

Webcrawling metodo taikymas Aplinkos apsaugos departamente automatizuotai nelegalių skelbimų paieškai

Parengė:

Monika Barauskaitė

Benas Švedas

Turinys

| | |
|---|---|
| Skaitmeninės Aplinkos apsaugos departamento problemos..... | 3 |
| Kas yra webcrawler? | 3 |
| Aplinkos apsaugos departamento nelegalių skelbimų paieškos ir patikros vykdymo procesai | 3 |
| Dabar egzistuojantis procesas | 3 |
| Siekiamas procesas ateityje | 4 |
| Webcrawler sprendžiama Aplinkos apsaugos departamento problema..... | 4 |
| Draudžiamų prekių ir paslaugų specifika | 5 |
| Projekto prioritetinės sritys | 5 |
| Lietuvos institucijų praktikos taikant webcrawling technologiją | 5 |
| Projekto tikslas..... | 7 |
| Planuojamas projekto įgyvendinimas | 7 |
| Projekto įvykdymo kriterijai (acceptance criteria)..... | 8 |
| Projekto rizikos, priklausomybės | 8 |
| Laukiamas projekto rezultatas ateityje..... | 8 |

Skaitmeninės Aplinkos apsaugos departamento problemos

2022-aisiais tebeegzistuoja atliekų vežimas į mišką – vien per pirmus aštuonis 2022 m. mėnesius buvo gauta **virš 850 pranešimų** dėl atliekų netvarkymo. Taip pat internete neatsakingai pardavinėjami invazinių ar saugomų, retų rūšių augalai ir gyvūnai ar jų produktai. Dėl interneto platumų ir nuolatinio informacijos pertekliaus pastebėti šias veiklas internete nėra paprasta. Efektyviai prižiūrėti veiklą internete **reikalingi inovatyvūs technologiniai įrankiai**.

Šiuo metu Aplinkos apsaugos departamente (AAD) vyraujantis nelegalių skelbimų stebėjimo internete procesas yra gana primityvus (žr. *skyrių Procesai -> Esama situacija*). Dabartinė skelbimų **paieška vykdoma nereguliariai, dideliais laiko intervalais**, pavyzdžiui, kartą per mėnesį arba tik atsiradus poreikiui. Dar svarbus aspektas – visa per patikrinimus gauta medžiaga, protokolai yra **saugomi nevienodai**, todėl, norint peržiūrėti atliktų patikrų istoriją, reikia kreiptis į skirtingus departamento skyrius atskirai. Taip pat, **nekaupiama ir statistika** apie tai, kiek iš tiesų buvo užfiksuota nelegalios veikos atvejų, kuomet buvo vykdoma planinė ar neplaninė patikra, todėl itin **sunku suvokti nelegalių skelbimų internete mastą**. Vien šiandien (2022-10-28), pagrindiniuose skelbimų portaluose (skelbiu.lt ir alio.lt) yra virš 1 milijono skelbimų, o AAD dominančiose kategorijose – daugiau kaip 126 tūkstančiai.

Be aktualios informacijos ir visos situacijos matymo, **sunku** vykdyti ne tik operatyvinę veiklą, bet ir **priimti kokybiškus, duomenimis grįstus valdymo sprendimus, vertinti galimą bendrą internete sutinkamų nelegalių veikų poveikį** aplinkai, į valstybės biudžetą **nesumokamų mokesčių sumą** (leidimų mokesčiai, gyventojų pajamų mokestis, pelno mokestis), sudėtinga užtikrinti *teršėjas moka* principo įgyvendinimą.

Kas yra webcrawler?

