

Navodila za program Zajem

O programu

Zajem je odprtokodni program za skeniranje, vpis atributov in izvoz dokumentov v obliki TIFF ali PDF ter izvozno datoteko, ki omogoča uvoz v arhivske sisteme.

Skeniranje deluje s pomočjo paketa SANE tako da program deluje samo na LINUX operacijskem sistemu.

Napisan je v programskem jeziku python3, za grafični vmesnik uporablja PyQt4, za podatkovno bazo v kateri so shranjene nastavitve program in trenutni dokumenti pa SQLite.

Nastal je na podlagi večletne uporabe komercialnih programov za skeniranje. Njegove prednosti, če izvzamemo ceno, so zelo enostavno povezovanje listov v dokument in enostavno dodajanje modulov za izvoz in branje podatkov dokumenta iz podatkovne baze, saj so moduli prav tako napisani v python-u.

Program omogoča tudi prepoznavanje besedila (tesseract,cuneiform) in črtne kode (zbarimg).

Program je v fazi testiranja. Delovanje je zaenkrat preverjeno na operacijskem sistemu Fedora 19 in skenerju Canon DR-9080C.

Če odkrijete kakšno napako jo lahko sporočite na mladen.kuntner@triera.net in jo bom poskušal odpraviti.

Če program uspešno uporabljate pa bi vas prosil da na zgornji naslov sporočite tip skenerja.

Licenca

Licenca programa je GNU General Public License 3 ali na kratko GPLv3.

Več o tem na:

<http://www.gnu.org/copyleft/gpl.html>

Namestitev

Pogram za svoje delovanje potrebuje:

- python3 (delovanje programa)
- pyqt4 (uporabniški vmesnik)
- sqlite3 (podatkovna baza)
- sane , sane-backends (skeniranje)
- sane-backends-drivers-scanners (gonilniki za skener)
- netpbm (obdelava skena)
- libtiff (združitev več skenov v en tiff in pdf)

Opcijsko za prepoznavanje besedila:

- tesseract in/ali cuneiform

Opcijsko za prepoznavanje črtne kode:

- zbarimg

Namestitev:

- naredite novo mapo za program
- arhiv programa razširite v narejeno mapo
- zaženete program - python3 Zajem.py
- prijavite se kot privzeti uporabnik (administrator , zajem)
- končate program
- premaknete program DemoDatabaseModul.py v mapo PluginsDatabase
- premaknete program DemoExportModul.py v mapo PluginsExport

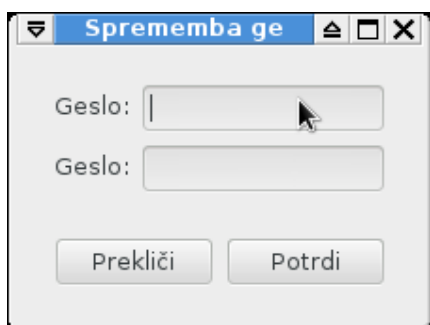
Namestitev programa je s tem končana.

Program je ob prvem zagonu naredil vse potrebne mape in podatkovno bazo.

Uporaba

V tem poglavju bom poskusil napisati, ne samo kako se izvede določena akcija v programu ampak tudi zakaj.

Zamenjava gesla – Nastavitve -> Spremeni geslo



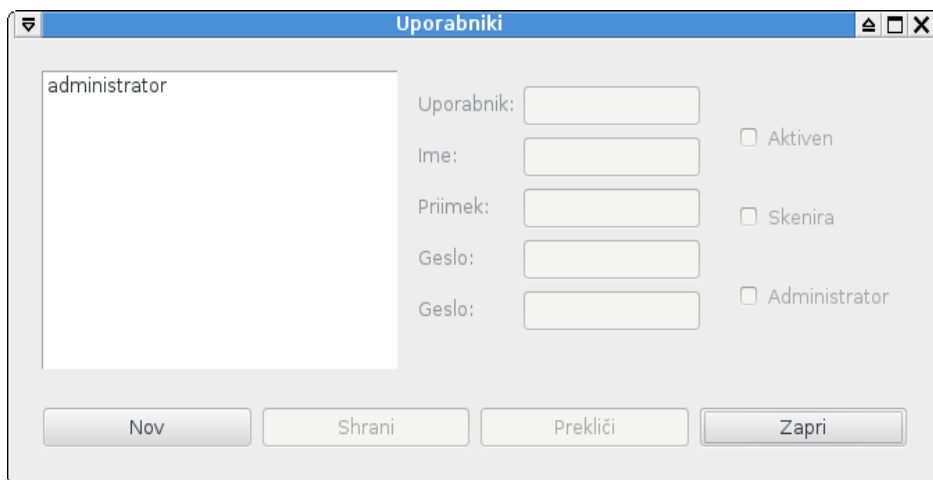
Tukaj si lahko uporabniki zamenjajo geslo, ki jim ga je dodelil administrator programa. Uporabniško ime in geslo v tem programu ne služita varnosti podatkov ampak bolj identifikacija uporabnika za delo s programom. Kljub temu se uporabniško geslo v bazo shranjuje kot soljen hash (SHA256) saj imajo uporabniki slabo navado, da isto geslo uporabljajo za več stvari.

! Da si ne bi uporabniki programa ali kakšna druga oseba dodelila pravice je važno, da ob prvi prijavi zamenjamo geslo uporabnika »administrator«.

