**Patento teksto XML failo sudarymas**

Failo pavyzdys:

<?xml version="1.0" encoding="UTF-8" standalone="yes"?>

**<simple-patent-document lang="lt" status="n" date-publ="20130903" kind="B" doc-number="5958" country="LT">**

**<description lang="lt">**

<doc-page file="LT5958B.pdf" wi="210" he="297" type="pdf" ppf="2" ppl="6"/>

**</description>**

**<description lang="lt">**

**<p num="0001">**Išradimas priskiriamas prie gyvūnų pašarų, būtent prie ėdesių priedų, skirtų naminiams gyvūnams. **</p>**

**<p num="0002">**Gyvūnų mitybos srityje ėdesiai ir ėdesių priedai yra labai svarbūs, nes į juos pridėta svarbių gyvūno sveikatą gerinančių bioaktyvių junginių. <img file="imgb0001.tif" wi="201" he="157" img-content="chem" img-format="tif"/> **</p>**

**</description>**

**<claims lang="lt">**

<doc-page file="LT5958B.pdf" wi="210" he="297" type="pdf" ppf="7" ppl="7"/>

**</claims>**

**<claims lang="lt">**

**<claim num="0001">**

**<claim-text>**Naminių gyvūnėlių ėdesių priedas, apimantis rozmarino ir cikorijų šaknų ekstraktus, b esi s k i r i a n t i s tuo, kad priedas apima mikronizuotą silicio-aliuminio ceolitą, rozmarino ekstraktą bei cikorijų šaknų ekstraktą, esant sekančiam ingredientų santykiui: mikronizuotas silicio-aliuminio ceolitas 20-60 masės %, rozmarino ekstraktas 20-40 masės %, cikorijų šaknų ekstraktas 20-40 masės % **</claim-text>**

**</claim>**

**<claim num="0002">**

**<claim-text>**Priedas pagal apibrėžties 1 punktą, b e s i s k i r i a n t i s tuo, kad mikronizuoto silicio-aliuminio ceolito dalelių dydis **&lt**;= 3 µm. **</claim-text>**

**</claim>**

**</claims>**

**<drawings>**

**<figure num ="1">**

<img file="imgf0001.tif" wi="201" he="157" img-content="drawing" img-format="tif"/>

**</figure>**

**<figure num ="2">**

<img file="imgf0002.tif" wi="171" he="153" img-content="drawing" img-format="tif"/>

**</figure>**

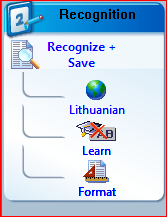
**</drawings>**

**</simple-patent-document>**

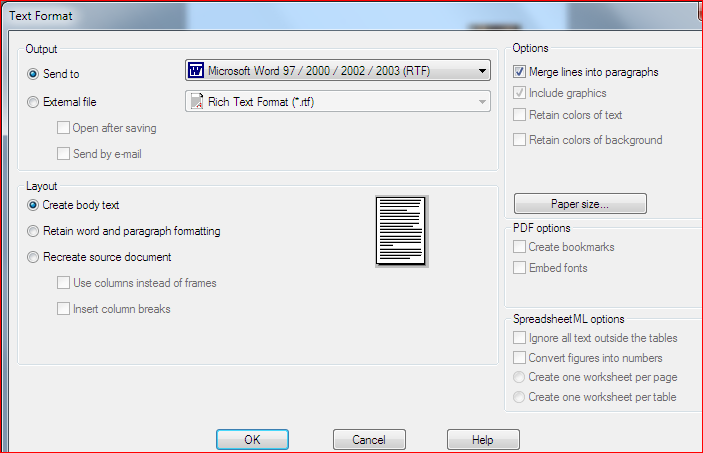
Patento tekstas iš skenuoto dokumento ištraukiamas naudojantis teksto atpažinimo programomis. Čia aprašyta darbo tvarka su Readiris Pro 11 programine įranga.