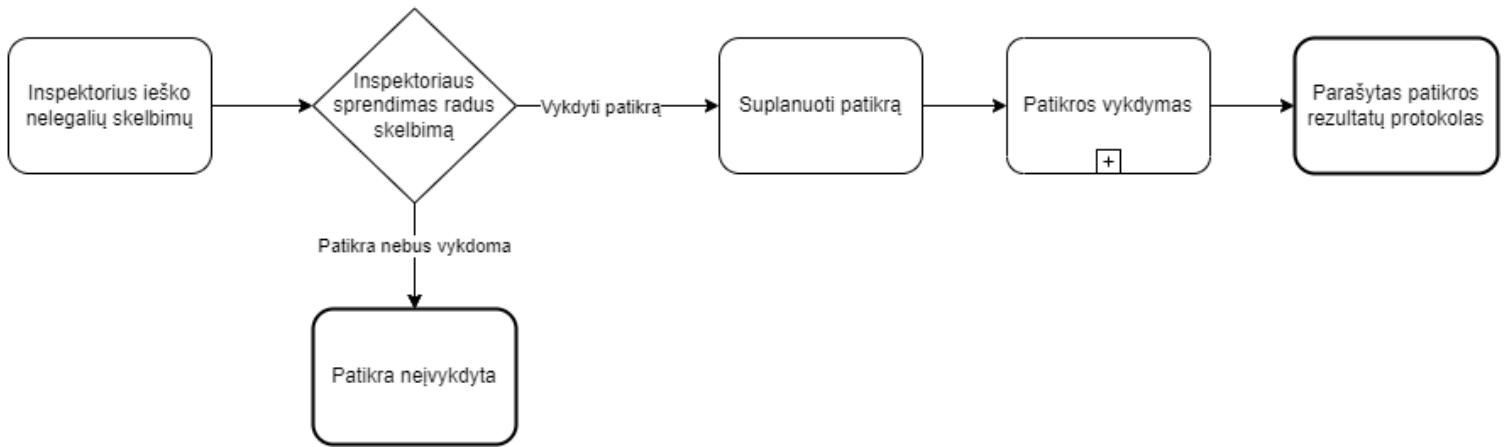
Webcrawler – skaitmeninis įrankis skirtas skenuoti nurodytus interneto tinklapius ieškant iš anksto nurodyto tikslo, kuris gali būti tekstinis arba vaizdinis. Radęs užduotą surasti tikslą webcrawler įrankis nurodo radimo vietą ir laiką. Plačiai sutinkamas webcrawling technologijos panaudojimas - kainų palyginimo portalai, kurie būtent šiuo principu nuskaito informaciją iš skirtingų el.parduotuvių.

Dažnai panašiam kontekste sutinkamas ir **webscraping** terminas. Tai yra tas pats webcrawler, tačiau, skirtumas tas, kad webscraper, radęs nurodytą tikslą ne tik išsaugo jo buvimo vietą ir laiką, bet išsaugo (parsiuočia) ir rastą informaciją (tekstą, nuotraukas).

Aplinkos apsaugos departamento nelegalių skelbimų paieškos ir patikros vykdymo procesai