Geslo zamenjamo tako, da dvakrat vpišemo novo geslo in zamenjavo potrdimo s klikom na gumb »Potrdi«.

Če si premislimo lahko akcijo prekinemo s klikom na gumb »Prekliči«.

Upravljanje z uporabniškimi računi – Nastavive -> Uporabniki



V seznamu na levi strani okna vidimo vse uporabnike programa. Polja za vpis podatkov so onemogočena dokler ne kliknemo na gumb »Nov« ali izberemo uporabnika iz seznama s klikom na njegovo uporabniško ime.

Na desni strani okna so pravice uporabnika in pomenijo sledeče:

Aktiven – uporabnik se lahko prijavi v program.

Skenira – uporabnik lahko skenira, vpisuje attribute dokumentov, povezuje liste in izvozi dokumente.

Administrator – uporabnik lahko spreminja nastavitve za uporabnike, profile skeniranja in tipe dokumentov.

Vidimo da so uporabniki ločeni na administratorje in uporabnike programa. Izkušnje so pokazale, da je primerno, da imajo uporabniki, ki skenirajo na voljo čim manj možnosti in s tem manj možnosti za morebitne napake.

Potek najbolj pogostih akcij:

Kreiranje novega uporabnika

Kliknemo gumb »Nov«, vpišemo podatke uporabnika v sredinskem stolpcu (vsi podatki so obvezni), dodelimo mu pravice v desnem stolpcu, potrdimo s klikom na gumb »Shrani«, uporabnik se doda v seznam.

Brisanje obstoječega uporabnika

! Uporabnikov v programu ni možno brisati, dosežemo pa to tako, da odstranimo kljukico v polju »Aktiven«

Kliknemo na uporabnika, kliknemo v polje »Aktiven (odstanimo kljukico), kliknemo gumb »Shrani«

Sprememba gesla obstoječega uporabnika

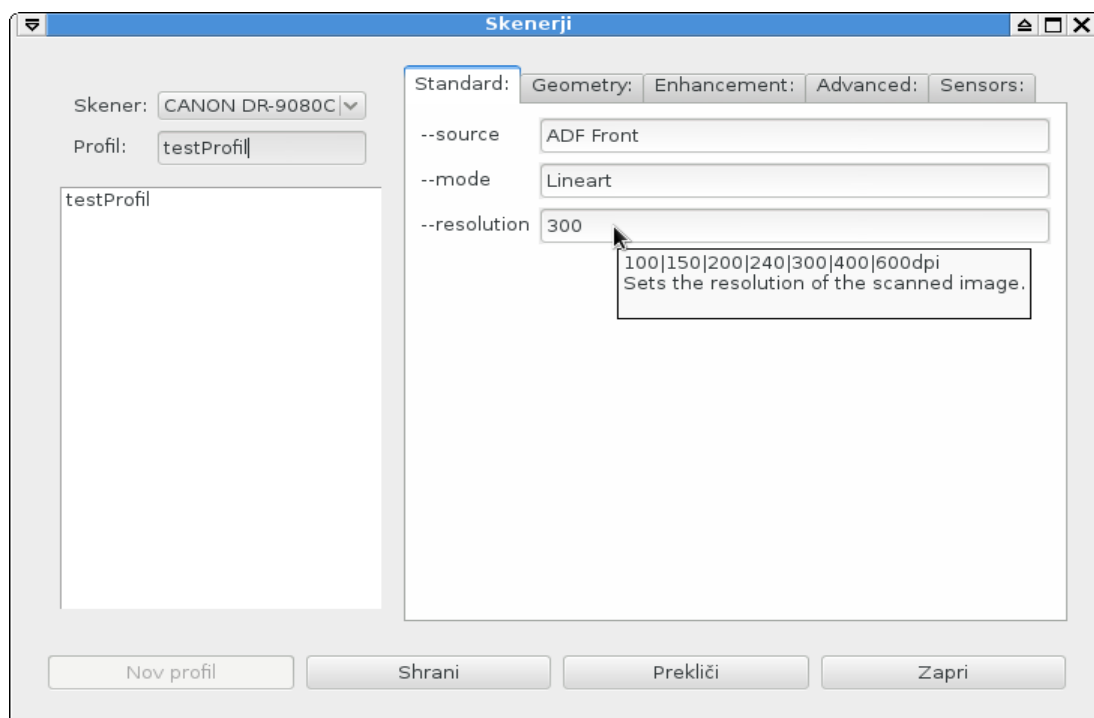
V primeru, da je uporabnik pozabil soje geslo mu ga lahko administrator programa spremeni.

Kliknemo na uporabnika, dvakrat vpišemo novo geslo, kliknemo gumb »Shrani«

Sprememba pravic obstoječega uporabnika

Kliknemo na uporabnika, spremenimo pravice v desnem stolpcu, kliknemo gumb »Shrani«

Upravljanje s skenerji in profili skeniranja – Nastavitve -> Skenerji



! Program omogoča delo samo z enim skenerjem. V podatkovni bazi je lahko več skenerjev z različnimi profili vendar lahko uporabljamo le enega naenkrat.

!! Skener mora biti ob nastavljanju profilov priključen in aktiven.

Profil skeniranja – so vse nastavitve za skener (barvna globina, kvaliteta skeniranja, velikost skenirane strani, svetlost, kontrast ...). Pri tipih dokumentov samo nastavimo skener in profil skeniranja in s tem določimo način skeniranja za posamezni dokument, uporabniku pa onemogočimo možnost napačne izbire.

