkki. Lisäksi OCR Onlinen kielivaihtoehtona on mikä tahansa kieli. Ohjelmassa on asetuksista lisättävä suomen kieli, koska listassa on aluksi vain 15 kieltä. Yhteensä kieliä on 153. Vilkaisulla ei käännöksestä (kuva 14) virheitä paljoa huomaa, mutta tarkemmin katsoessa virheitä näkyy useita.



kin aina joku toytyn.ee varuskunnasta. Ne olisi saatava esille. Kokeiltuina kc:noina tähän esim. seuraavat. Viipurin varuskunnassa ennen sotia toimi Viipurin Sotilaitten Kristillinen Yhdistys r.y., iolla oli kcnan viikossa kutsuiltansa. Näissä illanvietoissa oli ohjelmana paitsi taiteellisia esityksiä ja yhteislaulua, hengellisia puheita, joita pitivät monet yhdistyksen toiminnan piiriin joutuneet uskovat miehet. Esitykset eivät aina ulkonaiselta esitystavaltaan täyttäneet ehkä kovinkaan korkeita mittoja, mutta uscin sai niissä kuulla rakennuksekseen harvinaisen alkuvoimaista julistusta. Yhdistyksessä kouliintui täten useampiakin julistajia, ja niinpä, kun yhdistys mm. teki kesäisin runsaasti retkiä Kannaksen seuraskuntiin, oli niissä pidetyissä tilaisuuksissa omasta takaa virkistäviä puheita runsaasti, puhumattakaan siitä, että näiden miesten asetovereilleen pitämät todistukset varmaankin saivat olla suureksi siunaukseksi. — Sotilaiden Kristillisen Yhdistyksen aate elää ja toteutunee tulevaisuudessa ehkä laajemmaltikin. — Toinen keino puhujan saamiseksi osille ovat hengelliset keskustelupiirit, joihin ajanmittaan varmaankin kertyvät ne. joille on oman itsensä lisäksi läheisistä tovereistaan vastuu tunnollaan painamassa. Samoin löytynee joku todistaja iltahartauden pitäjiksi perusyksiköissä määrätyistä ^uehistä, joiksi pastorin yhdessä yksiköiden päällikköjen kanssa olisi löydettävä sopivimmat miehet. Näille miehille pastorin viikottain pitämien valmennusti-laisuuks'cn tulisi muodostua paitsi käytännölliseksi varmennukseksi po. tehtävään, myös hengelliseksi syventymiseksi. Ehkä heistä jotkut ajallaan voisivat tuila pastorille hyväksikin avuksi puheenaolevien illanviettojen ohjelmansuorittajina.

Hengellisten tilaisuuksien musiikkiohjelman valtaosan muodostaa yhteislaulu. Laulukirjana olisi pyrittävä käyttämään kirkkomme virsikirjaa, jota jokaisella sotilaspastorilla nykyisin lienee riittävästi käytettävissään tällaisia tilaisuuksia varten. Pienemmissä puitteissa voidaan tietenkin ,käyttää myös Hengellisiä lauluja ja virsiä sekä Siionin Virsiäkin. Säestyskysymys saattaa tuottaa vaikeuksia, jos omassa joukossa ci ole soittotaitoista. Tällöin olisi joillakin määrärahoilla saatava vaikka tuntipalkkioin o.t.o. kanttori, joka lisäksi hoitaisi virsilauluharjoitusten musiikkipuolen sekä mahdollisen kuorotoiminnan johdon. Erinomaisen lisän näihin ohjelmiin antaisi kuorolaulu. Useimmilla varuskuntapaikkakunnilla lieneekin jokin kuoro käytettävissä. Varmin olisi tietenkin varuskunnan oma kuoro — vaikkapa vain hengellisiin tarkoituksiin ja pienempikin. Se antaisi hyvää virikettä hengellisiin harrastuksiin. Taiteellisen yksinlaulun ja musiikin saaminen ohjelmistoihin vastinee nykyisin huomattavia palkkioita taiteilijoille. Kun on kuitenkin kysymys esiintymisestä maanpuolustajille, tehnevät useat taiteilijat palkkioihin nähden poikkeuksen. Näin tapahtui ainakin ennen Viipurissa, iramasn ei kuitenkaan ainakaan Turussa ole vielä päästy. — Ennen kaikkea on tietenkin haravoitava joukko-osastoista kaikki musiikkiiekijät. Samoin *on* muistettava ne taiteilijat, jotka Kenttä-piispantoimisto voi välittää Kristillisen Palveluskeskuksen kautta. Mikäli cdeilämainittuihin tarvittavia määrärahoja ei muuten olisi käytettävissä, voitaneen niitä jossakin määrin hankkia esim. pää-symaksuUisia arvokasohjelmaisia kirkkojuhlia järjestämällä, joilla voidaan päästä tässäkin suhteessa hyviinkin tuloksiin, kutcn kokemus on osoittanut.

Eräs selkka hengellisen musiikin kohdalla on kuitenkin vielä syytä muistaa. Sotilassoittokuntiemme hengellisen musiikin ohjelmistoa olisi huomattavasti lisättävä, sillä se lienee useimmilla soittokunnilla verraten heikko. YlikapeHimestarin sanan mukaan ei tällaisen ohjelmiston saaminen tuottaneksaan entyisiä vaikeuksia, kunhan asiaan vain virallisestikin kajottaisiin.

2. Isänmaallisvalistukselliset juhlatila"suudet,

Kansallisina juhlapäivinä on ollut tapana viettää päivän aiheeseen liittyviä juhlatilaisuuksia joko joukko-osaston tai sotilaskodin toimesta tai tavallisimmin molempien yhteisesti järjestäminä. Näiden tilaisuuksien ohjelmat — puheet, runot, laulut, musiikki ym. — muodostuvat päivän aiheen mukaan. Ohjelma on vakavahenkistä ja sisältyy siihen tavallisesti sotilaspastorin osuutena päivän aiheen mukainen puhc, joka saa aineksensa kansamme historian opetuksista kristillisessä valaistuksessa. Nämä tilaisuudet, mikäli ne ovat iltaisin, päättyvät tavallisimmin juuri pastorin puheeseen, joka suloin sisältää samalla iltahartauteen kuuluvan sanottavan, ja ilta-hartausvirteen Näiden tilaisuuksien taiteellista ohjelmaa hankittaessa on syytä aiotuille esiintyjille ilmoittaa tilaisuuden henki ja muu ohjelma. Silloin voidaan välttyä ehkä muutoin aiheutuvista mahdollisista ristiriitaisuuksista.

3. Ajanvietteelliset virkistystilaisuudet,

Päättynyttä sotaa edeltäneenä aikana joutui pastori valistusup-seerina toimiessaan järjestämään tavallisesti yhteistyössä sotilaskodin kanssa silloin tällöin myöskin viihdytyksellisiä vir'kistystilai-suuksia sotilaille. Ohjelman valinta oli tällöin pastorin tehtävänä. Tästä toiminnasta mainitsen seuraavan esimerkin.

V. 1940 syksyllä sain tällaisia rannikkolinnakkeilla järjestettäviä tilaisuuksia varten käyttööni muutamaksi päiväksi silloisen Merivoimien pastorin, sotilaspastori Alitalon Merivoimien joukko-osastoista kokoaman 12-miehisen kiertueen. Siihen kuului puhuja, lau-

—3—

Kuva 14. OCR Online -ohjelman tekstintunnistusnäyte

Ohjelmalla on mahdollisuus muuttaa kuvatiedosto (Taulukko 3, s.17) neljään (4) eri tiedostomuotoon (Taulukko 4, s.18). Testissä käytettiin testiaineistoa 3 (Taulukko 2, s.17). Virheiden korjaaminen niiden pieni määrä huomioon ottaen käy nopeasti.

5.1.8 OCR terminal -ohjelma

OCR terminal -ohjelma liittyi myöhemmin ABBYY Finereader Onlineen /9/.

Kieliä tässä ohjelmassa on suomen lisäksi pari kymmentä (Taulukko 1, s.16). Ohjelma vaatii rekisteröitymisen. Rajoituksena on 20 ilmaista sivua kuukaudessa (Taulukko 5, s.18). Ohjelma tukee TIFF-, PDF-, JPEG- ja muita kuvaformaatteja (Taulukko 3, s.17). Ohjelma muuttaa kuvan neljään (4) eri tiedostomuotoon (Taulukko 4, s.18).

Seuraavanlaisesti tulostui käännös (Kuva 15) testiaineistosta 4 (Taulukko 2, s.17); Punaisella ovat virheet, joissa ohjelma on jättänyt sanasta tunnistamatta tai tunnistanut väärin enemmän kuin yhden merkin. Keltaisella ovat virheet, joissa ohjelma on tunnistanut sanasta tai tunnistanut väärin vain yhden merkin. Punaisia virheitä löytyi 1 kappaletta. Keltaisia virheitä löytyi 1 kappaletta.

keskinäiseen polemiikkiin, vaan sen sijaan olisi pyrittävä siihen hedelmälliseen yhteistyöhön, jota usko ja psykoterapeuttinen tekniikka sielunlääkinnälle tarjoavat. Niinpä voi uskonnollinen kokemus äkkiä avata potilaan silmät näkemään synnin synniksi, mikä puolestaan auttaa häntä löytämään sen lähteenä olevan kompleksin, ja toisaalta saattaa jälleen psykoanalyyttinen käsittely selvittää hänelle hänen kompleksinsa, mikä voi johdattaa hänet tuntemaan siitä aiheutuvan synnin, jolloin hänelle avautuu myös Jumalan anteeksiantamuksen tie. Lääkärin on vain tehtävä oikea diagnoosi voidakseen valita sen edellyttämän joko hengellisen tai psykoterapeuttisen hoitotavan.