**1.** Meniu *File -> Open* pasirenkame patento pdf failą;

**2.** *Recognition* lange nurodome kalbą (Lithuanian), pasirenkame *Format*:

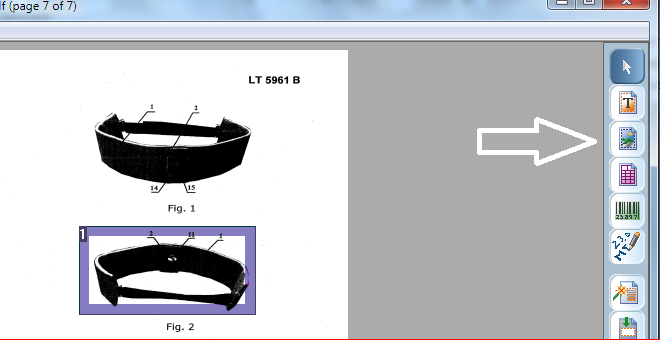


ir Text Format lange nurodome *RTF, Create body text*, spaudžiame *OK* ir atpažinimo procesui pasibaigus tekstinį dokumentą išsaugome.



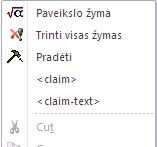
**3.** Visus patento paveikslus – brėžinius, lenteles, chemines ir matematines formules išsaugome tif formatu. Paveikslus, esančius tekste nuosekliai pavadiname imgb0001.tif, imgb0002.tif ir t.t., o brėžinių skyriaus paveikslus – imgf0001.tif, imgf0002.tif ir t.t.

Programos lange nurodome paveikslą pasirinkę *Draw graphics window*, jį pažymimeir išsaugome *File -> Save Graphics*.

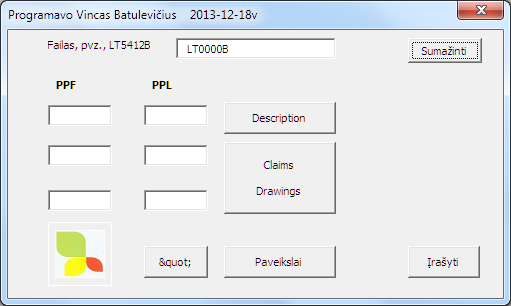


**4.** XML dokumentas sukuriamas naudojantis Fulltext20131023.docm failu su makrokomandomis.

Paspaudus dešinį pelės klavišą pasirenkame *Pradėti*:



Atsidaro langas:



PPF langeliuose nurodomi patento pdf failo aprašymo, apibrėžties ir brėžinių pradžios puslapio numeris, o PPL langeliuose – pabaigos puslapio numeris. Jei patente brėžinių skyriaus nėra, į atitinkamus langelius nieko nerašoma.

**5.** Į atidarytą Fulltext20131023.docm failą nukopijuojamas aprašymo (description) tekstas, sutvarkomas taip, kad būtų sudarytos tvarkingos pastraipos, o paveikslai (formulės, lentelės ir t.t.) ištrinami. XML failuose neleidžiami simboliai <, > ir kt., todėl juos reikia pakeisti atitinkamais kodais, žr. <http://www.w3schools.com/tags/ref_entities.asp>. Dažniausiai pasitaikančius simbolius galima pakeisti paspaudus mygtuką ***&quot***;

Jei aprašymo (description) tekste yra lentelių, formulių, paveikslų ir t.t., atitinkamų pastraipų pabaigoje įterpiamos žymos naudojantis kontekstinio meniu punktu *Paveikslo žyma*. Žymos formatas yra **\_\_\_imgb0001\_\_\_.** Įterpus naują žymą, jos numeris padidėja vienetu. Jei buvo neteisingai nurodytos žymos, jas reikia ištrinti pasirinkus *Trinti visas žymas* ir įterpti iš naujo.

Spaudžiame *Description* mygtuką ir taip suformuojami aprašymo XML elementai.