Zavihki na desni strani niso prevedeni, ker vsi podatki – tako imena zavihkov kot opcije in besedilo pomoči pridejo direktno iz programa SANE. Ker so odvisni od Proizvajalca in modela skenerja v večini primerov tudi ne bodo takšni kot so na sliki.

Možnosti posamezni opcij se ne da izbirati ampak jih moramo sami vpisati v polje. Da je ta postopek lažji postavimo miškin kazalec na polje, ki ga želimo spremeniti in pojavi se besedilo z možnimi vnosi za to opcijo in kratkim opisom opcije (primer na sliki za resolucijo skeniranja).

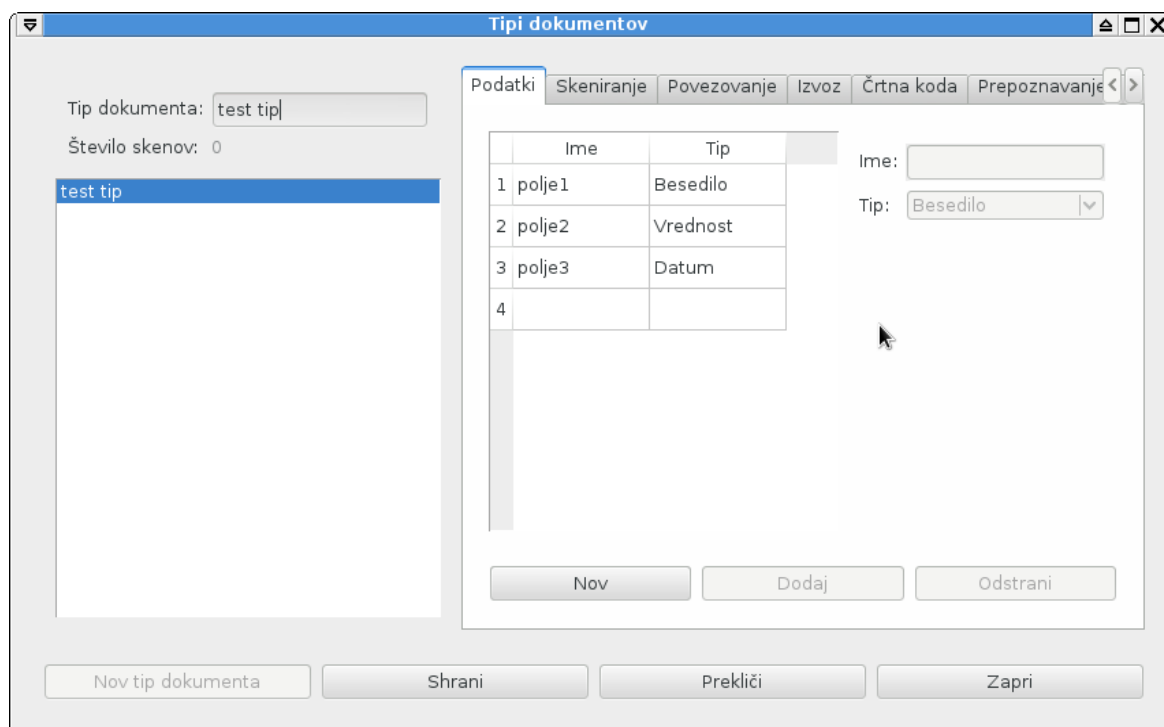
Kreiranje novega profila

Če imamo več kot en skener izberemo skener. V polje »Profil« vpišemo naziv profila. Spremenimo nastavitve skeniranja v zavihkih na desni strani okna (če privzete nastavitve niso primerne). S klikom na gumb »Shrani« shranimo profil.

Sprememba obstoječega profila

Izberemo profil iz seznama na levi strani okna, spremenimo nastavitve v zavihkih na desni strani okna in potrdimo spremembe s klikom na »Shrani«.

Upravljanje s tipi dokumentov – Nastavitve -> Tipi dokumentov



V levem delu okna je seznam tipov dokumentov, v desnem delu pa so zavihki v katerih nastavljamo posamezne lastnosti za izbran tip dokumenta. V nadaljevanju je opisan vsak zavihek.

V levem delu okna je še podatek »Število skenov«, ki pove koliko listov smo že poskenirali za ta tip dokumenta. Listi se pri skeniranju številčijo avtomatsko. Zaenkrat program omogoča skeniranje 9999 listov za en tip dokumenta.

Podatki – v tem zavihku nastavimo imena in tipe atributov za ta tip dokumenta. Program pozna tri tipe atributov: Besedilo (poljuben tekst), Vrednost (številčna vrednost z predznakom in decimalno vejico), Datum (datum v formatu dd.mm.llll)