Tekniikan ja uskon välinen vuorovaikutus tapahtuu tällöin seuraavasti. Tekniikka, psykoanalyysi, tutkii ja tuo päivänvaloon ihmissielun probleemit, usko, armo, hävittää ne tavalla, jota emme osaa edes selittää. Tekniikkaa voidaan oppia, ja se on käyttökelpoinen maaperän muokkauksessa, mutta hengellisiä kokemuksia sillä ei kyetä aikaansaamaan. Kun Tournier'lta on pyydetty sielunhoito-opillisia neuvoja, hänen on täytynyt tunnustaa, ettei hän osaa niitä antaa, sillä hän ei tiedä, kuten hän sanoo, mitä on tehtävä todellisessa sielunhädässä olevan ihmisen auttamiseksi. Hän on suorastaan sitä <m:eltä, ettei tuollaisessa tapauksessa ole tehtävä mitään, sillä jos ihminen siinä johonkin ryhtyy, se jo on itseensä eikä Jumalaan luottamsta. Siksi siinä ei olekaan tähdellistä se, mitä me teemme, vaan mitä me olemme. Jos itse olemme yhteydessä Jumalaan, ja tämä yhteys opettaa meidät näkemään viheliäisyytemme ja oman kykenemättömyytemme auttaa toisia, silloin jo tällä paikalla olemisemme on omiaan johtamaan heidät uskonnollisiin kokemuksiin. Tekniikkaan nähden voimme toisia neuvoa, mutta kun usko on kysymyksessä, ei neuvoistamme ole mitään hyötyä. Tekniikka auttaa ihmistä kuitenkin tutkimaan itseään ja syventämään hänen moraalisia arviointiperusteitaan. Tällöin ei normina ole enää muodollinen moraali, joka pitää silmällä vain sitä, mitä ihminen tekee, vaan syvällisempi moraalikäsitys, joka kysyy, minkälainen ihminen on. Tässäkin voi psykoanalyysi auttaa meitä löytämään kanssaihmisistämme heidän sydämensä usein salassa olevat hyvät ominaisuudet, ja mitä tulee meihin itseemme, se on omiaan paljastamaan sisimpämme pohjamutiaan myöten.

Muodollinen moraali, jonka normina siis on, onko jokin teko itsessään hyvä vai paha, onko se hyödyllinen vai ei ja mitä ihmiset siitä sanovat, johdattaa orjuuttavaan ja seurauksiltaan monta kertaa tuhoisaan moralismiin. Elämällä on kuitenkin lakinsa, ja siksi esim. kirkon on asetettava moraalille tietyt vaatimukset, sillä kuinka hyvät ihmisen tarkoitusperät ovatkin, paha vaanii häntä yhtenään ja saa hänet helposti saaliikseen, ellei kirkko ole opetuksellaan häntä opastamassa. Mutta tämän moraalin normina tulee olla Jumalan

— 50 —

Kuva 15. OCR Terminal –ohjelman tekstintunnistusnäyte testiaineistosta 4

Tuloksista löytyi erittäin vähän virheitä Wordissa. Jotkut kirjaimet ovat muuttuneet eikä teksti ole samalla tasolla kuin kuvassa, joten väliviivoja voi olla alkuperäisen kuvan takia joissain kohdissa, missä niitä ei pitäisi olla. Kuvan kuitenkin täytyy olla riittävän iso, jotta ohjelma tunnistaa merkit. PDF-tiedostossa ei sanahaulla tunnista hajanaisia sanoja.

5.2 Koneelle asennettavia OCR-ohjelmia

Kun testattavat selainpohjaiset ohjelmat oli käyty läpi, testattiin tekstintunnistusohjelmia, jotka asennetaan työasemalle. Testattavia ohjelmia kokeiltiin viisi (5) kappaletta ja tutkittiin, kuinka hyvä lopputulos niillä saadaan.

Testiaineistoista (Taulukko 7) pääasiassa testiaineistoa 2 käytettiin, mutta vaihtelevuuden vuoksi käytettiin myös testiaineistoja 1 ja 5. Useimmat testattavista ohjelmista olivat vain kokeiluversioita, joiden käyttöikä oli rajallinen, ja joissakin ohjelmissa myös tallennettavien tiedostojen lukumäärä oli rajattu. Yhteen työhön keskitettynä rajallinen aika kuitenkin riittää. Lukumääräisesti rajattujen ohjelmien ongelma on, että tulostettavia tiedostoja voi olla liikaa. Ohjelmissa oli eroavaisuuksia tuettujen kielten kanssa (Taulukko 6). Jotkut ohjelmat tukivat monia kieliä ja jotkut tunnistivat vain englanninkielestä löytyviä merkkejä ja sanoja. Ohjelmat erosivat myös siinä, mitä kuvaformaattia ne tukevat (Taulukko 8) ja mihin tiedostomuotoon käsitelty tiedosto tulostuu (Taulukko 9). Tuettuja kuvaformaatteja oli runsaasti: PDF, GIF, TIFF, JPEG ja PNG, BMP ja PPT (Power Point Table), joka on Microsoftin PowerPoint-ohjelman käyttämä tiedostomuoto ja PCX (Personal Computer eXchange), joka on vähenevässä käytössä oleva kuvaformaatti.

Tiedostomuotoja, jotka voi valita käsittelyn jälkeiseksi formaatiksi, oli myös muutamia: DOC, PDF, TXT, RTF, XLS ja HTML (Hypertext Markup Language), joka on hypertekstin merkintäkieli.

Ohjelma	suomi	englanti	ruotsi	muut kielet yht.
Acrobat pro 10	x	x	x	25
Autobahn		x		
Free OCR 3.0		x		
Finereader 10	x	x	x	186
Smart OCR		x		
PDF Converter 7		x		

Taulukko 6. Työasemalle asennettavien tekstintunnistusohjelmien kielet

Käsittelyissä käytettiin kolmea (3) eri testiaineistoa (Taulukko 7), jotta saataisiin erilaisia näkökulmia ja vaihtelevuutta.

Ohjelma	Testiaineisto 1	Testiaineisto 2	Testiaineisto 5
Acrobat pro 10	x		
Autobahn		x	
Free OCR 3.0		x	
Finereader 10		x	
Smart OCR		x	
PDF Converter 7			x

Taulukko 7. Ohjelmissa käytetyt testiaineistot

Ohjelma	PDF	GIF	TIFF	JPEG	BMP	PNG	PPT	PCX
Acrobat pro 10			x	x				
Autobahn	x		x	x				
Free OCR 3.0	x	x	x	x	x			
Finereader 10	x	x	x	x	x	x		x
Smart OCR	x	x	x	x	x	x		
PDF Converter 7	x	x	x	x	x	x	x	

Taulukko 8. Työasemalle asennettavien tekstintunnistusohjelmien käsiteltävät kuvaformaatit

Ohjelma	DOC	PDF	TXT	RTF	XLS	HTML
Acrobat pro 10		x				
Autobahn		x				
Free OCR 3.0	x		x			
Finereader 10	x	x	x	x	x	
Smart OCR	x	x	x	x		x
PDF Converter 7		x				

Taulukko 9. Työasemalle asennettavien tekstintunnistusohjelmien tulostettavat tiedostomuodot

Ohjelma	Ei rajoituksia	Hinta (€)	Kokeilu-aika (vrk)	Tiedostorajoitus (kpl)	Vesileima
Acrobat pro 10		199 dollaria	30		
Autobahn		960 euroa			x
Free OCR 3.0	x				
Finereader 10		129 euroa	15	50	
Smart OCR		99,90			x
PDF Converter 7		99,99 dollaria	30		

Taulukko 10. Työasemalle asennettavien tekstintunnistusohjelmien rajoitukset

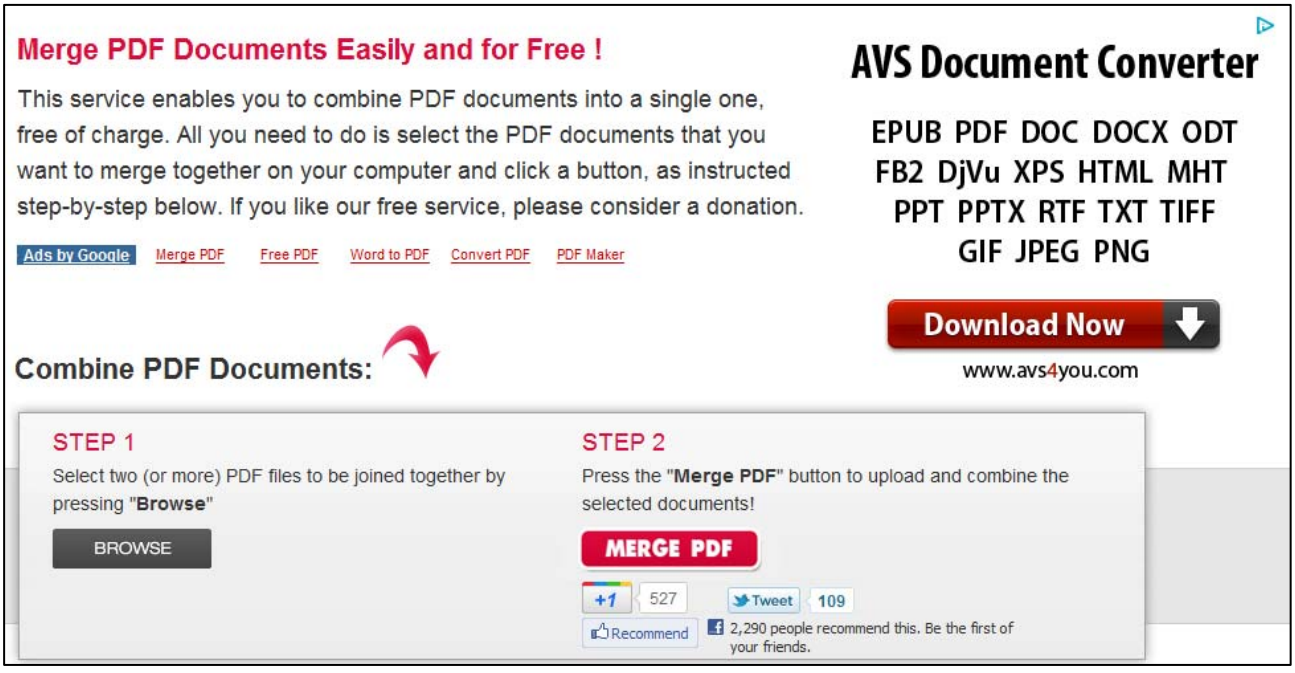
Rajoituksettomia ohjelmia oli vain yksi, mutta ohjelman tuottaman tunnistuksen laatu ei ole onnistunutta. Paras lopputulos oli Finereader 10:llä, mutta tiedosto-rajoitusten takia projektia ei saanut ohjelmalla tehtyä. Toiseksi paras ohjelma oli PDF Converter 7, jonka kuukauden koeaika oli riittävä selvityksen tekoon.