**6.** Ištrinamas aprašymo tekstas ir į jo vietą nukopijuojamas sutvarkytas apibrėžties tekstas. Jei tekste yra specialių simbolių, spaudžiamas mygtukas ***&quot***; Jei apibrėžties tekste yra lentelių, formulių, paveikslų ir t.t., atitinkamų pastraipų pabaigoje įterpiamos žymos naudojantis kontekstinio meniu punktu *Paveikslo žyma*. Reikia žiūrėti, kad žymų numeracija tęstų aprašymo paveikslų numeraciją.

Jeigu apibrėžties punktuose yra sąrašų, XML žymas <claim num="X"><claim> ir <claim-text></claim-text> reikia pridėti naudojantis kontekstinio meniu punktais <claim> ir <claim-text> pažymėjus reikiamas pastraipas. Taip sutvarkyti apibrėžties punktai turi atrodyti taip:

<claim num="0001">

<claim-text>1. Šviesos impulsų suminio dažnio generavimo būdas, kai į kvadratinio jautrio netiesinę medžiagą nukreipia šviesos impulsus, kurioje, tenkinant fazinio sinchronizmo sąlygas, vyksta suminio dažnio generavimas, b e s i s k i r i a n t i s tuo, kad šviesos impulsus į kvadratinio jautrio netiesinės medžiagos įėjimo galą nukreipia poromis nuosekliai vienas po kito taip, kad </claim-text>

<claim-text>- kiekvienos poros pirmas šviesos impulsas pirmą kartą praeina kvadratinio jautrio netiesinę medžiagą joje nesukeliant netiesinės sąveikos, </claim-text>

<claim-text>- praėjusį per minėtą netiesinę medžiagą pirmą impulsą grąžina atgal ir vėl nukreipia į minėtos kvadratinio jautrio netiesinės medžiagos tą patį įėjimo galą, </claim-text>

<claim-text>- antrą kartą sklisdamas grąžintas minėtas pirmas šviesos impulsas persikloja su pirmą kartą sklindančiu per kvadratinio jautrio netesinę medžiagą tos pačios poros antruoju šviesos impulsu, užtikrinant tarp šių impulsų fazinio sinchronizmo sąlygas, </claim-text>

<claim-text>- išpildžius minėtoje kvadratinio jautrio netiesinėje medžiagoje tarp minėtų impulsų fazinio sinchronizmo sąlygas vyksta efektyvi netiesinė sąveika ir vyksta minėtų pirmą ir antrą kartą kvadratinio jautrio neteisinėje medžiagoje sklindančių kiekvienos poros impulsų apsijungimas į suminio dažnio šviesos impulsus. </claim-text>

</claim>

<claim num="0002">

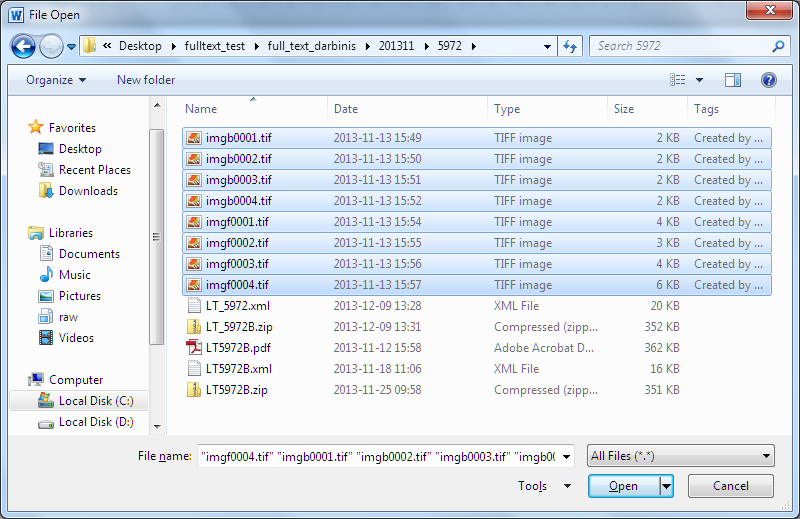
<claim-text>2 .Šviesos impulsų suminio dažnio generavimo būdas pagal 1 punktą, b e s i s k i r i a n t i s tuo, kad laiko intervalas tarp porą sudarančių šviesos impulsų yra lygus trukmei T, per kurią poros pirmasis šviesos impulsas pradedant nuo kvadratinio jautrio netiesinės medžiagos įėjimo galo prasklinda per minėtą kvadratinio jautrio netiesinę medžiagą, grąžinamas atgal ir nukreipiant vėl patenka į minėtos kvadratinio jautrio netiesinės medžiagos tą patį įėjimo galą. </claim-text>