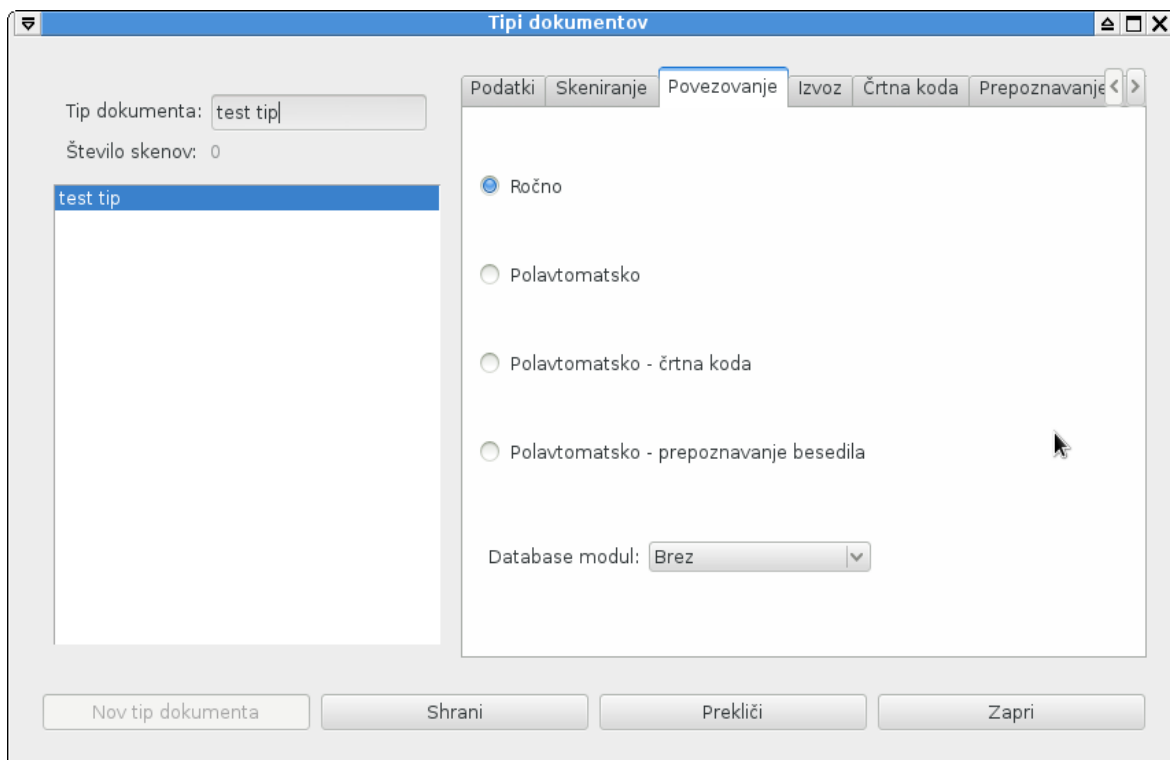
Primer: skeniramo prejete račune, zanima nas naziv firme ki je izdala račun, datum računa in vrednost računa. Naredili bi sledeče attribute:

Naziv firme Besedilo

Vrednost računa Vrednost

Datum računa Datum

Skeniranje – v tem zavihku izberemo skener in profil skeniranja za ta tip dokumenta.



Povezovanje – v tem zavihku izberemo način povezovanja posameznih listov v dokument in morebitno prepoznavanje črtne kode ali besedila ter iskanje podatkov za attribute dokumenta v podatkovni bazi.

Ročno – je primerno če skeniramo več dokumentov hkrati in bomo liste povezovali v dokumente šele po koncu skeniranja vseh dokumentov.

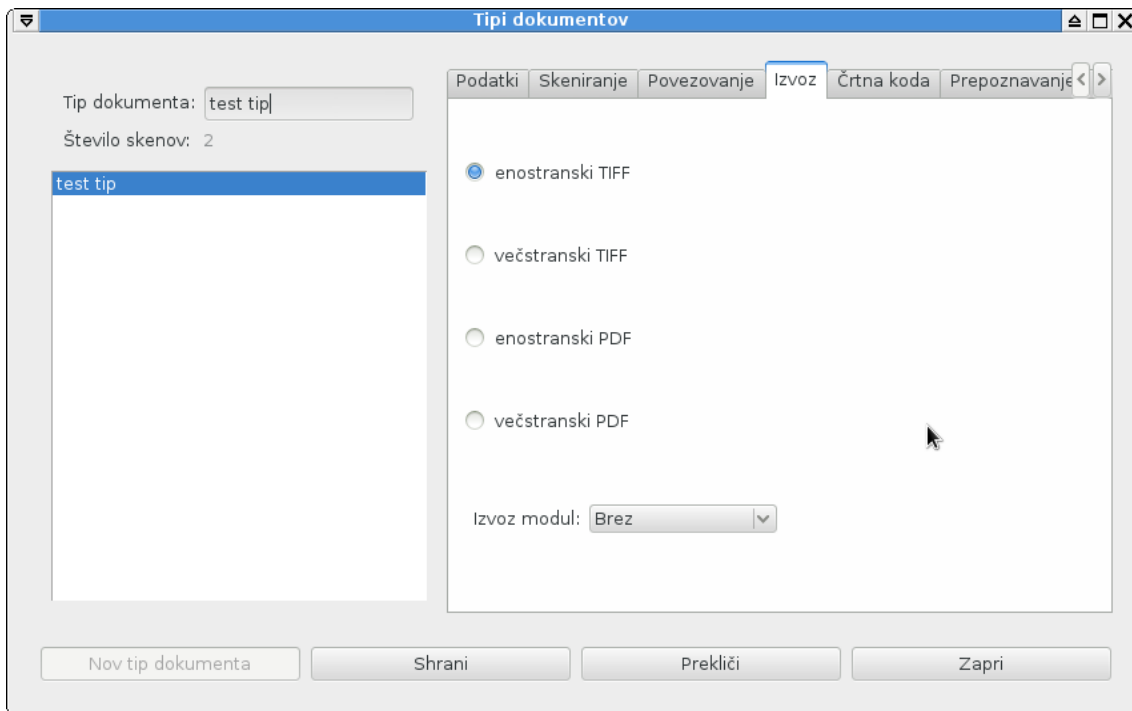
Polavtomatsko – je primerno če skeniramo en dokument, čigar liste bomo takoj povezali. Listi se avtomatsko označijo, tako da samo vpišemo attribute dokumenta in kliknemo na »Spni liste«