5.2.5 Abbyy Finereader 10 -ohjelma

Abbyy Finereader 10 -ohjelmasta löytyy 189 kieltä (Taulukko 6), joka on paljon. Ohjelman /10/ tunnistuksesta löytyi erittäin vähän virheitä.

Ohjelman ilmaisversiolla (Taulukko 10) ei saa kuin 50 sivua tallennettua, eikä se toimi kuin 15 vuorokautta. Ohjelma olisi vaihtoehto työn suorittamiselle, mutta rajoitustensa takia ohjelmaa ei voi käyttää. Aivan virheettömästi ohjelma ei tekstiä tunnistanut, mutta virheitä oli erittäin vähän. Ohjelman hinta on 129 euroa.

Jos joku haluaisi kyseisellä ohjelmalla vastaavan työn suorittaa ja haluaisi yhdistää tiedostot yhdeksi, PDF-tiedostot saa helposti yhdistettyä selaimesta toimivalla ilmaisella Merge PDF -ohjelmalla (Kuva 16) /11/. Myös Word-tiedosto oli riittävän onnistunut.

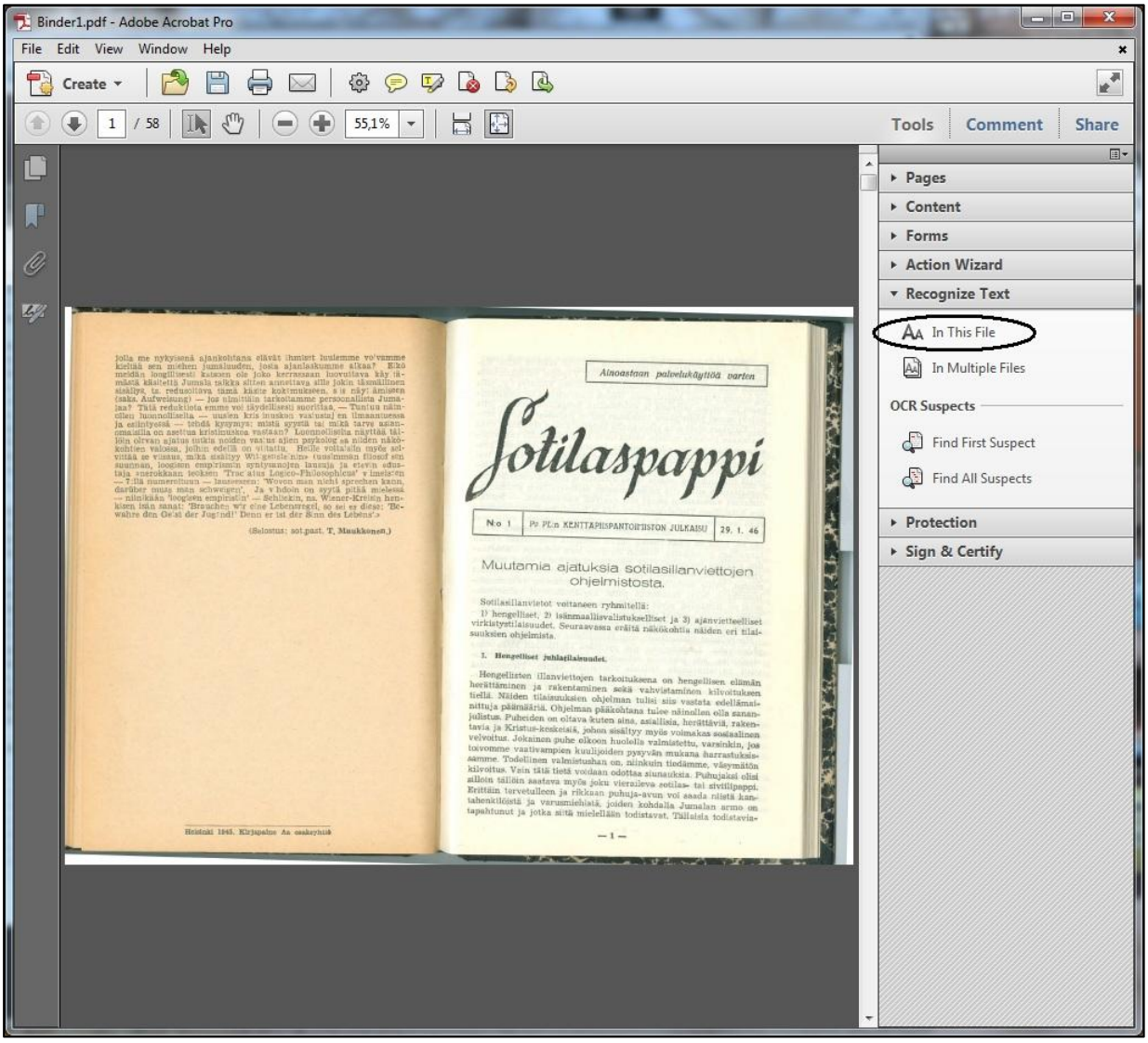


Kuva 16. Merge PDF -selainpohjaisen ohjelman näkymä

Ohjelman käsittelemien tiedostojen lopputulokset olivat selvityksen ohjelmista parhaat. Virheitä oli erittäin vähän eikä niiden korjaamiseen mene kauaa.

5.2.6 Adobe Acrobat pro 10 -ohjelma

Adobe Acrobat pro 10 -ohjelmasta oli käytössä kolmenkymmenen (30) päivän kokeiluversio /12/, mutta toisin kuin Finereaderissa (k. 5.2.1), tässä ohjelmassa ei ollut rajoituksia dokumenttien tallennusten suhteen (Taulukko 10 s.33). Ohjelman alussa muutettiin ensimmäiseksi haluttu määrä skannattuja kuvia eli tässä tapauksessa 46. vuosikerta yhdeksi PDF-tiedostoksi, jonka jälkeen käskettiin ohjelmaa tunnistamaan teksti (kuva 17).



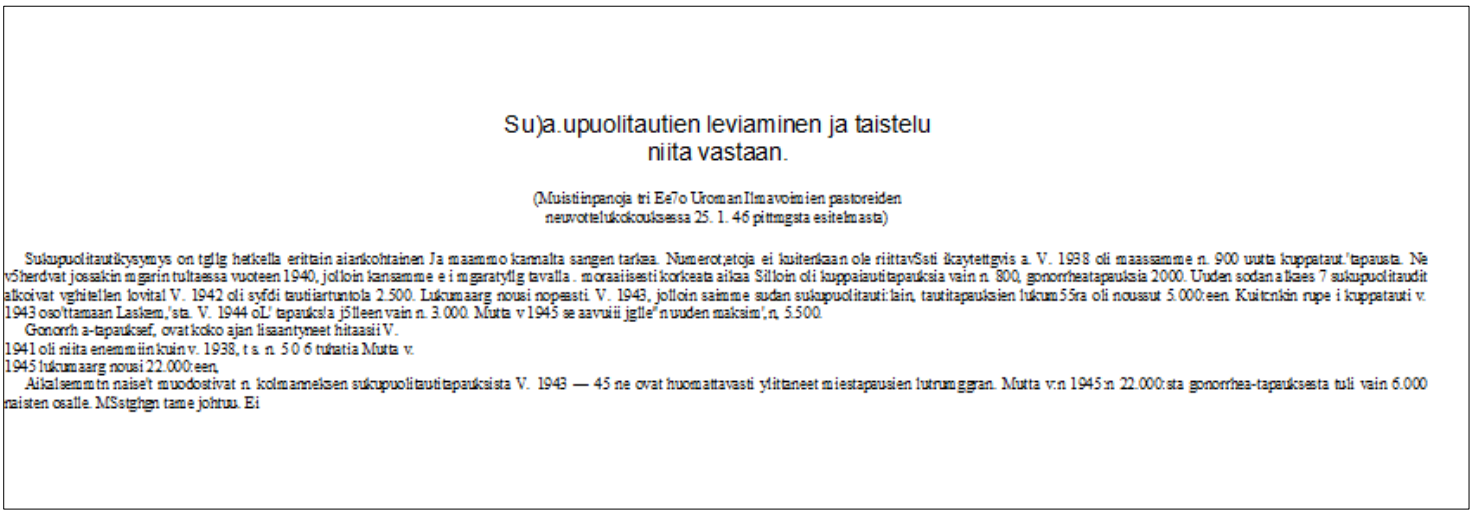
Kuva 17. Adobe Acrobat pro 10 -ohjelman tekstintunnistusnäyte

Tekstintunnistuksen valmistuttua tarvitsi vain tallentaa tiedosto halutun nimiseksi ja vuosikerta oli valmis. Testissä käytettiin testiaineistoa 1 (Taulukko 7, s.32). Virheitä oli useita, eikä ohjelma löytänyt vuosikerran lukuja. Ohjelma kääntää kuvan vain PDF-tiedostoksi, joten virheitä ei voi korjata.

5.2.7 Autobahn DX -ohjelma

Autobahn DX -ohjelmalla /13/ pystyy yhdistämään niin PDF-, kuin TIFF-tiedostoja. Ennen tekstintunnistusta tuli muistaa valita OCR options -valikosta Searchable PDF -vaihtoehto. Sen valitsemalla ohjelma (Kuva 20) ymmärtää tunnistaa tekstin kuvan sisällä.

Lopputulos ei ollut onnistunut, ja vaikka olisi ollutkin, Wordissa (Kuva 18) tekstissä oli paljon virheitä ja PDF:n (Kuva 19) päällä oli vesileima, joka johtuu ohjelman kokeiluversiosta (Taulukko 10, s.33). Vesileima ei silti vaikuttanut päätökseen olla käyttämättä ohjelmaa selvityksessä, sillä lopputulos ei ollut riittävän hyvä.



Su)a.upuolitautien leviaminen ja taistelu
niita vastaan.