Dabar egzistuojantis procesas

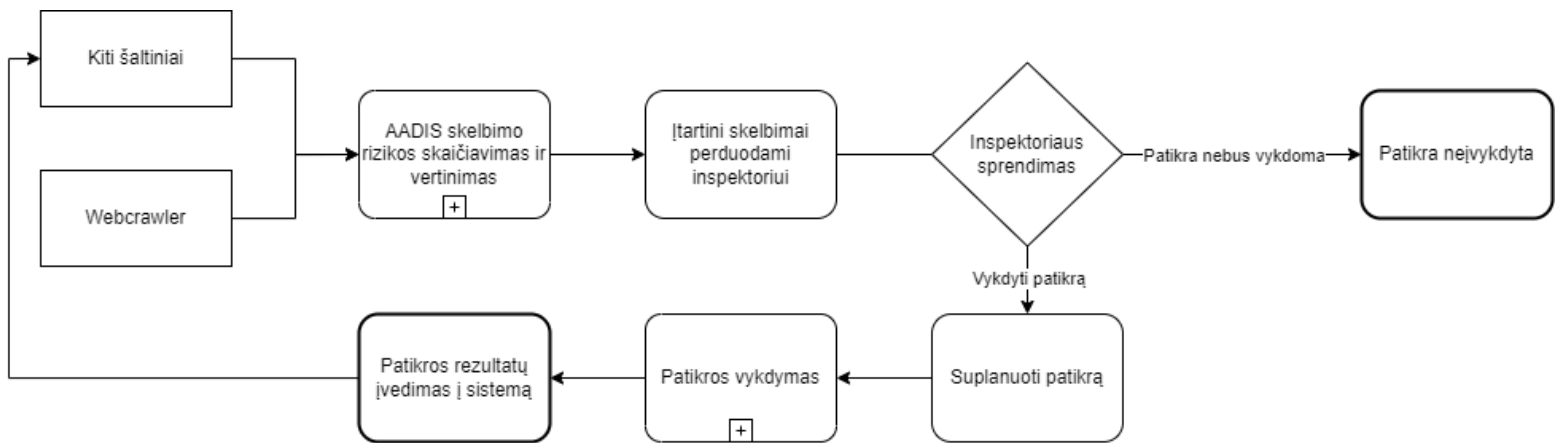
Šiuo metu AAD inspektoriai dažniausiai patys ieško galimai nelegalių skelbimų internete. Aptikus tokį skelbimą, inspektoriai patys ir priima sprendimą, kaip reaguoti į jį – kokio pobūdžio patikrą vykdyti, ar jos išvis reikia. Įvykdžius patikrą, visais reikalingais dokumentais ir protokolais nebūna pasidalinama su kitais padaliniais – patikros rezultatai būna fiksuojami popieriniu formatu arba lokaliai inspektoriaus kompiuteryje, todėl nebūna matomi jokioje bendroje sistemoje. Tokiu būdu pasidalinimas informacija tarp įstaigų ar inspektorių tampa ganėtinai sudėtingu procesu.



1 pav. *Esamas nelegalių skelbimų paieškos ir patikros vykdymo procesas Aplinkos apsaugos departamente.* Schema kurta autorių, 2022 m. spalio mėn.

Siekiamas procesas ateityje

AAD kuriama bendra informacinė sistema optimizuotų dabartinį nelegalių skelbimų paieškos procesą. Šioje sistemoje, pasitelkus webcrawler įrankį, pagal raktažodžius būtų surasti visi galimai nelegalūs skelbimai, esantys Lietuvos internetinėje erdvėje ir, panaudojus informaciją ir iš kitų šaltinių, įvertinama to skelbimo rizika. Inspektorius, gavęs ne tik skelbimą, bet ir to skelbimo riziką, kad jis yra nelegalus, galėtų žymiai efektyviau priimti sprendimus, vykdyti patikras ir, svarbiausia, kad tos patikros rezultatai būtų matomi bendroje sistemoje vėlesniuose skelbimų rizikos vertinimo skaičiavimuose.



2 pav. *Siekiamas nelegalių skelbimų paieškos ir patikros vykdymo procesas Aplinkos apsaugos departamente.* Schema kurta autorių, 2022 m. spalio mėn.

Webcrawler sprendžiama Aplinkos apsaugos departamento problema

Išsiaiškinus esamą situaciją, galima teigti, jog pagrindinė problema yra ta, kad šiuo metu **AAD atliekama galimai nelegalių skelbimų paieška vykdoma rankiniu būdu, todėl yra lėta, neišnaudojamos IS efektyvumo galimybės, neefektyviai naudojami žmogiškieji ištekliai.**

Webcrawler įrankio pagalba būtų automatizuojama šiuo metu departamente vykdoma galimai nelegalių skelbimų paieška. O integravus šį įrankį į visą AAD informacinę sistemą (AADIS), inspektoriams būtų pateikiama išsamesnė informacija apie skelbimo keliamą riziką. Tuo pačiu, inspektoriams daugiau laiko galėtų skirti pačių patikrinimų planavimui ir vykdymui, jiems nebereikėtų pildyti popierinių dokumentų – viskas automatiškai būtų užfiksuojama sistemoje.