Polavtomatsko – črtna koda – je primerno če skeniramo en dokument, ki ima na prvem listu črtno kodo. Črtna koda se avtomatsko prečita v prvi atribut dokumenta, listi so označeni za spenjanje. Prepoznajo se samo tipi črtne kode, ki so omogočeni v zavihku »Črtna koda«. Če ni omogočen doben tip črtne kode le ta ne bo prepoznana.

Polavtomatsko – prepoznavanje besedila – je primerno, če skeniramo en dokument. Glede na nastavitve v zavihku »Prepoznavanje besedila« se v prvi atribut zapiše prepoznano besedilo iz prvega lista dokumenta.

Database modul – omogoča avtomatsko polnjenje ostalih atributov dokumenta iz podatkovne baze glede na podatek v prvem atributu. Primer funkcije, ki to omogoča je DemoDatabaseModul.py. Za vaš tip dokumenta lahko spremenite ta program ali pa napišete svojega. Pogoji, da ga program ponudi na izbiro, je da se program nahaja v mapi PluginsDatabase.

Deluje tako, da se v trenutku, ko polje za vnos prvega atributa izgubi fokus, pokliče izbrani program ki vrne listo s podatki za ostale attribute, ki se vpišejo v za to namenjena polja.



Izvoz – v tem zavihku določamo izhodno obliko dokumenta ter obliko izvozne datoteke. Izvozna datoteka je datoteka, ki vsebuje vse dokumente (skenirane liste, attribute dokumentov, morebitno prepoznano besedilo dokumentov) v obilki primerni za uvoz v kakšen drug program – najpogosteje elektronsko arhivo.

Enostranski TIFF – vsak list dokumenta se izvozi kot en TIFF dokument.

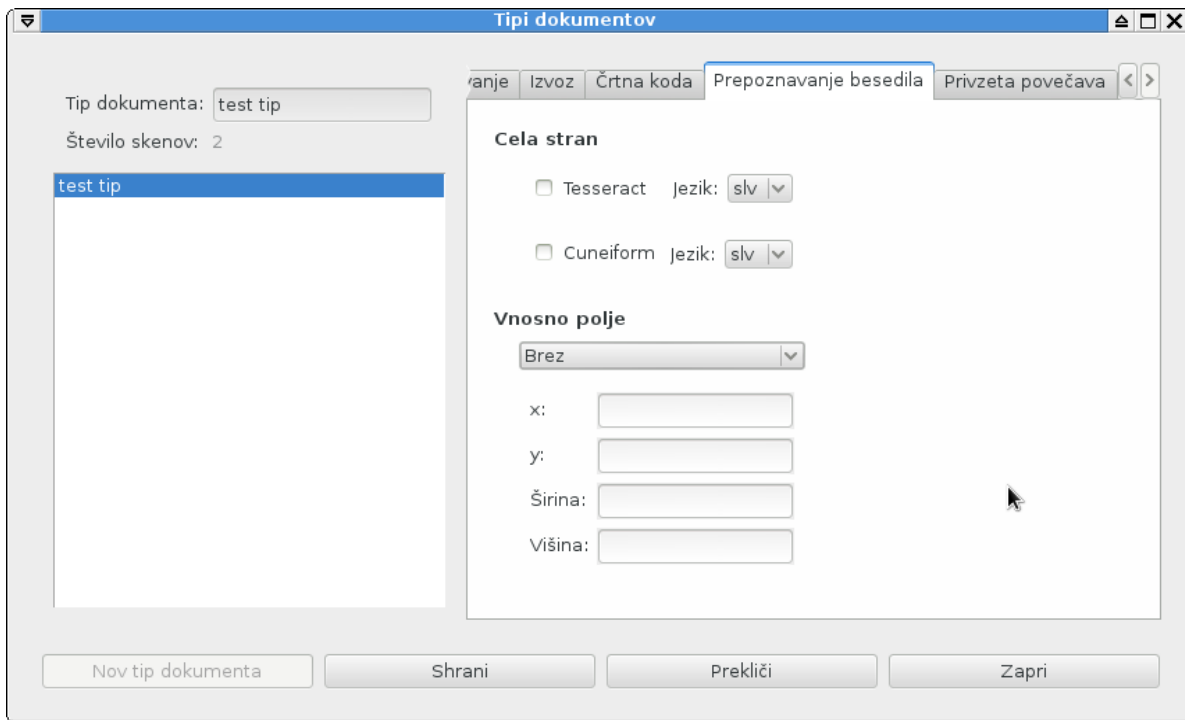
Večstranski TIFF – vsi listi dokumenta se izvozijo kot en večstranski TIFF dokument.

Enostranski PDF – vsak list dokumenta se izvozi kot en PDF dokument.