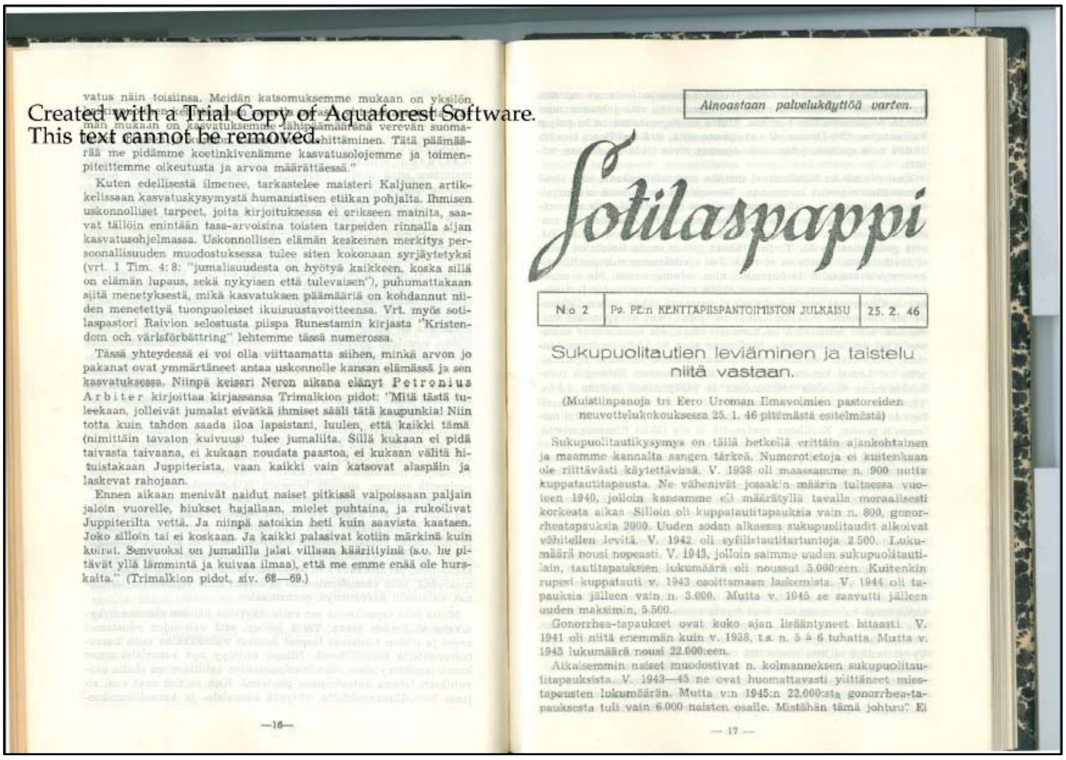
(Muistiinpanoja tri Ee7o UromanIlmavoimien pastoreiden
neuvottelukokouksessa 25. 1. 46 pitmgista esitelmasta)

Sukupuolitautikysymys on tgllg hetkella erittain aiankohtainen Ja maammo kannalta sangen tarkea. Numerot;etoja ei kuiterkaan ole riittavSsti ikaytettgvis a. V. 1938 oli maassamme n. 900 uutta kuppataut.'tapausta. Ne v5herdvat jossakin mgarin tultaessa vuoteen 1940, jolloin kansamme e i mgaratytlg tavalla . moraaiisesti korkeata aikaa Silloin oli kuppaiautitapauksia vain n. 800, gonorrheatapauksia 2000. Uuden sodan alkaes 7 sukupuolitaudit alkoivat vghitellen lovital V. 1942 oli syfdi tautiartuntola 2.500. Lukumaarg nousi nopeasti. V. 1943, jolloin saimme sudan sukupuolitauti:lain, tautitapauksien lukum55ra oli noussut 5.000:een. Kuitcnkin rupe i kuppatauti v. 1943 oso'ttamaan.Laskem,'sta. V. 1944 oL' tapauksia j5lleen vain n. 3.000. Mutta v 1945 se aavuiii jgtle" nuuden maksin',n, 5.500.

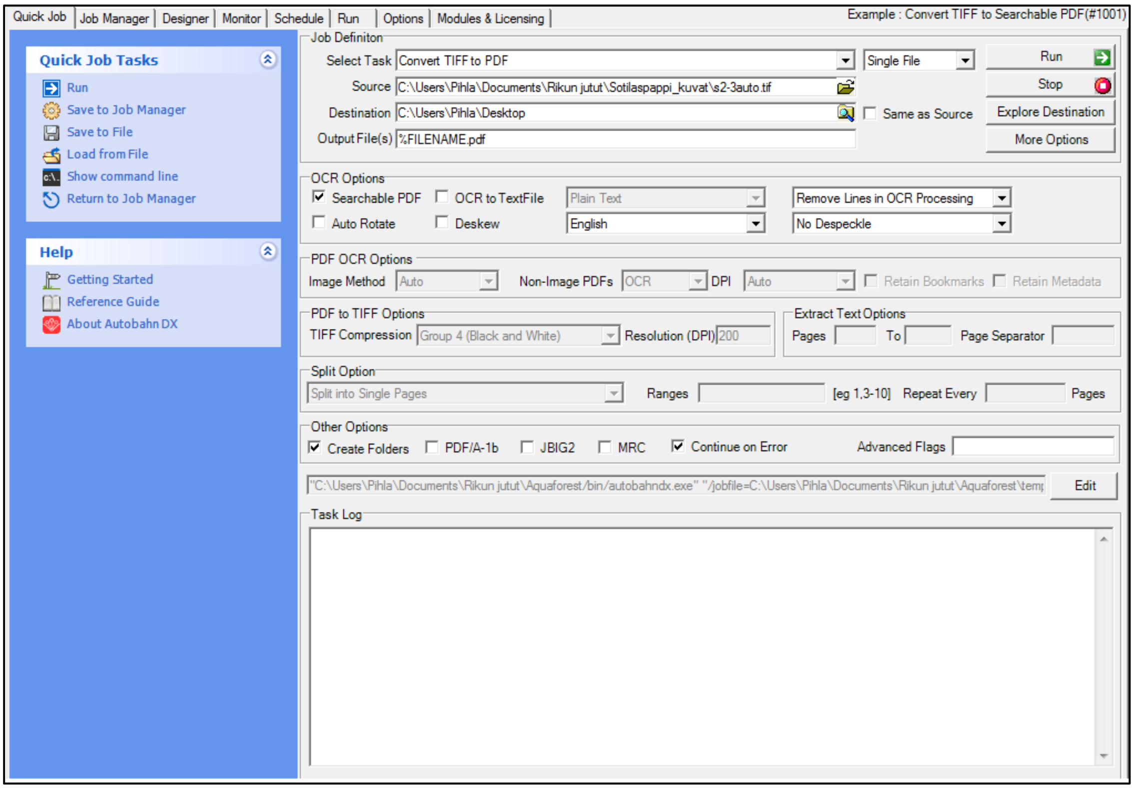
Gonorrh a-tapauksef, ovat koko ajan lisaantyneet hitaasii V.
1941 oli niita enemmiin kuin v. 1938, t s. n. 5 0 6 tuhatia Mutta v.
1945 lukumaarg nousi 22.000:een,

Aikalsemmtn naiset muodostivat n. kolmanneksen sukupuolitautitapauksista V. 1943 — 45 ne ovat huomattavasti yiittaneet miestapausien lutrumggran. Mutta v:n 1945:n 22.000:sta gonorrhea-tapauksesta tuli vain 6.000 naisten osalle. MSstghgn tame johtuu. Ei

Kuva 18. Autobahn DX:n .doc-version tekstintunnistusnäyte



Kuva 19. Autobahn DX:n PDF –version näkymä

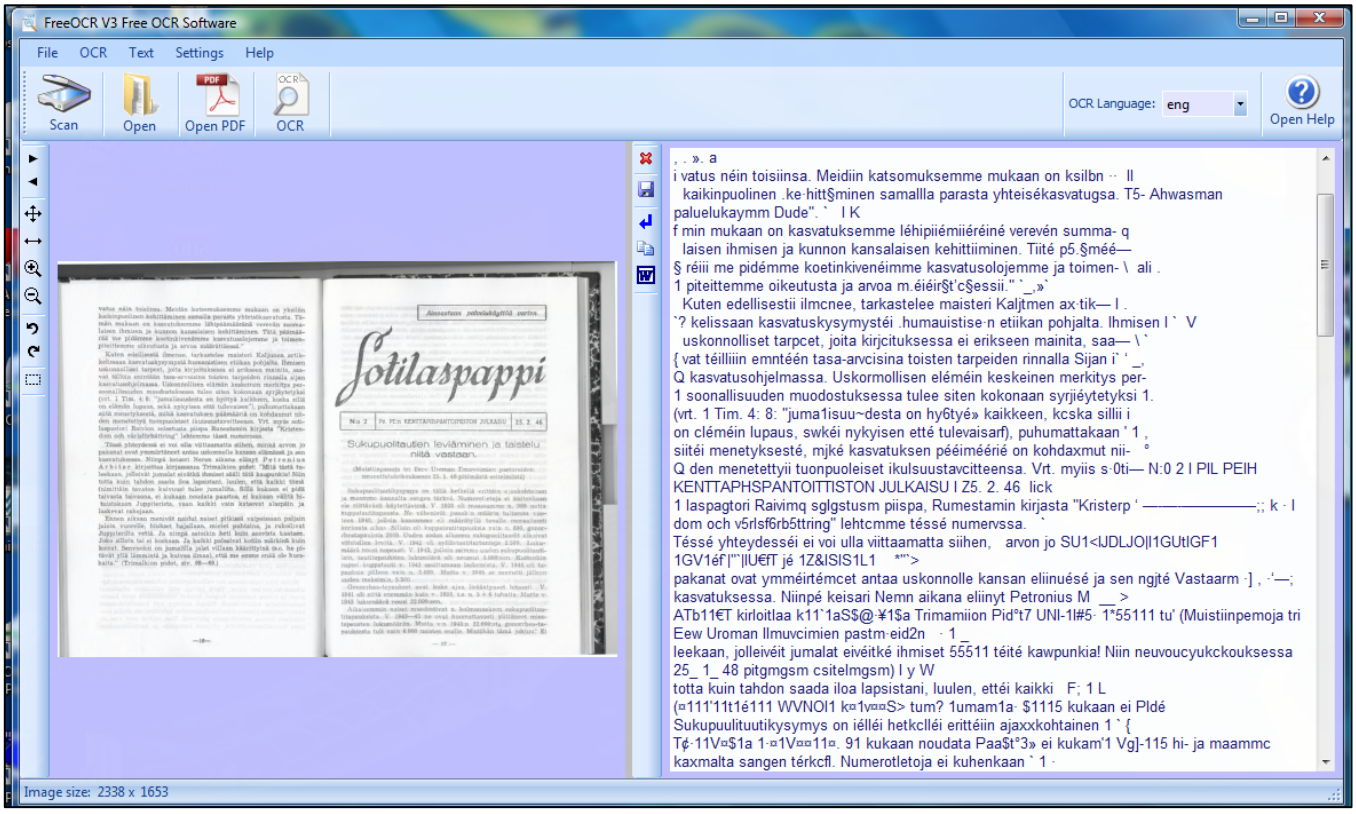


Kuva 20. Autobahn DX –ohjelman näkymä

Testissä käytettiin testiaineistoa 2 (Taulukko 7, s.32). Teksti vaati paljon korjausta, mikä olisi vienyt liikaa aikaa.