Draudžiamų prekių ir paslaugų specifi­ka

Gamtos išteklių naudojimas ir prekyba Lietuvoje **nėra absoliučiai draudžiama**, tačiau, teisės aktuose yra numatyta tvarka ir neretu atveju tam reikia gauti **leidimą**. Leidimai gali būti išduodami ir draudžiamoms prekėms, pavyzdžiui, žvejybos tinklams, tačiau, tokie atvejai labai reti. Tad, dažnu atveju, aptikus parduodamos invazinės arba saugomos rūšies gyvūno ar augalo, draudžiamo medžioklės ar žūklės įrankio skelbimą, tampa **akivaizdu, jog aptikome neteisėtą veiką**.

Verta pastebėti ir tai, jog Lietuvoje yra nedidelė prekybos laukiniais gyvūnais ir saugomais augalais rinka, tad dažnai skelbimai yra **unikalūs** ir tikėtina, kad greitai nepasikartosiantys, todėl nestebint **tendencijų** sunku aptikti šiuos skelbimus. Be to, kokybiškam automatizuotos skelbimų paieškos veikimui, kiekvienam stebimam gyvūnui ar augalui reikalingi **ne tik lietuviškas ir lotyniškas pavadinimai** (vien jų CITES konvencijoje yra 2,4 tūkst.), bet reikia žinoti ir kaip šie gyvūnai vadinami prekeivių **žargone ir tarmiškai**.

Situacija dar sudėtingesnė kontroliuojant tas sritis, kur leidimų turėtojų yra nemažai (pvz.: nuotekų ir atliekų išvežimo ir tvarkymo paslaugas teikiantys ūkio subjektai, automobilių remonto dirbtuvės). Vien Aplinkosaugos leidimų išdavimo sistemoje (ALIS) galima užsisakyti **17 skirtingų leidimų** ir, deja, ši sistema nevienija visų Aplinkos ministerijos prerogatyvoje esančių sričių leidimų. Radus vienos iš šių veikų skelbimą internete, **būtina įsitikinti skelbimo autoriaus (ne)buvimu atitinkamos srities leidimų duomenų bazėje**, nes kitaip galima nesunkiai apkaltinti legaliai veikiančias įmones ar nepastebėti niekuo neįtartiną skelbimą, kurio autorius visgi neturi leidimo vykdomai veiklai.

Projekto prioritetinės sritys

Taikant webcrawler įrankį ir siekiant patikrinti jo naudojimo perspektyvas, taikymą planuojama pradėti nuo sričių, kurių kontrolei nebūtina kurti raktažodžių kombinacijų ar daryti sudėtingų integracijų su skirtingomis duomenų bazėmis – būtų pasirenkamos sritys, kurios yra **akivaizdžiai nelegalios**. Šiuo atveju tos sritys būtų:

- Prekyba draudžiamais (laukiniais, saugomais, invaziniais) gyvūnais ir jų produktais;
- Prekyba draudžiamų (saugomų, invazinių) rūšių augalais ir jų produktais;
- Prekyba draudžiamais žūklės ir medžioklės įrankiais ir jų naudojimas.

Lietuvos institucijų praktikos taikant webcrawling technologiją

Webcrawling įrankis nėra naujiena – versle šis metodas plačiai taikomas jau kurį laiką, dažniausiai su komerciniu tikslu, pavyzdžiui, konkurentų kainų palyginimui. Dėl savo galimybių šis **įrankis buvo pastebėtas ir įvairių Lietuvos institucijų**, įgalinęs juos skirti laiką ne varginančiam rankiniam duomenų rinkimui internete, o tų duomenų analizei, kur tikslingai panaudojamos darbuotojų kompetencijos.