Večstranski PDF – vsi listi dokumenta se izvozijo kot en PDF dokument.

Izvoz modul – določa obliko izvozne datoteke. Če ni izbran se izvozna datoteka ne kreira. Ob izvozu dobimo samo TIFF ali PDF datoteke s skeni dokumentov ter morebitne datoteke s prepoznanim besedilo, podatki za attribute dokumentov pa so izgubljeni!

Primer programa za izvoz je DemoExportModul.py in naredi XML datoteko. Tega lahko po vaših željah spremenite ali napišete svojega. Da ga program prepozna mora biti v mapi PluginsExport.



Prepoznavanje besedila – v tem zavihku določamo ali se bo pri izvozu izvedla prepoznavna besedila za vse liste dokumentov in s katerim programom, ter s katerim programo se bo izvajala prepoznavna besedila v prvi atribut pri polavtomatskem skeniranju s prepoznavo besedila.

Za prepoznavo besedila se uporabljata odprtokodna programa Tesseract in Cuneiform. Kljukica pri imenu programa pomeni, da se bo s tem programom za vsak list izvedla prepoznavna besedila in rezultat zapisal v txt datoteko. Če je kljukica pri obeh programih dobimo za vsak list dve txt datoteki.

! optična prepoznavna besedila je lahko časovno zelo zahtevna in je odvisna od lastnosti računalnika in vsebine skeniranih strani.

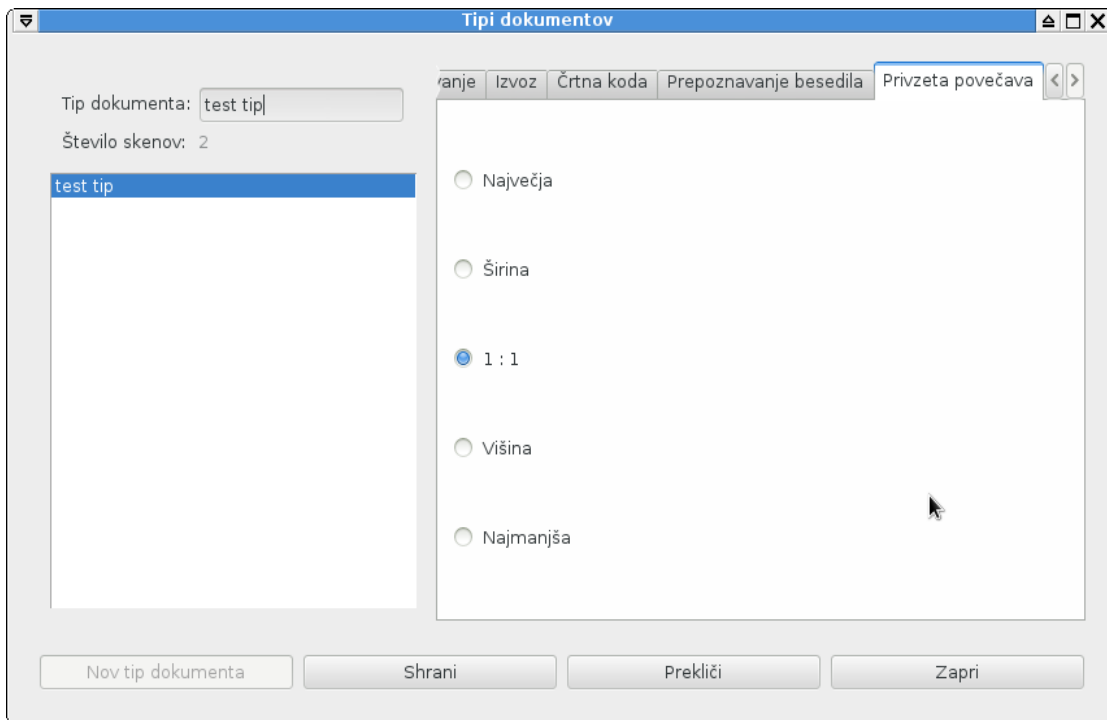
V spodnjem delu tega zavihka lahko določimo s katerim programom se bo izvajala prepoznavna besedila za prvi atribut dokumenta ter velikost področja za prepoznavo. Vse vrednosti so v procentih velikosti skenirane lista na katerem se dela prepoznavna.

Pogosto se dogaja, da podatek, ki nas zanima, ni vedno na istem mestu. V tem primeru seveda ne bo pravilno prepozna ali pa sploh ne bo.

Pomagamo si lahko tako, da z miško kliknemo na levi zgornji kot besedila, ki ga želimo prepoznati (izberemo zgornji levi kot področja za prepoznavo, velikost pa ostane tista ki smo jo določili).

V primeru, da tudi velikost besedila ni pravilna, z miško kliknemo a levi zgornji kot besedila in kazalec premaknemo v spodnji desni kot besedila in šele sedaj spustimo tipko na miški (z miško označimo velikost področja za prepoznavo).

! Možnost prepoznavne besedila na danem področju s pomočjo miške deluje tudi, če ni nastavljeno na zavihku »Povezava«, le da ne deluje avtomatsko ampak imamo zraven opcij naštetih zgoraj še možnost klika na dokumet z desno tipko na miški, ki prepozna besedilo iz naprej nastavljenega področja.



Privzeta povečava – v tem zavihku izberemo povečavo v kateri bo prikazan skeniran list.