5.2.8 Free OCR 3.0 -ohjelma

Free OCR 3.0 -ohjelma /14/ tuki vain englannin kieltä, joten se ei tunnistanut ä- ja ö-kirjaimia (Kuva 21), mutta tunnistustaso ei ole ohjelmassa muutenkaan onnistunutta, joten tämä ongelma ei ole tärkeä.

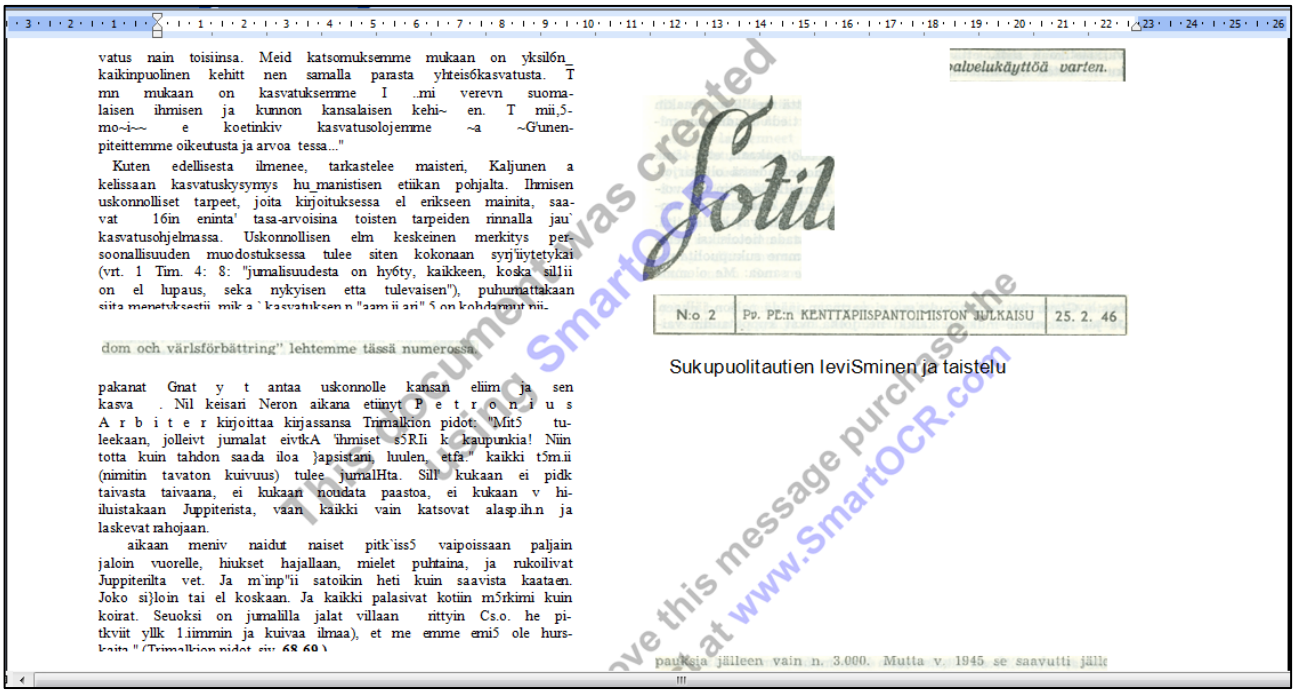


Kuva 21. Free OCR 3.0 -ohjelman tekstintunnistusnäyte

Yhdestä aukeamasta ohjelma tunnisti muutaman sanan. Testissä käytettiin testiaineistoa 2 (Taulukko 6, s.31). Ohjelman laatu ei tunnistuksen perusteella ollut lainkaan riittävä. Tekstin korjaamiseen olisi käytännössä mennyt yhtä kauan kuin koko tekstin uudelleen kirjoittamiseen.

5.2.9 Smart OCR -ohjelma

Smart OCR -ohjelma /15/ siirsi kuvan haluttuun tiedostomuotoon, mutta ei tunnistanut merkkejä riittävästi ja sivustolla lukee selvästi, että ohjelma jättää vesileiman (Taulukko 10, s.33) lopputulokseen (Kuva 22), joten pelkästään sen takia ohjelma ei täytä vaatimuksia.



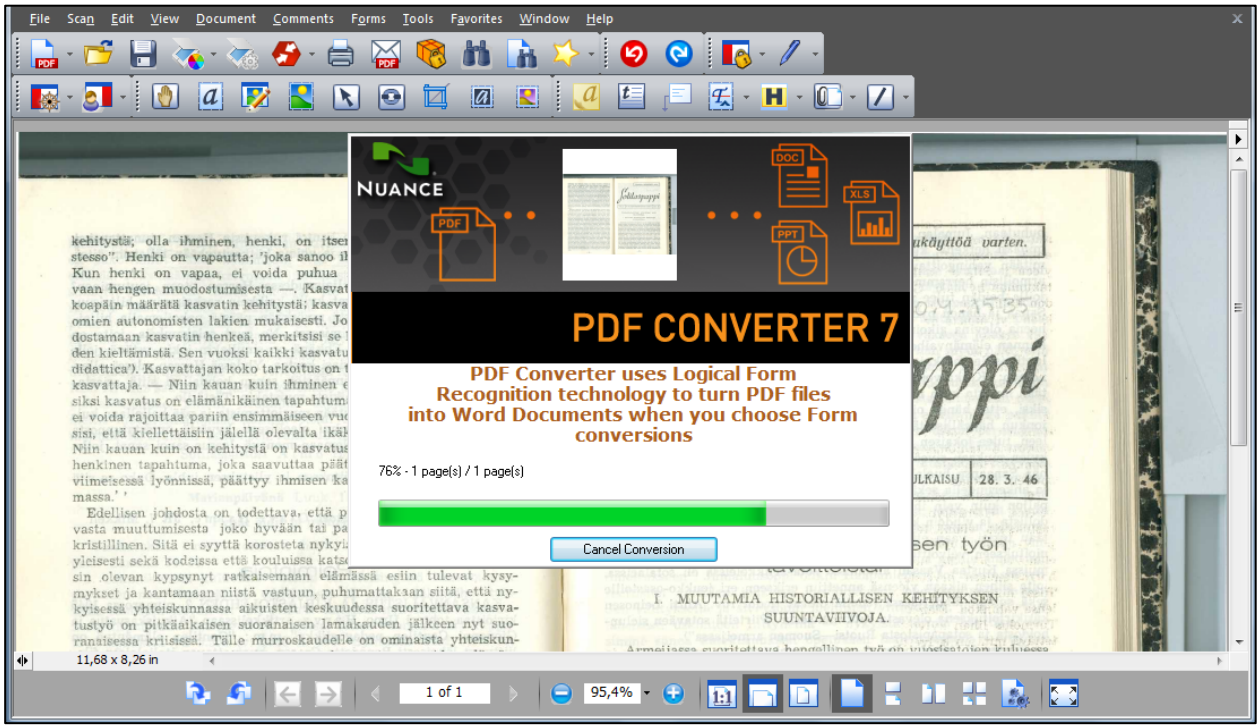
Kuva 22. Smart OCR -ohjelman tekstintunnistusnäyte

Testissä käytettiin testiaineistoa 2 (Taulukko 7, s.32). Tunnistuksen suurin ongelma on se, että ohjelma ei edes tunnista koko tekstiä, vaan parhaimmillaan hieman yli puolet tekstistä tulostuu.

5.2.10 PDF Converter Professional 7 -ohjelma

PDF Converter Professional 7 -ohjelmasta oli tarjolla 30 vuorokauden kokeiluversio /16/.

Lopputulos Word-dokumentiksi muuttamisen jälkeen oli tällä ohjelmalla vaihtelevaa, sillä sanat, joita ohjelma ei Word-dokumentiksi muutettaessa täysin tunnistanut, tunnistettiin PDF-muotoon muutettaessa (Kuva 23). Ohjelman tulostama Word-dokumentti ei tunnistanut ä- eikä ö- kirjaimia, mikä olisi tietenkin suotavaa. Tämä on huono asia, sillä toisin kuin PDF-tiedosto, Word-dokumentti tunnisti harvalla kirjoitetut sanat. PDF-formaatissa ä- ja ö -kirjaimet eivät onneksi ole ongelma, koska PDF tunnistaa ä- ja ö- kirjaimet. Testissä käytettiin testiaineistoa 5 (Taulukko 6, s.31).



Kuva 23. PDF Converter 7 -ohjelman tekstintunnistus

Ohjelma pystyy yhdistämään kuvatiedostot yhdeksi PDF-tiedostoksi, mikä on suositeltavaa ennen tekstintunnistuksen suorittamista. Pakollista tämä ei ollut, mutta nopeutti ja helpotti tunnistusprosessia huomattavasti. Word-tiedostoksi kääntäminen onnistui vasta, kun kuvatiedostot olivat ensin muutettu PDF-muotoon. Ohjelma oli laadultaan selvityksen käytettävin, mutta silti puutteelli-

nen. Ensimmäisessä tarkastelussa virheitä ei huomattu monta, mutta tarkemman tarkastelun jälkeen virheitä löytyi useita.