Valstybinės maisto ir veterinarijos tarnybos (VMVT) veikla šiuo metu apima daugiau nei 117 000 kontroliuojamų subjektų, kai kurie iš jų **vykdo prekybą ir internetinėje erdvėje**. Šiai priežiūrai šiuo metu VMVT naudoja „Web Content Extractor“ sistemą, kuri, pagal iš anksto nustatytus raktažodžius, atrenka norimą skelbimų informaciją iš konkrečių internetinių puslapių. Šis įrankis VMVT padeda surinkti skelbimų informaciją ir, pagal atliktą statistiką, **matomas įrankio indėlis į užblokuotų skelbimų kiekį, bei padidėjusį kiekį veisėjų, įregistravusių savo veiklą**. Tačiau, gauti rezultatai nebūna automatiškai sugrupuojami ir įvertinami, pagal tai, kurie iš jų yra legalūs, o kurie ne. Dėl to, visa tai šiuo metu atliekama rankiniu būdu – VMVT darbuotojai gautą informaciją patys sutikrina turimose duomenų bazėse ir atrinktus duomenis nukreipia atitinkamiems tarnybos departamentams tolimesniam kontrolės vykdymui.

Valstybinė mokesčių inspekcija (VMI) taip pat **prižiūri komercinės veiklos kontrolę elektroninėje erdvėje**. VMI naudojamas įrankis „Web Content Extractor“ padeda jiems greitai be papildomo programavimo atlikti nesudėtingos

struktūros tinklapių skanavimą. Įrankis palaiko suplanuotų užduočių funkciją (geba startuoti pats). Tačiau atliekant sudėtingų/didelės apimties tinklapių skanavimą, pvz. „skelbiu.lt“ (kadangi skelbimų kiekis labai didelės apimties), šio įrankio pagalba ne visuomet pavyksta nuskauduoti visus skelbimų portale esančius skelbimus. Taip pat, skanuojant skelbimų tinklapius pasitaiko atvejų, kai skelbimų portaluose naudojami algoritmai nustato, kad tai nėra žmogaus atliekama paieška, ir ima blokuoti sistemos IP adresą bei apriboja informacijos surinkimą. Atliekant tinklapių skanavimą panaudojant „Web Content Extractor“ įrankį, nustatomi ir kiti skanavimo apribojimai, pavyzdžiui, šiuo metu, „skelbiu.lt“ portale norint sužinoti skelbime patalpintą telefono numerį, reikia papildomai paspausti mygtuką, kad būtų atvaizduojamas pilnas telefono numeris, o „alio.lt“ skelbimų portale skelbimai atvaizduojami dinamiškai ir yra rodomi tik pirmi 20 skelbimų – norint nuskauduoti daugiau skelbimų, tenka naudoti papildomas „slinkti žemyn“ funkcijas. Tai kelia papildomų problemų dabartiniam įrankio efektyvumui ir naudingumui, todėl **VMI šiuo metu ieško geresnių automatizacijos sprendimo būdų**, kurie apjungtų ne tik webcrawling metodą, bet ir dirbtinį intelektą.

Valstybinės darbo inspekcijos (VDI) sprendžiamas iššūkis internetinėje erdvėje – **nelegalaus darbo apraiškų paieška, įmonių sukčiavimo atvejai**, pavyzdžiui, mokant darbuotojams mažesnę algą nei yra skelbiama darbo pasiūlymuose. VDI inspektoriai, ieškodami viešai prieinamos informacijos apie įmones ar pažeidimus, pastebėjo šio **proceso automatizacijos poreikį**. Šiuo metu kuriamas įrankis dar tik bus testuojamas. Tačiau, žinoma, kad šis įrankis **apjungs webcrawling ir dirbtinį intelektą**, kuris, pasitelkdamas viešai prieinamus duomenis iš valstybės institucijų bei interneto portalų, įvertins kiekvieno įtartino atvejo nelegalaus darbo apraiškos **riziką**. Apskaičiuota rizika bus pateikiama VDI inspektoriui, taip pat bus suteikiama galimybė **vizualiai matyti platesnį susietų tarpusavyje duomenų** apie įmones ir viešai pateikiamos informacijos ryšių tinklą, leis imtis efektyvesnių priešūros veiksmų.