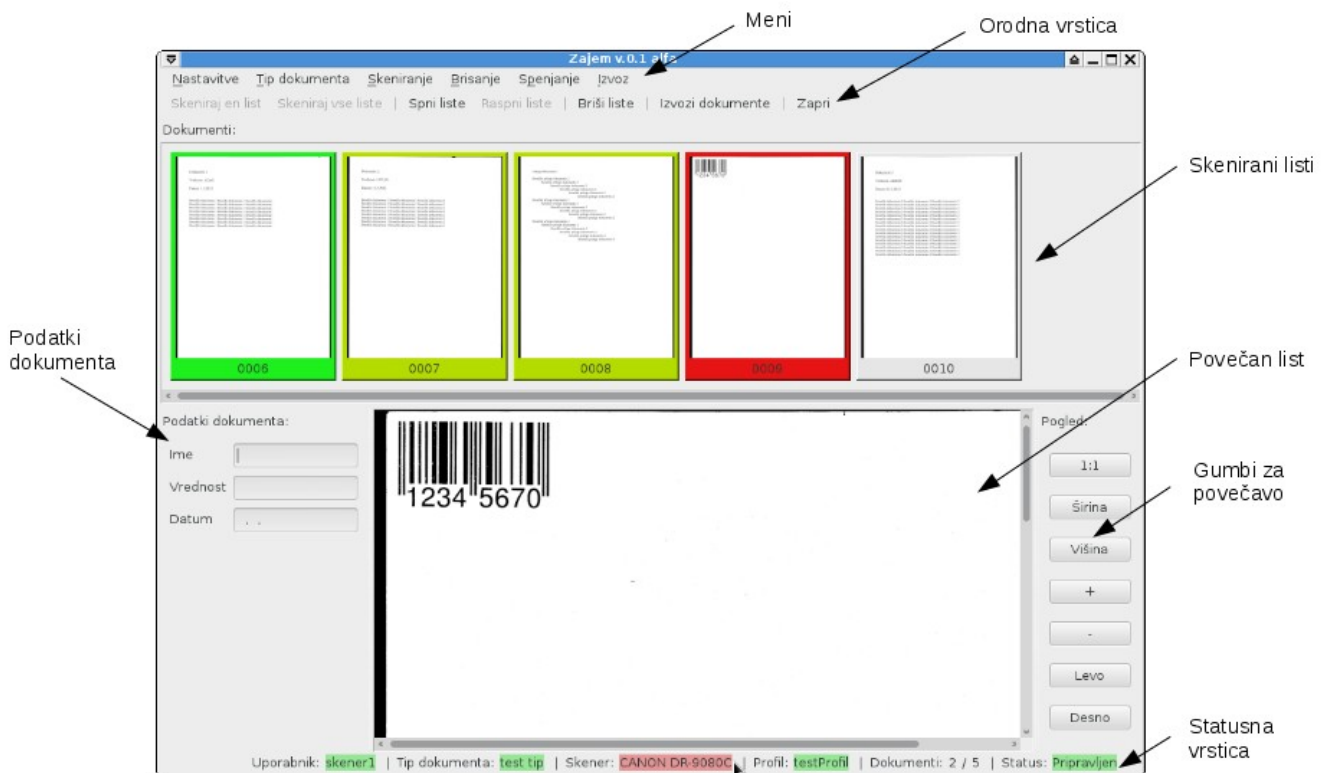
Največja (gumb »+«) - največja povečava, ena skenirana točka je enaka eni prikazani točki na ekranu

Širina (gumb »Širina«) - program velikost slike skeniranega lista prilagodi tako, da je viden cel dokument po širini.

Višina (gumb »Višina«) - program velikost slike skeniranega lista prilagodi tako, da je viden cel dokument po višini.

Najmanjša (gumb »-«) - najmanjša povečava, velikost slike skeniranega lista je 20% izvirne velikosti.

1:1 (gumb »1:1«) - je privzeta privzeta povečava, program skuša glede na velikost in resolucijo zaslona ter resolucijo skeniranja sliko skeniranega lista povečati/pomanjšati tako, da je na zaslonu enake velikosti kot je list v fizični obliki.



Grafični vmesnik – podroben opis

Meni – meni programa je odvisen od pravic prijavljenega uporabnika. Na sliki je meni, kot ga vidi uporabnik, ki ima pravice za skeniranje. Posamezne črke v postavkah menija so podčrtane, kar pomeni da lahko to postavko izberemo tudi s kombinacijo tipk »Alt« + črka.

Orodna vrstica – izgled je odvisen od pravic prijavljenega uporabnika in stanja programa. Pomemben je zato, da so akcije dosegljive z enim klikom na gumb. Akcije, ki jih trenutno ni možno izvesti so označene sveto sivo.

Skenirani listi – v tem oknu so prikazni vsi trenutno poskenirani listi za izbran tip dokumenta. Listi označeni z zeleno-modrimi odtenki so že povezani v dokument, listi označeni z rdečo so izbrani za naslednjo akcijo, vsi ostali listi so označeni s sivo barvo.

Za izbiro listov se uporablja miška. Kot je v navadi v drugih programih leva tipka izbere list, »Ctrl« + leva tipka doda/odstrani list k izbranim, »Shift« + leva tipka doda/odstrani liste od prvega izbranega do tistega na katerega smo kliknili. Da ni potrebna uporaba tipkovnice je dodan še klik z desno tipko na miški, ki deluje enako kot kombinacija »Shift« + leva tipka.