6 Tulosten esittely

Opinnäytetyön teko aloitettiin skannaamalla Sotilaspappi-aikakausilehden vuosikerta aukeama kerrallaan. Seuraavaksi mietittiin sopiva testiaineisto, jossa olisi tarpeeksi erilaista tekstiä ja erilaisia merkkejä. Vaihtelevuuden ja kattavuuden vuoksi päätestiaineiston lisäksi valittiin vielä neljä testiaineistoa.

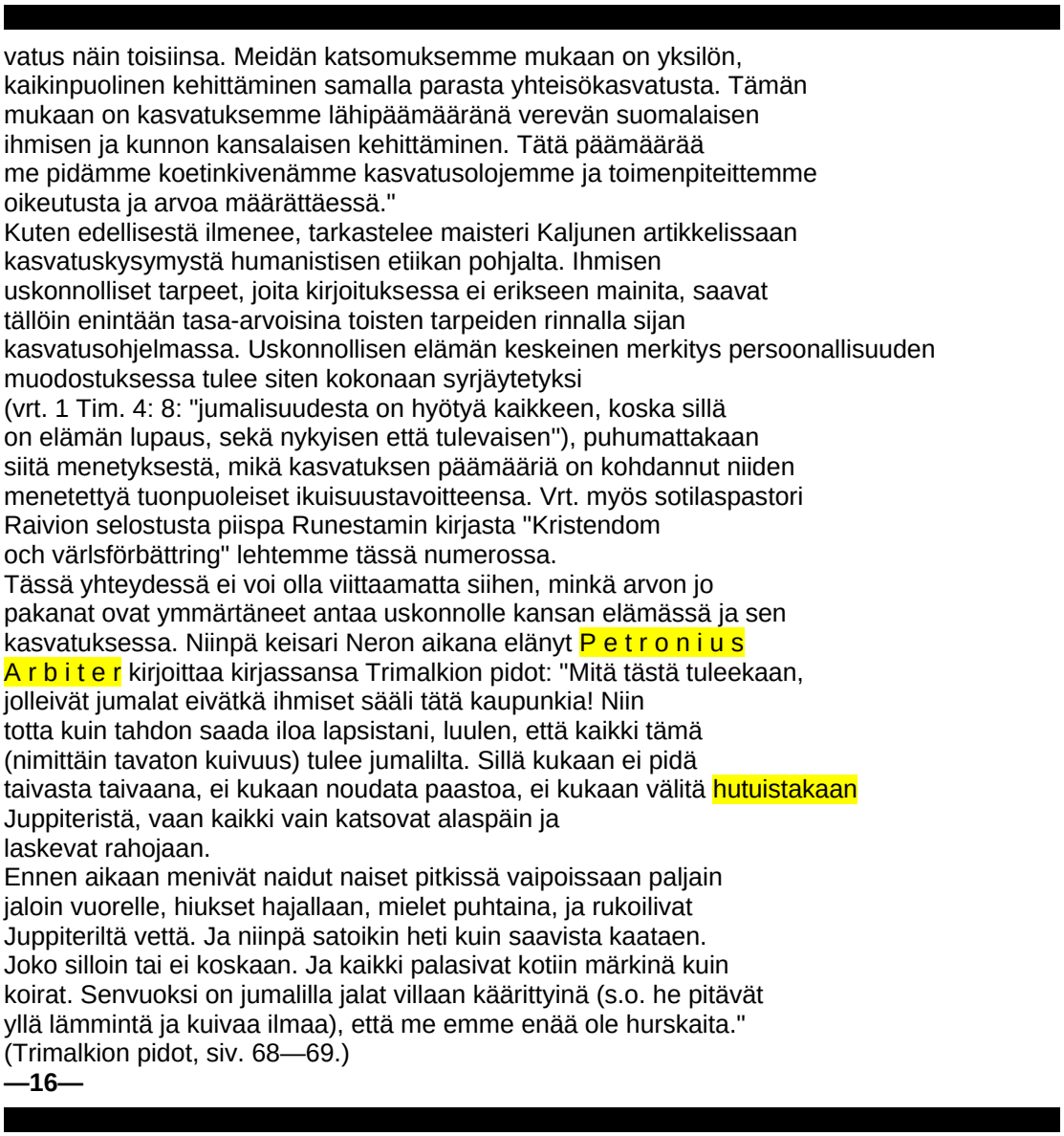
Ohjelmat etsittiin Internetistä ja kaikki löydetyt ohjelmat testattiin, jonka jälkeen selvitettiin paras ohjelma eri ohjelmien virheiden määriä vertaamalla. Paras ohjelma oli sellainen, jossa oli vertailun jälkeen pienin määrä virheitä. Käyttökelpoisin vaihtoehto oli PDF Converter professional 7 -ohjelma, joka asennettiin työasemalle. Ohjelmalla tehtiin vuosikerran kuvatiedostoja yhdistämällä yksi PDF, jonka jälkeen valittiin ohjelmasta tekstintunnistus, joka tunnisti PDF:stä merkit. Tämän jälkeen tallennettiin vuosikerta-PDF ja nimettiin se (Sotilaspappi vuosikerta 46).

Testimateriaalina toimiva aikakausilehden vuosikerta oli käsin ladottua tekstiä, joka aiheutti sanoissa epätasaisuutta. Toinen vaikeuksia tuottava ongelma oli harvaan kirjoitetut sanat, joiden tunnistaminen osoittautui ongelmalliseksi. Kyseessä oli PDF-tiedosto, joten virheiden korjaaminen oli mahdotonta.

Vertailin tuloksia kolmen eri ohjelman välillä. Testiaineistona käytettiin testiaineisto 2:n (Taulukko 7, s. 32) ensimmäistä sivua. Vertailun ohjelmat valittiin niistä ohjelmista, jotka kääntävät lähdeaineiston suoraan PDF -tiedostoksi. Vertailussa olivat selvityksen parhaiten toimivaksi ohjelmaksi osoittautunut Abbyy Finereader 10 -ohjelma, selvityksen käyttökelpoisimmaksi ohjelmaksi osoittautunut PDF Converter Professional 7 -ohjelma ja Adobe acrobat pro 10 -ohjelma. Virheiksi ei merkitty ä- ja ö-pisteiden puuttumista, koska ne eivät estäneet sanojen löytymistä haussa.

Seuraavanlaisesti tulostui käännös (Kuva 24) Finereader 10 -ohjelmalla (k. 5.2.5, s. 34), Punaisella ovat virheet, joissa ohjelma on jättänyt sanasta tunnistamatta tai tunnistanut väärin enemmän kuin yhden merkin. Keltaisella ovat

virheet, joissa ohjelma on tunnistanut sanasta tai tunnistanut väärin vain yhden merkin. Punaisia virheitä löytyi 0 kappaletta. Keltaisia virheitä löytyi 3 kappaletta.