Valstybinė vartotojų teisių apsaugos tarnyba (VVTAT) internetinėje erdvėje ieško tiek **nesaugių gaminių**, tiek ir **nelegalių pardavėjų**. Šiuo metu tam atlikti naudojamos kelios skirtingos programos ir interneto naršyklių plėtiniai, nes **nėra bendro įrankio**, kuris gebėtų atlikti visas reikiamas funkcijas. Pavyzdžiui, viena iš naudojamų sistemų fiksuoja pakeitimus internetiniame puslapyje, įrašydama tikslų laiką, tačiau nesugeba susekti nukreipimų į naujai sukurtus tinklapius. Taip pat, daugeliu atveju, paiešką reiktų vykdyti **ne tik pagal raktažodžius, bet ir pagal nuotraukas**, nes neretu atveju aprašymai tai pačiai prekei gali būti labai skirtingi. Gera žinia ta, kad šiuo metu Europos komisija kaip tik kuria **bendrą įrankių paketą**, kuris būtų naudingas ir automatizuotų dalį darbų tokiose įstaigose kaip VVTAT.

Žurnalistų etikos inspektoriaus tarnyba (ŽEIT) – dar viena institucija, kuri panaudojo webcrawler įrankį savo veikloje – aptinkant **neapykantos kalbą internete**. Prieš tai dešimtmetį naudota pagal raktažodžius veikusi paieškos sistema buvo netinkama naudoti - ŽEIT atliko tyrimą, kurio metu išanalizavo 6 tūkst. sistemos pateiktų komentarų iš kurių nei viename nerasta neapykantos kalbos. Šiuo atveju buvo reikalingas įrankis mokantis vertinti kontekstą. Toks įrankis buvo sukurtas *Codami* įmonės, ŽEIT dalyvaujant GovTech laboratorijos iššūkyje. Šis įrankis pasižymėjo **sėkmingu dirbtinio intelekto (DI) ir mašininio mokymosi** panaudojimu, **dideliu rezultatų tikslumu** bei ženkliai žmogiškųjų resursų efektyvumo padidėjimu.

Ryšių reguliavimo tarnyba (RRT) žinodami, kad nelegalaus turinio yra daugiau, nei sulaukiama pranešimų, dalyvavo GovTech iššūkių serijoje ir kėlė ypatingą uždavinį - **aptikti internete ir atpažinti vaikų seksualinio išnaudojimo vaizdinę medžiagą ir pornografiją** Lietuvos interneto erdvėje (IP ruože). Uždavinys susidėjo iš **trijų sudėtingų dalių**: identifikuoti tinklapius patalpintus Lietuvoje – ne tik “.lt”, bet ir kitus domenus, kurie patalpinti Lietuvoje ir gali būti patraukti atsakomybėn ar blokuojami; rasti tinklapius, kuriuose talpinamas draudžiamas turinys tikrinant visą ilgą sąrašą sudarytą nustatant tinklapio veikimo vietą; atpažinti vaizdus naudojant dirbtinį intelektą, bet negalint duoti įrankiui vaizdų mokymuisi, nes tokio turinio turėjimas yra draudžiamas. Šių uždavinių kompleksiskumas ir sudėtingumas lemia, kad šis įrankis įprastų RRT procesų, tokių kaip karštoji linija, nepakeičia, o juos **papildo ir plečia matomą interneto dalį**. Šiuo metu Ryšių reguliavimo tarnyba tęsia įrankio tobulinimą ir ieško DI modelių, galinčių patikimai atpažinti vaikų išnaudojimo vaizdus ir gebėti mokytis, turint labai ribotas pirminio apmokymo galimybes.

Pabendraurus su minėtomis priešūros institucijomis, kurios turi patirties naudojant webcrawling, galima apibendrinti šiuos bendrus pastebėjimus:

- Paprasti rinkoje laisvai prieinami universalūs įrankiai nepatenkina priežiūros institucijų poreikių, neįgalina efektyviai ir patogiai vykdyti jų funkcijų. Tačiau, svarbu suprasti, kad sukurtas pajėgus skaitmeninis įrankis, pvz. webcrawler, su minimaliomis modifikacijomis tikų ne vienai institucijai.
- **Dirbtinis intelektas – vis reikalingesnė sąlyga**, norint pasiekti aukštų rezultatų su webcrawling įrankiais.
- Dažnai galvojama kaip apeiti tam tikras skelbimų portalų funkcijas, tačiau visada verta suinteresuotoms institucijoms išreikšti poreikį dėl tam tikrų **skelbimų portalų** funkcijų, pvz. **papildomas privalomas laukas** įrašyti individualios veiklos ar verslo liudijimo, aplinkosaugos leidimų numerius. Tokios praktikos populiarėjimas suteiktų vartotojams galimybę priimti informuotą sprendimą, taip skatinant atsakingesnį vartojimą.
- Daugeliui valstybės priežiūros institucijų reikia **Lietuvoje talpinamų tinklapių sąrašo**. Ne tik .lt domenu, bet visų realiai Lietuvoje veikiančių tinklapių. Šis sąrašas galėtų būti sudaromas vienoje vietoje, nuolatos atnaujinamas ir pasiekiamas visoms priežiūros institucijoms.
- Bendra išvada – **žinojimas ką veikia panašios institucijos gali padėti** greitai išspręsti problemą, pasidalinant bendrai tinkančiais įrankiais arba **identifikuojant bendrus poreikius** ir taip sudarant prielaidas priimti efektyvesnius sprendimus.

Projekto tikslas

Iš problemos kylantis mūsų projekto tikslas – **automatizuoti galimai nelegalių skelbimų paiešką ir pateikti rezultatus, kurie tikų integracijai į kitas sistemas.**

Pasiekus iškeltą tikslą, būtų pagerintos sąlygos AAD operatyviai, efektyviai ir kokybiškai vykdyti savo funkciją – aplinkos apsaugos kontrolę. Taip pat lengviau pavyktų identifikuoti žmones ar įmones, kurios vykdo nelegalią veiklą ir tuomet tikslingai komunikuoti jiems, perspėti apie jų vykdomos nelegalios veiklos matomumą.

Tuo tarpu Lietuvoje būtų užtikrintas strateginių tikslų, tokių kaip Lietuvos Nacionalinė aplinkos apsaugos strategija, įgyvendinimas, nuslėpimų mokesčių surinkimas, aplinkosaugos reikalavimų laikymasis, tarptautinių įsipareigojimų, tokių kaip CITES ir EU Green deal, vykdymas.

Planuojamas projekto įgyvendinimas

Tikslui pasiekti suplanavome tokį įgyvendinimą:

1. Surinkti prioretizuotų sričių, kurių internetiniuose skelbimuose galima pritaikyti webcrawler įrankį, sąrašą:
 - 1.1. Aukštesnį prioritetą turėtų akivaizdžiai nelegalūs skelbimai (pvz.: draudžiama prekybinė veikla, kuriai nėra leidimų, draudžiamų gyvūnų, augalų pardavimas ir pan.).
2. Išanalizuoti skelbimų, pasitvirtinusių kaip nelegalių, specifiką (raktažodžiai, nuotraukos, telefono numeriai);
3. Sukurti identifikatorių grupes (raktažodžių sąrašus) kiekvienai iš prioritetinių sričių;
4. Išsiaiškinti dėl integracijos galimybių į aplinkosaugos leidimų IS (<https://www.alisas.lt/permits>), kad webcrawler galėtų pasiimti leidimų informaciją iš šios IS;
5. Išsiaiškinti dėl galimybės pridėti papildomus privalomus laukus skelbimų portaluose tai veiklai, kuriai privaloma turėti aplinkosaugos leidimus;
6. Suderinti webcrawler rezultatų formatą su AADIS ir įrankio kūrėjais;
7. Suderinti techninius ir organizacinius projekto įgyvendinimo aspektus su webcrawler įrankio kūrėjais;
8. Pilotinės webcrawler versijos, kuri apimtų aukštesnio prioriteto sritis, parengimas naudoti ir testavimas;
9. Atsižvelgus į naudotojų ir susijusių asmenų atsiliepimus, pagal poreikį, padaryti pakeitimai po pilotinės webcrawler versijos ir paleistas daugiau sričių apimantis webcrawler, kuris turėtų integraciją į ALIS;
10. Numatyti webcrawler priežiūros, raktažodžių atnaujinimo procesą.

Projekto įvykdymo kriterijai (acceptance criteria)

Įgyvendinimui pamatuoti išsikėlėme tokius kriterijus:

1. Surinktas prioretizuotų sričių, kurių skelbimuose galima pritaikyti webcrawler įrankį, sąrašas;
2. Išanalizuota skelbimų, pasitvirtinusių kaip nelegalių, specifiška ir paruošti jų aprašymą su įvardintais geriausius rezultatus atnešusiais metodais:
 - 2.1. Išanalizuotos ir aprašytos skirtingų Lietuvos institucijų patirtys naudojant webcrawling metodą;
 - 2.2. Išanalizuotos ir aprašytos specialistų, dirbančių su nelegaliais aplinkosaugos skelbimais, patirtys.
3. Sukurtos identifikatorių grupės (raktažodžių sąrašai) kiekvienai iš prioritetinių sričių;
4. Skelbimų portaluose pridėti papildomi privalomi laukai toms sritims, kurių veiklai privaloma turėti aplinkosaugos leidimus;
5. Užtikrinta galimybė integruoti ALIS leidimų tikrinimą į webcrawler su susijusiomis šalimis ir įrankio kūrėjais;
6. Paleistas webcrawler:
 - 6.1. kuris periodiškai pateiktų galimai nelegalių skelbimų sąrašus;
 - 6.2. teiktų ataskaitas tinkamu formatu peržiūrai ir importavimui į AADIS bei kitas išorines sistemas.
7. Sukurtas dokumentuotas procesas, leidžiantis užtikrinti tęstinumą - techninė priežiūra ir periodiškas raktažodžių atnaujinimas.

Projekto rizikos, priklausomybės

Išanalizavus esamą situaciją departamente, kitų Lietuvos institucijų patirtis, matome tokias grėsmes:

- Automatizuotos sistemos, iš kurios būtų galima gauti informaciją apie asmens/įstaigos turimus/turėtus leidimus skirtingose institucijose, nebuvimas: daugelio AAD prižiūrimų veiklų vykdymui yra reikalingi leidimai iš įvairių institucijų, visos jos informaciją saugo skirtingais formatais ir gali būti sunku integruoti šią informaciją į webcrawler. Be galimybės automatiškai patikrinti skelbime esančią kontaktinę informaciją su leidimų registrais bus gaunamas didesnis skelbimų kiekis, kuriuos teks rankiniu būdu papildomai peržiūrėti AAD darbuotojams;
- Webcrawler įrankio kūrėjai – noras bendradarbiauti, galimybės sukurti AAD tinkamą produktą;
- Skelbimų portalai – noras bendradarbiauti dėl papildomų privalomų laukų skelbimuose sukūrimo;
- Tęstinumo užtikrinimo trūkumas – neaišku kas, kaip prižiūrės įrankį ilgalaikėje perspektyvoje.

Laukiamas projekto rezultatas ateityje

Ilgalaikėje perspektyvoje, įtraukus webcrawler įrankį į AADIS sistemą visoms sritims, ir gaunant kokybiškus rezultatus, tikėtini tokie rezultatai:

- Automatizuotas skelbimų surinkimas ir paskirstymas per AADIS, su galimybe patikrinti inspektoriaus darbą didina skaidrumą ir tai komunikuojant galima didinti pasitikėjimą AAD;
- Prižiūrimi lietuviški interneto skelbimų portalai nuo nelegalios veikos prieš aplinką;
- Pateikiama informacija kokybiškam sprendimų priėmimui aukščiausiame vadybos lygmenyje ir operatyvinio darbo organizavimui;
- Efektyvesnis žmoniškųjų resursų valdymas/naudojimas - specialistas galės daugiau laiko skirti aukštos kompetencijos reikalaujančioms užduotims;
- Bus tiksliai identifikuojami žmonės, kurie vykdo nelegalią veiklą ir vykdoma tikslinės informacinio poveikio kampanijos.