Podatki dokumenta - vnosna polja za podatke posameznega dokumenta. Odvisna so od izbranega tipa dokumenta. Vrednosti, ki jih lahko vnesemo v posamezno vnosno polje so odvisne od tipa vnosnega polja. V polje tipa Vrednost lahko vnašamo samo številke s predznakom in decimalno vejico. V polje tipa Datum pa samo datume v obliki dd.mm.llll. Če vnos ni pravilne oblike se vnosno polje obarva z odtenkom rdeče, ko želimo liste povezati v dokument.

Povečan list – v tem oknu vidimo izbran list povečan glede na privzeto povečavo za izbran tip dokumenta. Povečavo lahko spreminjamo z gumbi za povečavo. Če je nastavljena prepoznavna besedila jo lahko izvedemo z miško (desni klik - privzeto področje, levi klik - izbran levi zgornji kot, velikost privzeta, izbira področja z levo tipko – izbrano področje)

Gumbi za povečavo – s temi gumbi lahko povečamo/pomanjšamo sliko dokumenta ali sliko zavrtimo za 90 stopinj.

»1:1« - je privzeta povečava – list na ekranu bi naj bil enake velikosti kot list v fizični obliki.

»Širina« – list je prikazan tako da je viden cel po širini, po višini se lahko premikamo z drsnikom ali kolesčkom na miški.

»Višina« – list je prikazan tako, da je viden cel po višini.

»+« - list je prikazan v največji povečavi, ena skenirana točka ustreza eni točki na ekranu

»-« - list je prikazan v najmanjši povečavi, slika je 20% velikosti skenirane slike

»Levo« - zavrti sliko za 90 stopinj v smeri nasprotni urinemu kazalcu.

»Desno« - zavrti sliko za 90 stopinj v smeri urinega kazalca.

! Listi, ki smo jih zavrteli, ostanejo v takšnem položaju tudi pri izvozu dokumentov.

Statusna vrstica – prikazuje stanje programa. Od leve proti desni vidimo – prijavljenega uporabnika, tip dokumenta, skener, profil skeniranja, trenutno število dokumentov/listov in status programa. Če je prišlo do napake je posamezen status obarvan v odtenku rdeče (na sliki vidimo da program ni zaznal skenerja).

! **Miškin kazalec** – ko se izvajajo dolgotrajna opravila (skeniranje, izvoz) se miškin kazalec spremeni in s tem pokaže, da naj do trenutka, ko se spremeni nazaj, ne izvajamo nobenih akcij.

Primer postopka skeniranja

Prijavimo se kot uporabnik, ki ima pravico skeniranja.

V opravilni vrstici izberemo tip dokumenta.

Odvisno od tipa skenerja kliknemo na gumb »Skeniraj en list« ali »Skeniraj vse liste« in počakamo, da se akcija izvede.

Označimo liste, ki sestavljajo pri dokument in vpišemo podatke.

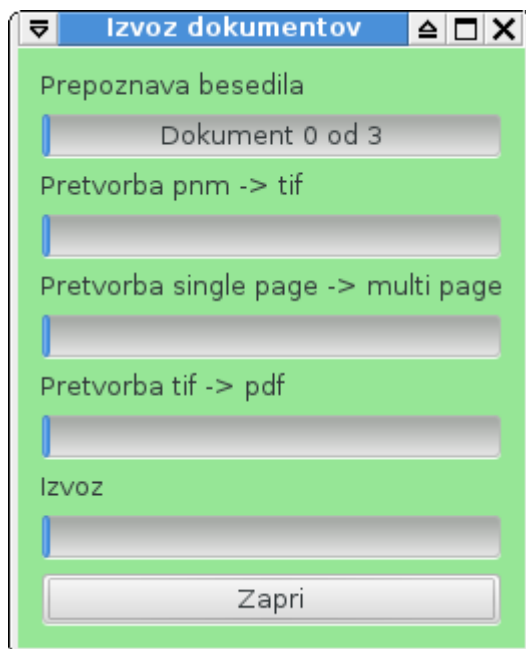
Spnemo liste v dokument s klikom na »Spni liste«

To ponovimo za vse dokumente.

Če smo se pri spenjanju zmotili lahko dokument raspnemo nazaj v liste s klikom na »Raspni liste«.

Če je bil poskeniran kakšen list odveč ga lahko pobrišemo s klikom na »Pobriši liste«.

Ko so dokumenti pripravljene jih izvozimo s klikom na »Izvozi dokumente« in počakamo da se akcija izvede. Izvoz lahko glede na število dokumentov in možnost prepoznavanja besedila traja dalj časa. Potek izvoza je viden v oknu na sliki:



Glede na nastavitve se določene akcije ne izvedejo. Če je na primer za tip dokumenta nastavljen izvoz v TIFF obliko bo vrstica Pretvorba tif -> pdf ostala prazna.

Ko se miškin kazalec vrne v prvotno obliko kliknemo gumb »Zapri« in s tem končamo izvoz.

! Listi ki niso speti v dokument se ne izvozijo!

! Slike izvoženih dokumentov in morebitne txt datoteke s prepoznanim besedilom se nahajajo v mapi Izvoz/Dokumenti, datoteka za izvoz (njeno ime je odvisno od programa za izvoz) pa v mapi Izvoz.