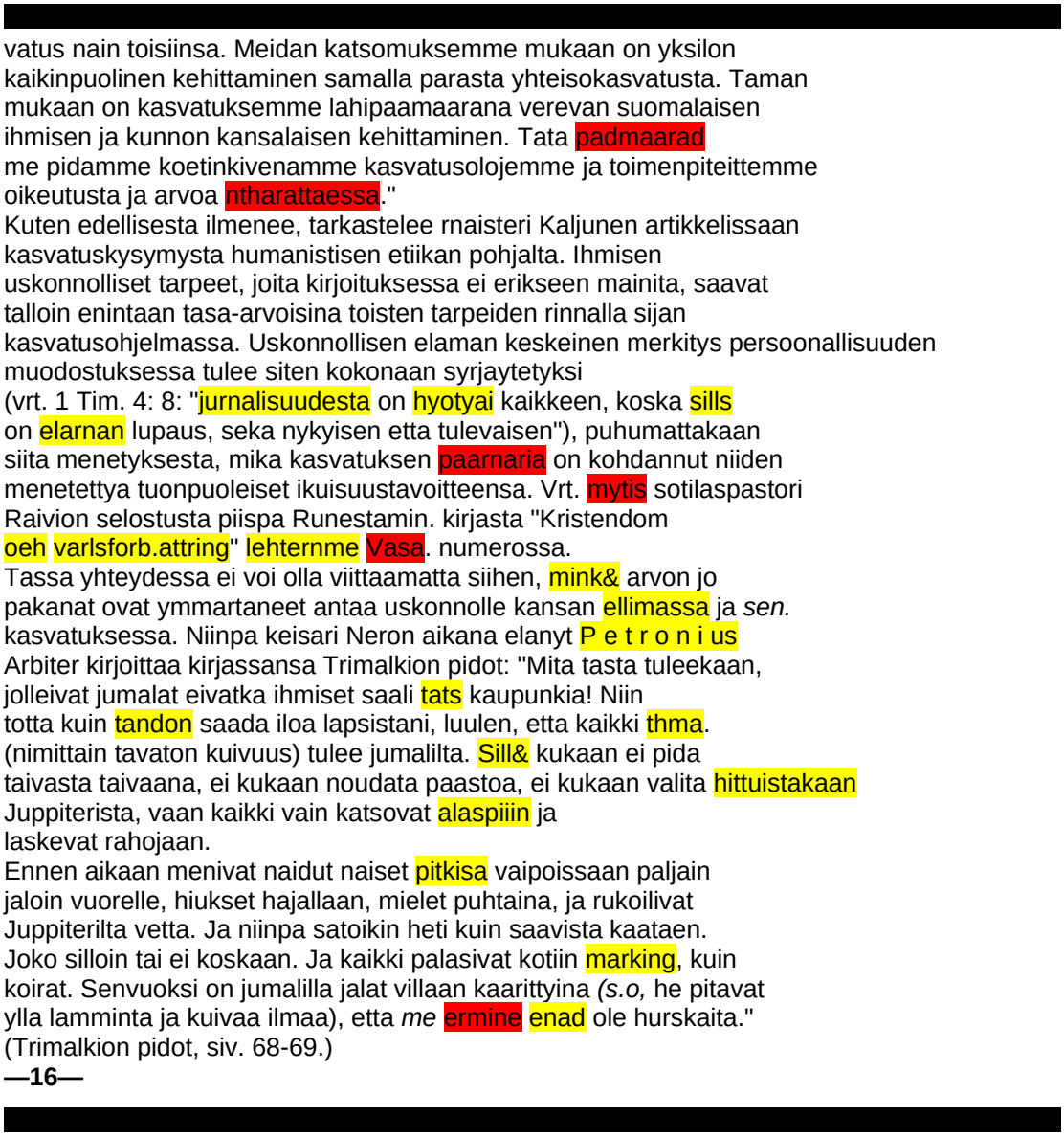
vatus näin toisiinsa. Meidän katsomuksemme mukaan on yksilön,
kaikinpuolinen kehittäminen samalla parasta yhteisökasvatusta. Tämän
mukaan on kasvatuksemme lähipäämääränä verevän suomalaisen
ihmisen ja kunnon kansalaisen kehittäminen. Tätä päämäärää
me pidämme koetinkivenämme kasvatusolojemme ja toimenpiteittemme
oikeutusta ja arvoa määrättäessä."
Kuten edellisestä ilmenee, tarkastelee maisteri Kaljunen artikkelissaan
kasvatuskysymystä humanistisen etiikan pohjalta. Ihmisen
uskonnolliset tarpeet, joita kirjoituksessa ei erikseen mainita, saavat
tällöin enintään tasa-arvoisina toisten tarpeiden rinnalla sijan
kasvatusohjelmassa. Uskonnollisen elämän keskeinen merkitys persoonallisuuden
muodostuksessa tulee siten kokonaan syrjäytetyksi
(vrt. 1 Tim. 4: 8: "jumalisuudesta on hyötyä kaikkeen, koska sillä
on elämän lupaus, sekä nykyisen että tulevaisen"), puhumattakaan
siitä menetyksestä, mikä kasvatuksen päämääriä on kohdannut niiden
menetettyä tuonpuoleiset ikuisuustavoitteensa. Vrt. myös sotilaspastori
Raivion selostusta piispa Runestamin kirjasta "Kristendom
och värlsförbättring" lehtemme tässä numerossa.
Tässä yhteydessä ei voi olla viittaamatta siihen, minkä arvon jo
pakanat ovat ymmärtäneet antaa uskonnolle kansan elämässä ja sen
kasvatuksessa. Niinpä keisari Neron aikana elänyt P e t r o n i u s
A r b i t e r kirjoittaa kirjassansa Trimalkion pidot: "Mitä tästä tuleekaan,
jolleivät jumalat eivätkä ihmiset sääli tätä kaupunkia! Niin
totta kuin tahdon saada iloa lapsistani, luulen, että kaikki tämä
(nimittäin tavaton kuivuus) tulee jumalilta. Sillä kukaan ei pidä
taivasta taivaana, ei kukaan noudata paastoa, ei kukaan välitä hutuistakaan
Juppiteristä, vaan kaikki vain katsovat alaspäin ja
laskevat rahojaan.
Ennen aikaan menivät naidut naiset pitkissä vaipoissaan paljain
jaloin vuorelle, hiukset hajallaan, mielet puhtaina, ja rukoilivat
Juppiteriltä vettä. Ja niinpä satoikin heti kuin saavista kaataen.
Joko silloin tai ei koskaan. Ja kaikki palasivat kotiin märkinä kuin
koirat. Senvuoksi on jumalilla jalat villaan käärittyinä (s.o. he pitävät
yllä lämmintä ja kuivaa ilmaa), että me emme enää ole hurskaita."
(Trimalkion pidot, siv. 68—69.)
—16—

Kuva 24. Finereader 10 -ohjelman tekstintunnistusnäyte

Ohjelma oli selvästi paras vaihtoehto työn suoritukseen, mutta rajoitustensa (Taulukko 10, s. 33) takia ohjelmaa ei voitu käyttää.

Seuraavanlaisesti tulostui käännös (Kuva 25) PDF Converter 7 -ohjelmalla (k. 5.2.10, s. 40), Punaisella ovat virheet, joissa ohjelma on jättänyt sanasta tunnistamatta tai tunnistanut väärin enemmän kuin yhden merkin. Keltaisella ovat

virheet, joissa ohjelma on tunnistanut sanasta tai tunnistanut väärin vain yhden merkin. Punaisia virheitä löytyi 6 kappaletta. Keltaisia virheitä löytyi 19 kappaletta.



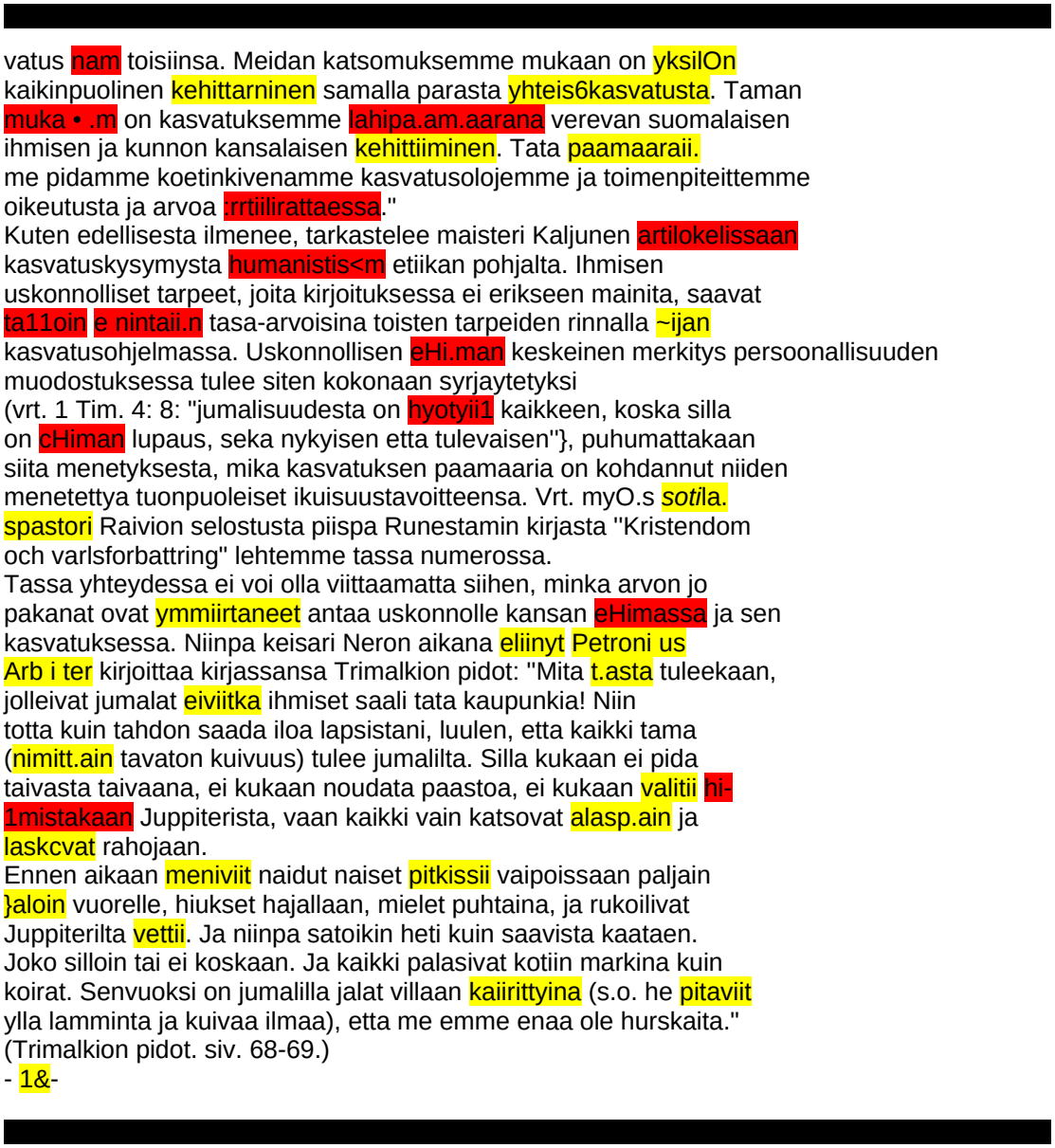
vatus nain toisiinsa. Meidan katsomuksemme mukaan on yksilon
kaikinpuolinen kehittaminen samalla parasta yhteisokasvatusta. Taman
mukaan on kasvatuksemme lahipaamaarana verevan suomalaisen
ihmisen ja kunnon kansalaisen kehittaminen. Tata padmaarad
me pidamme koetinkivenamme kasvatusolojemme ja toimenpiteittemme
oikeutusta ja arvoa ntharattaessa."
Kuten edellisesta ilmenee, tarkastelee rnaisteri Kaljunen artikkelissaan
kasvatuskysymysta humanistisen etiikan pohjalta. Ihmisen
uskonnolliset tarpeet, joita kirjoituksessa ei erikseen mainita, saavat
talloin enintaan tasa-arvoisina toisten tarpeiden rinnalla sijan
kasvatusohjelmassa. Uskonnollisen elaman keskeinen merkitys persoonallisuuden
muodostuksessa tulee siten kokonaan syrjaytetyksi
(vrt. 1 Tim. 4: 8: "jurnalisuudesta on hyotyai kaikkeen, koska sills
on elarnan lupaus, seka nykyisen etta tulevaisen"), puhumattakaan
siita menetyksesta, mika kasvatuksen paarnaria on kohdannut niiden
menetettya tuonpuoleiset ikuisuustavoitteensa. Vrt. mytis sotilaspastori
Raivion selostusta piispa Runestamin. kirjasta "Kristendom
oeh varlsforb.attring" lehternme Vasa. numerossa.
Tassa yhteydessa ei voi olla viittaamatta siihen, mink& arvon jo
pakanat ovat ymmartaneet antaa uskonnolle kansan ellimassa ja *sen.*
kasvatuksessa. Niinpa keisari Neron aikana elanyt P e t r o n i us
Arbiter kirjoittaa kirjassansa Trimalkion pidot: "Mita tasta tuleekaan,
jolleivat jumalat eivatka ihmiset saali tats kaupunkia! Niin
totta kuin tandon saada iloa lapsistani, luulen, etta kaikki thma.
(nimittain tavaton kuivuus) tulee jumalilta. Sill& kukaan ei pida
taivasta taivaana, ei kukaan noudata paastoa, ei kukaan valita hittuistakaan
Juppiterista, vaan kaikki vain katsovat alaspiiin ja
laskevat rahojaan.
Ennen aikaan menivat naidut naiset pitkisa vaipoissaan paljain
jaloin vuorelle, hiukset hajallaan, mielet puhtaina, ja rukoilivat
Juppiterilta vetta. Ja niinpa satoikin heti kuin saavista kaataen.
Joko silloin tai ei koskaan. Ja kaikki palasivat kotiin marking, kuin
koirat. Senvuoksi on jumalilla jalat villaan kaarittyina *(s.o,* he pitavat
ylla lamminta ja kuivaa ilmaa), etta *me* ermine enad ole hurskaita."
(Trimalkion pidot, siv. 68-69.)
—16—

Kuva 25. PDF Converter 7 -ohjelman tekstintunnistusnäyte

Ohjelma oli virheineen selvityksen kelpuutettavin ohjelma, jonka rajoitukset eivät estäneet työn toteuttamista.

Seuraavanlaisesti tulostui käännös (Kuva 26) Adobe Acrobat pro 10 -ohjelmalla (k. 5.2.6, s. 35), Punaisella ovat virheet, joissa ohjelma on jättänyt sanasta tunnistamatta tai tunnistanut väärin enemmän kuin yhden merkin. Keltaisella ovat

virheet, joissa ohjelma on tunnistanut sanasta tai tunnistanut väärin vain yhden merkin. Punaisia virheitä löytyi 14 kappaletta. Keltaisia virheitä löytyi 25 kappaletta.



vatus nam toisiinsa. Meidan katsomuksemme mukaan on yksilOn
kaikinpuolinen kehittarninen samalla parasta yhteis6kasvatusta. Taman
muka • .m on kasvatuksemme lahipa.am.aarana verevan suomalaisen
ihmisen ja kunnon kansalaisen kehittiiminen. Tata paamaaraii.
me pidamme koetinkivenamme kasvatusolojemme ja toimenpiteittemme
oikeutusta ja arvoa :rrtiilirattaessa."
Kuten edellisesta ilmenee, tarkastelee maisteri Kaljunen artilokelissaan
kasvatuskysymysta humanistis<m etiikan pohjalta. Ihmisen
uskonnolliset tarpeet, joita kirjoituksessa ei erikseen mainita, saavat
ta11oin e nintaii.n tasa-arvoisina toisten tarpeiden rinnalla ~ijan
kasvatusohjelmassa. Uskonnollisen eHi.man keskeinen merkitys persoonallisuuden
muodostuksessa tulee siten kokonaan syrjaytetyksi
(vrt. 1 Tim. 4: 8: "jumalisuudesta on hyotyii1 kaikkeen, koska silla
on cHiman lupaus, seka nykyisen etta tulevaisen"}, puhumattakaan
siita menetyksesta, mika kasvatuksen paamaaria on kohdannut niiden
menetettya tuonpuoleiset ikuisuustavoitteensa. Vrt. myO.s *soti*la.
spastori Raivion selostusta piispa Runestamin kirjasta "Kristendom
och varlsforbattring" lehtemme tassa numerossa.
Tassa yhteydessa ei voi olla viittaamatta siihen, minka arvon jo
pakanat ovat ymmiirtaneet antaa uskonnolle kansan eHimassa ja sen
kasvatuksessa. Niinpa keisari Neron aikana eliinyt Petroni us
Arb i ter kirjoittaa kirjassansa Trimalkion pidot: "Mita t.asta tuleekaan,
jolleivat jumalat eiviitka ihmiset saali tata kaupunkia! Niin
totta kuin tahdon saada iloa lapsistani, luulen, etta kaikki tama
(nimitt.ain tavaton kuivuus) tulee jumalilta. Silla kukaan ei pida
taivasta taivaana, ei kukaan noudata paastoa, ei kukaan valitii hi-
1mistakaan Juppiterista, vaan kaikki vain katsovat alasp.ain ja
laskcvat rahojaan.
Ennen aikaan meniviit naidut naiset pitkissii vaipoissaan paljain
}aloin vuorelle, hiukset hajallaan, mielet puhtaina, ja rukoilivat
Juppiterilta vettii. Ja niinpa satoikin heti kuin saavista kaataen.
Joko silloin tai ei koskaan. Ja kaikki palasivat kotiin markina kuin
koirat. Senvuoksi on jumalilla jalat villaan kaiirittyina (s.o. he pitaviit
ylla lamminta ja kuivaa ilmaa), etta me emme enaa ole hurskaita."
(Trimalkion pidot. siv. 68-69.)
- 1&-

Kuva 26. Adobe Acrobat pro 10 -ohjelman tekstintunnistusnäyte

Ohjelmaa ei virheiden määrän takia voitu valita selvityksen käyttökelpoisimmaksi.

Selvityksessä testatuista ohjelmista vertailtu Finereader 10 -ohjelma oli toimivin, mutta rajoitustensa (Taulukko 10, s. 33) takia ohjelman valitseminen työhön ei ollut mahdollista. Selvityksessä testatuista ohjelmista vertailtu PDF Con-

verter 7 -ohjelma oli puutteineen käyttökelpoisin, minkä takia ohjelma valittiin vuosikerran käsittelyyn.

7 Yhteenveto ja päätelmät

Työssä pääsin selvittämään, kuinka saada mahdollisimman tarkka digitaalinen versio käsin ladotusta kirjasta. Voi olla, että tulevaisuudessa tarvitaan tai halutaan tehdä toisesta kirjasta digitaalinen versio, josta on tarvetta hakea sanoja. Lopputuloksena sain PDF Converter 7 -ohjelmalla vuosikerran yhteen PDF-tiedostoon, johon pystyy tekemään sanahakuja. Ongelma lopputuloksessa oli harvaan kirjoitetut sanat, joita käytetyllä ohjelmalla oli vaikeuksia havaita. Vaikeuksia oli myös monen muun sanan kohdalla, mutta käsiteltäessä lähdetiedosto PDF-tiedostoksi, PDF Converter 7 oli käytettävissä olevista vaihtoehdoista käyttökelpoisin. Yritin selvittää, miten ohjelman saisi huomaamaan sanat, joita se ei tunnistanut tekstintunnistuksessa, mutta en saanut tähän kysymykseen vastausta. Word-muotoon muutettujen tiedostojen virheet on korjattava käsin, mutta parhaimmilla ohjelmilla käsiteltynä korjaukseen ei paljoa aikaa mene virheiden pienen määrän ansiosta. Haluttaessa löytää vuosikerrasta esimerkiksi luku 3, on hakukenttään kirjoitettava ”N:o 3”, jotta ohjelma löytää kyseisen luvun. Haluttaessa löytää luku päivämäärällä, on hakukenttään kirjoitettava esimerkiksi ”28. 3. 46” siten, että pisteen jälkeen tulee väli.

Tietokantavaihtoehdon poissulkeminen oli ikävää, mutta yritys saada siitä toimiva olisi ollut liian suuritöinen ja ehkä loppujen lopuksi myös turha. Vaikka työn ohessa ilmeni omat ongelmansa, oli ohjelmien erilaisuus ja laatu mielenkiintoista selvitettävää.

Työasemalle ladattavien ohjelmien tulokset olivat pääasiassa parempia - eli niissä oli vähemmän virheitä - kuin selaimessa toimivat ohjelmat. Poikkeustapauksena oli selaimessa toimiva OCR Terminal -ohjelma, jonka lopputulos oli samaa tasoa työasemalle asennettavan Abbyy Finereader 10:n kanssa. Ilman rajoituksiaan (Taulukko 10, s. 33) eli rajattua käyttöaikaa ja rajattua tiedostonkäsittelymäärää Finereader 10 olisi ollut paras vaihtoehto projektin toteutukseen.

Testiaineistoista suurimmassa käytössä oli testiaineisto 2, mutta vaihtelevuuden ja kattavuuden saavuttamiseksi käytettiin myös neljää (4) eri testiaineistoa. Testiaineisto 2 valittiin pääkäyttöön sen monipuolisen sisällön takia. Monipuolisella sisällöllä tarkoitetaan eri tavalla kirjoitettuja sanoja. Testiaineistosta löytyi harvalla kirjasimella kirjoitettuja sanoja, eri tasossa olevia sanoja, lihavoituja sanoja ja myös numeroita.

Kuvat

Taulukot

Lähteet

1. Tekstien digitointi http://www.digiwiki.fi/fi/index.php?title=Tekstien_digitointi (luettu 2.5.2011)
2. Free OCR http://www.free-ocr.com/ (luettu 3.5.2011)
3. Free Online OCR http://www.free-online-ocr.com/ (luettu 3.5.2011)
4. i2OCR http://www.sciweavers.org/free-online-ocr (luettu 4.5.2011)
5. New OCR http://www.newocr.com/ (luettu 4.5.2011)
6. OCR Convert http://www.ocrconvert.com/ (luettu 5.5.2011)
7. OCR Now! https://my.ocrnow.com/ (luettu 5.5.2011)
8. OCR Online http://www.ocronline.com/ (luettu 6.5.2011)
OCR terminal https://www.ocrterminal.com/ (luettu 6.5.2011)
(myöhemmin:
9. http://finereader.abbyyonline.com/en/Account/OcrTerminalWelcome) (luettu 19.7.2011)
10. Abbyy http://finereader.abbyy.com/ (luettu 11.7.2011)
11. Merge PDF http://foxyutils.com/mergepdf/# (luettu 14.7.2011)
12.Acrobat
https://www.adobe.com/cfusion/tdrc/index.cfm?product=acrobat_pro&loc=en
(luettu 14.7.2011)
13. Autobahn http://www.aquaforest.com/en/autobahn.asp (luettu 18.7.2011)
14. Free OCR http://www.freewarefiles.com/Free-OCR_program_34315.html (luettu 19.7.2011)
15. Smart OCR http://www.smartocr.com/ (luettu 20.7.2011)
16. Nuance PDF Converter http://www.nuance.com/products/pdf-converter-professional7/index.htm (luettu 10.8.2011)