</claim>

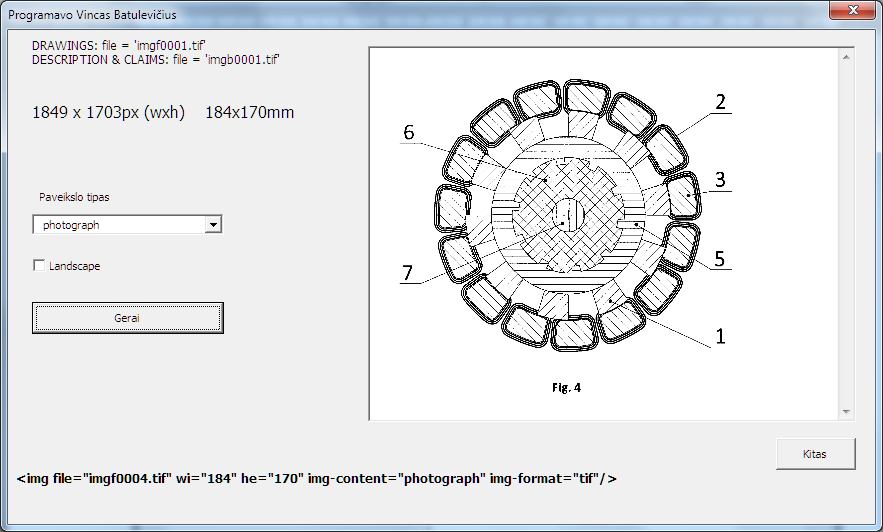
Jei sąrašų nėra, jokių XML žymų pridėti nereikia.

Mygtuku *Claims*/*Drawings* bus suformuoti reikiami elementai.

**7.** Jei patente yra paveikslų, spaudžiame mygtuką *Paveikslai* ir pasirenkame visus patento paveikslus:



Atsivėrusiame kitame lange nurodome paveikslo tipą (math, chem, flowchart, drawing ir t.t.), orientaciją (jei *landscape*), spaudžiame *Gerai* ir po to spaudžiame *Kitas* kol sutvarkome visus paveikslus.



**8.** Kai visos patento dalys apdorotos, spaudžiame *Įrašyti*. Sukuriamas patento XML failas ir išsaugomas.

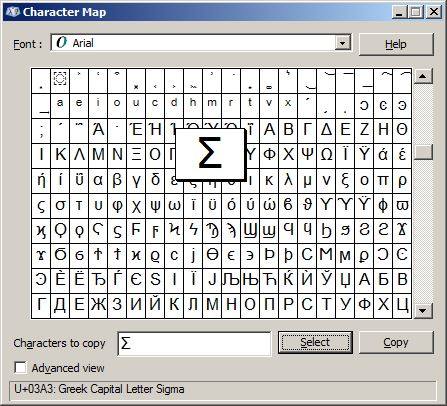
Paveikslų iš aprašymo <img> elementai (imgb000X.tif) automatiškai įrašomi į reikiamas pastraipas (<p> elementus) kaip buvo nurodyta žymomis, pavyzdžiui:

<p num="0002">Gyvūnų mitybos srityje ėdesiai ir ėdesių priedai yra labai svarbūs, nes į juos pridėta svarbių gyvūno sveikatą gerinančių bioaktyvių junginių. **<img file="imgb0001.tif" wi="201" he="157" img-content="math" img-format="tif"/>**  </p>

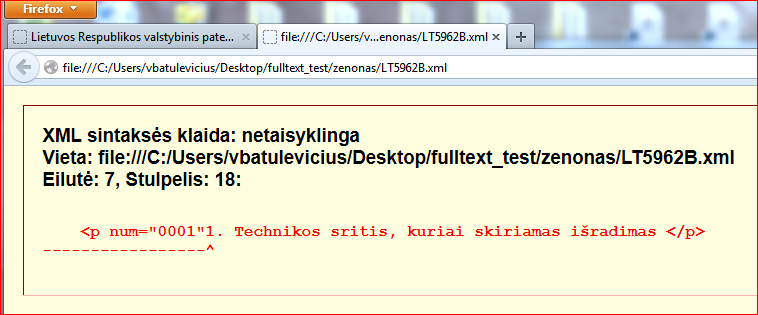
Būtinai reikia patikrinti ar faile nėra klaidų, neleidžiamų simbolių ir ar <img> elementai yra reikiamose vietose.

<figure> elemento atributas *num* turi atitikti patento originalo numeraciją.

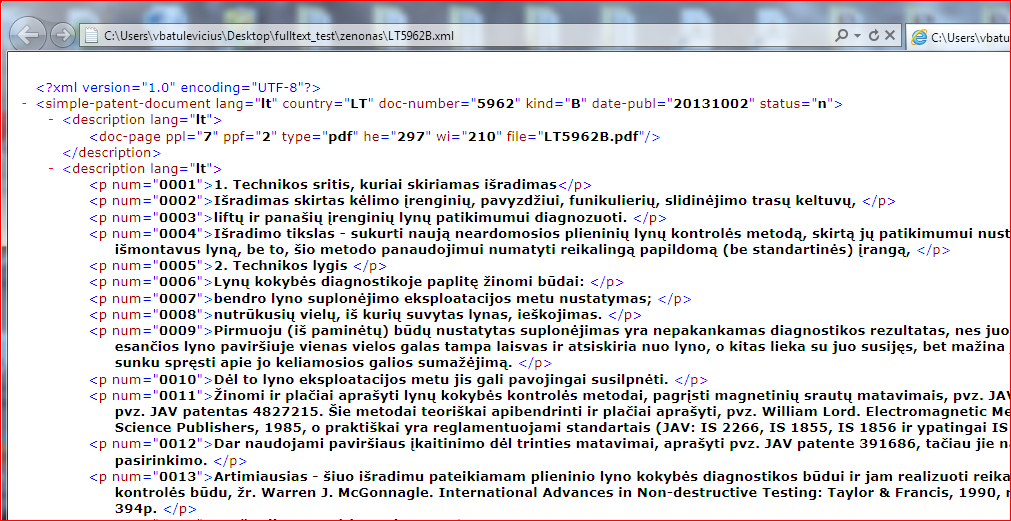
Tvarkant failą patogu naudotis programa Notepad++. Jeigu į tekstą reikia įterpti nelotyniškas raides ar kitus simbolius, pavyzdžiui, Celsijaus laipsnio ženklą (°), Notepad++ teksto redaktoriuje per meniu punktą *Run‑>charmap* reikia atidaryti langą *Character Map*, kuriame bus galima pasirinkti reikiamą simbolį:



Ar faile nėra XML sintaksės klaidų galima patikrinti atidarant failą naršyklėmis *Internet Explorer* arba *Firefox*. Jei faile yra klaidų, tai failo turinys bus nerodomas arba nurodoma vieta, kur yra klaidų.



Failas be klaidų turi atrodyti taip *Internet Explorer* naršyklėje:



*Firefox* naršyklėje:

