

Lähteülesanne. Tehisaruga terviktekstide koostamise analüüsi ja prototüübi tellimine

Hanke eesmärk on saada analüüs järelduste ja juhistega, kuidas aktid konverteerida XML õigusakti skeemile vastavalt ja koostada Riigi Teatajas avaldatavaid terviktekste tehisaru abil ja saada analüüsi rakendamiseks reaalsete andmetega töötav prototüüp Riigi Teatajas avaldamiseks esitatud aktide konverteerimiseks õigusakti XML skeemile vastavaks ja muudatuste alusel terviktekstide koostamiseks.

Prototüübina peame silmas universaalset õppimisvõimelist ja edasiarendatavat vahendit (*MVP minimum viable product*), mis suudab sisendiks antava DOCX valitud XSD kohase väljundina anda ja koostama reeglitele vastava valideeruva XMLi ja viima muudatusi olemasolevatesse tervikteksti sõnastustesse.

1. Analüüsi lähtekohad

1.1 Terviktekstide koostamise ja avaldamise alus

Riigi Teatajas avaldatakse akti andja poolt vastuvõetud kujul kõik aktid vastavalt osale kuupäevaliselt koos neid muudatusi sisaldavate terviktekstidega vt RTS [§ 2](#) ja [3](#).

1.2 Teksti liigid Riigi Teataja

Riigi Teatajas on kõik aktid liigitatud kolmeks alltoodud teksti liigiks.

1) Algtekid so kõik akti liigid vastuvõetud kujul.

2) Algtekst – terviktekst so vastuvõetud kujul akti tekst, mis ajalistes arvestuses on ka esimene terviktekst määratuna ainult akti liikide puhul, millest terviktekste koostatakse. Kuulub otsingu alusel nii algtekstide kui ka terviktekstide kategooriasse Riigi Teataja avaldamisel koos nende teksti liikidele lubatud ning lisatud seoste ja funktsionaalsusega. Eelkõige terviktekstide puhul Riigi Teatajas ettenähtud funktsionaalsete võimalustega.

3) Terviktekst so akti sõnastuse tekst, mis on koostatud ja avaldatud koos kehtivusandmetega ajakohaseid muudatusi sisaldavatena, mis seni koostatakse käsitsi ja edaspidi tehisaru poolt.

1.3 Akti liigid, millest terviktekstid koostatakse:

Terviktekstid koostatakse järgmistest akti liikidest:

- 1) Seadus, peale Riigikogu võib akti andjaks olla ka Ülemnõukogu, kui see on algselt tema poolt vastu võetud olgugi, et hiljem Riigikogu poolt muudetud, samuti võib olla akti andjaks rahvas so Riigi Teatajas kajastatud, kui rahvahääletusel vastuvõetud ;
- 2) Vabariigi Presidendi seadlus, mida ei ole seni vastu võetud ega avaldatud, ei erine ülesehituselt seadusest;

- 3) Vabariigi Valitsuse määrus;
- 4) ministrite määrus;
- 5) Eesti Panga presidendi määrus;
- 6) kohaliku omavalitsuse üksuse määrus, need on valla- ja linnavolikogu määrus ning valla- ja linnavalitsuse määrus;
- 7) Riigikogu otsus;
- 8) Vabariigi Valitsuse korraldus;
- 9) Riigi Teataja väljaandja võib koostada ja avaldada ka muude akti liikide terviktekste.

Varasemalt on avaldatud ka Vabariigi Valimiskomisjoni määrusi, kuid alates 2017. aasta 1. jaanuarist neid enam ei avaldata, kuna Vabariigi Valimiskomisjonil puudub pädevus määrust anda. Riigi Teatajas on leitavad Vabariigi Valimiskomisjoni määrused märkega kehtetu.

1.4 Terviktekstid peavad vastama XML-skeemile.

1.4.1 Aktide struktuurelemendid on laias laastus jaotatud kolmeks:

- metaandmestik - (nt etapp, akti nimi ehk pealkiri, teksti liik, dokumendi liik, vastuvõtmise-, avaldamise- ja jõustumise andmestik jms)
- sisuandmestik - õigusakti sisu sätted, muutmismärked, tabelid, pildid, allkiri
- lisad - õigusakti lisad PDF vormingus

1.4.2 Õigusaktides on ajaloos koostatud erineva struktuuriga ning leidub akte milles on elemente, mis on tekstiliselt keerukad, näiteks valemid ja tabelid, ja ei vasta 1:1 õigusakti skeemile. Seetõttu on XML koosseisus lubatud HTML konteiner, mis võimaldab ebastandardseid struktuure lisada selliselt nagu need on ajaloos koostatud. Seega on XML õigusakti struktuur loodud selliselt, et see võimaldaks lahendada kõik erisused.

1.4.3 Akti lisad, milles sisaldub tabel, vorm, skeem, pilt, kaart vm, vormistatakse PDF-vormingus. PDF vormingus lisad ja pildid on XML struktuuri koosseisus Base64 Binary koodina. Keerulisemad tabelid, vormid sisalduvad HTML konteineris html märgenditena. Kasutatakse ka tabelite puhul XML skeemis ettenähtud tabeli elemente.

1.4.4 Õigusakti XML skeemid ja vastavad xsd ` d XML skeemi kirjeldused ja mõisted on lahti seletatud järgmises asukohas: <https://varamu.riha.ee/#Oigusakt> → kolitakse andmed.eesti.ee

1.5 Avaldamise korraldus - taustateadmiseks

1.5.1 Üldise avaldamise korralduse ja terviktekstide koostamise eest vastutajad sätestab RTS [§ 9](#).

1.5.2 Avaldatavad aktid valmistatakse ette ja määratakse avaldamiseks Riigi Teataja haldusmoodulis.

1.5.3 Riigi Teatajasse esitatakse avaldamiseks algtekst DHX kaudu kontrollitud akti andja esindaja poolt digiallkirjastatuna või digitempliga, kellele on haldusmoodulis määratud esitamisõigus.

1.5.4 Süsteem kontrollib automaatselt digiallkirja või digitempli alusel avaldamisõigust ja asetab kontrollitud akti avaldamise ettevalmistamiseks kas 1) mitte XML töölauale või 2) XML töölauale. XML vormingus esitatakse praegu ainult KOVü määruste algtekstid ja algtekst-terviktekstid. Samuti esitatakse algteksti avaldamise käigus edaspidi kirjeldatud protseduuri alusel kõik terviktekstid juba xml vormingus vastavalt õigusakti liigi skeemile.

1.5.5 Mitte XML töölauale esitatud algtekstid konverteeritakse haldusmooduli väliselt skeemikohasele XML vormingusse Altova või VexPro abil.

Vabavaraline VexPro kasutab konverterit RTF → XML <https://vexpro.just.ee/>

1.5.6 Haldusmoodulis kavandatakse algteksti avaldamine ja määratakse algtekstile avaldamismärke. Avaldamismärke kohta vt [Riigi Teataja KKK punkt 1](#)

1.5.7 Algtekstide muutmise sätete alusel, selgitatakse välja ja määratakse, millised terviktekstid tuleb koostada ja saadetakse koos avaldamismärkega teade tervikteksti koostajale.

1.5.8 Tervikteksti koostaja peab tervikteksti avaldamiseks esitama DHX kaudu või Riigi Teataja väljaandja poolt otse haldusmoodulis vähemalt kaks tööpäeva enne algteksti avaldamist, mõnikord kiireloomuliselt ka varem.

1.5.9 Enne avaldamist kontrollitakse üle, kas kõik muudatused on avaldatavas tervikteksti õigesti tehtud ja vajadusel parandatakse.

1.5.10 Esitatud tervikteksti sõnastusele määratakse haldusmoodulis avaldamismärke ja avaldatakse samas andes st samas osas ja kuupäeval selle aluseks oleva algtekstiga.

Täpsem avaldamiskorraldus on vt lisa avaldamisjuhise.

1.6 Terviktekstide koostamine praegu

1.6.1 Terviktekstid koostatakse kas Altova või VexPro abil.

1.6.2 Terviktekst peab vastama õigusakti XML skeemile. Õigusakti XML skeemid seaduste ja määruste kohta on koostatud normitehnika nõuete alusel vt [HÕNTE](#).

1.6.3 Õigusakti XML skeemides on määratud elemendid eelkõige tervikteksti põhisealt, et teha need masinloetavaks ja elemendipõhise funktsionaalsuse arendamiseks.

1.6.4 Kuna seaduste ja määruste kohta on täpsed ülesehituse reeglid HÕNTE's, siis nende akti liikidele on koostatud ka täpsemad elemendipõhised XML skeemid.

Üksikaktide so Riigikogu korralduste ja Vabariigi Valitsuse korralduste ülesehituse nõuded tulenevad tavast ja praktikast ning on selle alusel ka skeemis määratud. Kuid silmas tuleb pidada, et üksikaktide ülesehitus võib väga suures osas tavadest erineda ja seetõttu ei saa alati lähtuda skeemist ning aluseks tuleb võtta üldine skeem, milles ülesehituse elemendid ei ole kindlaks määratud.

1.6.5 Terviktekstid koostatakse Riigi Teataja haldusmooduli väliselt ja ka tehisarul abil terviktekstide koostamiseks ei ole vaja Riigi Teataja haldusmooduli arendusi.

1.6.6 Algtekstide alusel terviktekstide koostamiseks laeb tervikteksti koostaja alla Riigi Teatajas avaldatud tervikteksti ja viib sinna sisse algtekstid tulenevad muudatused ja lisab iga muudetud sätte järel muutmismärke elemendi so algteksti avaldamismärke koos jõustumisajaga, mis käib just selle sätte kohta.

1.6.7 Silmas tuleb pidada, et ühest algtekstist võib tuleneda mitme akti tervikteksti koostamise vajadus ning samuti ühe ja sama akti mitu erineval ajal tulevikus jõustuvat tervikteksti sõnastust. Vt näiteks [Sotsiaalhoolekande seaduse ja teiste seaduste muutmise seadus–Riigi Teataja](#) ja sellest koostatud terviktekstid:

Väärtpaberite registri pidamise seadus	RT I, 30.12.2025, 37	Riigikogu	seadus
Sotsiaalseadustiku üldosa seadus	RT I, 30.12.2025, 36	Riigikogu	seadus
Sotsiaalmaksuseadus	RT I, 30.12.2025, 35	Riigikogu	seadus
Sotsiaalmaksuseadus	RT I, 30.12.2025, 34	Riigikogu	seadus
Sotsiaalhoolekande seadus	RT I, 30.12.2025, 33	Riigikogu	seadus
Sotsiaalhoolekande seadus	RT I, 30.12.2025, 32	Riigikogu	seadus
Sotsiaalhoolekande seadus	RT I, 30.12.2025, 31	Riigikogu	seadus
Sotsiaalhoolekande seadus	RT I, 30.12.2025, 30	Riigikogu	seadus
Sotsiaalhoolekande seadus	RT I, 30.12.2025, 29	Riigikogu	seadus
Sotsiaalhoolekande seadus	RT I, 30.12.2025, 28	Riigikogu	seadus
Tervishoiuteenuste korraldamise seadus	RT I, 30.12.2025, 27	Riigikogu	seadus
Tervishoiuteenuste korraldamise seadus	RT I, 30.12.2025, 26	Riigikogu	seadus

1.6.8 On võimalik, et muudetakse terves ühe akti liigi piires aga ka terves õiguskorras olemasolevaid sõnu. Sel juhul ei pruugita eraldi terviktekste koostada või avaldada vaid muudetakse terminit avaldatud tervikteksti sõnastuses. Vt VVS § 105¹⁹ alates [lõikest 10](#).

1.6.9 Kui muutmisvormel näeb ette paragrahvi või mõne muu pealkirjastatud jaotusüksuse kehtetuks tunnistamise, siis säilitatakse selle paragrahvi või muu jaotusüksuse pealkiri ja sinna alla lisatakse muutmismärke koos märkega kehtetu – näiteks:

§ 4¹. Maakonna piirid planeerimisseaduse tähenduses
[Kehtetu - [RT I, 04.07.2017, 1](#) - jõust. 01.01.2018]

1.6.10 Akti kehtetuks muutumisel koostatakse ja avaldatakse kehtetu märkega terviktekst, juhul kui:

- Akt tunnistatakse kehtetuks vt [Hoiu-laenuühistu seadus–Riigi Teataja](#)
- Akti kehtetuks muutumine on ette nähtud pealkirjas nimetatud tähtajaga vt [2025. aasta riigieelarve seadus–Riigi Teataja](#)
- Seaduses muutub kehtetuks määruse andmise aluseks olev volitusnorm vt [Riigi teaduspreemiate põhimäärus–Riigi Teataja](#)

1.6.11 Kehtetu märkega tervikteksti sõnastusel on sisu eemaldatud ja muutmismärkena viidatakse aktile, mis on kehtetuse esile kutsunud. Selliseid akte on lihtne koostada VexPro abil mõne klõpsuga eemaldades sisu ja pannes muutmismärke.

1.6.12 Kui koostatakse ja avaldatakse tervikteksti sõnastus, mis hakkab kehtima tulevikus samaaegselt juba avaldatud tervikteksti sõnastusega, siis Riigi Teataja süsteem asendab seni avaldatud tervikteksti sõnastuste aja jadas ja asendatud tervikteksti sõnastus ei ole enam otsinguga leitav ning tal on märges „Mitte jõustunud sõnastus“ vt näiteks: [Liiklusseadus–Riigi Teataja](#). Tervikteksti koostamise nõudeid see erisus ei mõjuta.

1.7 Lähtekohad terviktekstide koostamiseks tehisaru abil

1.7.1 Kõigepealt tuleb analüüsida ja näha ette reeglid (prompt) seaduse ja määruse XML skeemipõhise tervikteksti sõnastuse koostamiseks arvestades tänast praktikat ja tervikteksti koostamise juhendit. Hiljem saab analüüsida ka üksikaktide tervikteksti koostamist

1.7.2 Kehtetu märkega aktide terviktekstide sõnastuse koostamist tehisaru abil ei ole vaja põhjalikumalt analüüsida, kuna see on ka käsitsi tehtavalt lihtne ja algandmed vajavad inimese otsust. Võimalik, et kui akt tunnistatakse selge sõnaga kehtetuks, siis tehisaru koostab ka sellest tuleneva kehtetu märkega tervikteksti.

1.7.3 Sisulised nõuded tehisarule konverteerimiseks ja tervikteksti koostamise ülesande täitmiseks:

1) tervikteksti eelne tegevus, mis põhineb samuti xml-skeemile vastavaks viimisel, kuid ei sisalda sisulise muutmise vajadust - konverteerib talle antud RTF või DOCS vormingus algteksti või algtekst-tervikteksti vastava akti liigi xml-skeemi alusel xml vormingusse ja valideerib selle.

2) tehisarule antakse muutev akt koos selle avaldamismärkega;

3) otsib muutvas aktis sisalduvate muutmismuutvormelitel alusel Riigi Teatajast välja avaldatud muudetavad terviktekstid arvestades muutvas aktis sisalduvaid jõustumisaegu ja laeb need alla. Alla laadimiseks saab kasutada API't, kirjeldus Riigi Teataja lehel rubriigis „Abi ja teave“ alamrubriigis „KKK“ punkt 9 Avaandmed - [KKK-Riigi Teataja](#);

4) teeb muutmisvormelite alusel selles muudatused, lisab päisesse ja muudetud sätetel muutmismärked koos linkidega, ning asendab või eemaldab eelnevad muutmismärked vastavalt terviktekstide koostamise juhendile;

5) samuti tervikteksti koostamisel:

- eemaldab algtekst-tervikekstile järgneva tervikteksti koostamisel teiste aktide muutmise paragrahvide sisu. Näiteks:
2. jagu
Seaduste muutmine ja kehtetuks tunnistamine
§ 255. – § 268. [Käesolevast tekstist välja jäetud.];
- lisab algtekst-tervikekstile järgneva tervikteksti sõnastuse koostamisel pealkirja alla vormikohase vastuvõtmismärke, mis sisaldab algtekst-tervikeksti avaldamismärget lingiga ning vastuvõtmise ja jõustumise kuupäeva. Näide:

Alusharidusseadus

Vastu võetud 11.12.2024

[RT I, 09.01.2025, 1](#)

jõustumine 01.09.2025

- muudab algtekst-tervikekstile järgneva tervikteksti koostamisel teksti liigi terviktekstiks;
- muudab dokumendi staatuse välja väärtus „esitatud avaldamiseks“ väärtuseks „avaldatud“ – seda nõuet täpsustame analüüsi käigus;
- eemaldab eelnevas terviktekstis olevad veaparandusmärked, kuna järgneva teksti kohta ei ole need enam ajakohased. Näide veaparandusmärgest, mis tuleb eemaldada:
Näide: 15.01.2026 10:52
Veaparandus: Parandatud ilmne ebatäpsus, lisatud tühik sõnade „säilitada” „menetlustoimingu” vahele, Riigi Teataja seaduse § 10 lõike 4 alusel.;
- lisab metaandmetesse tervikteksti sõnastuse kehtivuse alguse kuupäeva, arvestades sõnastuses tehtud muudatusi;
- toob välja märgistatuna kirjavead, mitte neid parandades;
- toob välja ilmselged ülesehituse või loogika vead, näiteks lisatakse mitte numeratsiooni loogilises järjekorras olevaid sätteid – seda xml skeem ka ei luba;
- märgib ära kohad, kuhu ei saanud muudatusi sisse viia.

6) valideerib koostatud tervikteksti sõnastuse vastavust vastava liigi õigusakti XML skeemile;

7) jõustumisaegade arvestamiseks arvestab iga muudatuse tegemisel ja muutmismärkes märkimisel jõustumisnormis sätestatud jõustumistähtaegu. Kui muutvas aktis konkreetseid kuupäevalisi jõustumisaegu ei ole või on sätestatud, st „jõustub üldises korras“, siis arvestab tervikteksti sõnastuse ja muudatuste jõustumisaegu järgmiselt:

- seaduse puhul – muutva akti avaldamismärkes sätestatud avaldamise kuupäevast kümnendal päeval;
- määruse puhul – muutva akti avaldamismärkes sätestatud avaldamise kuupäevast kolmandal päeval;
- Riigikogu otsuse ja Vabariigi Valitsuse korralduse puhul märgib jõustumise ajaks muutva akti kuupäeva. Tavaliselt nendel aktidel jõustumisnormi ei ole.
- sündmuse saabumisega seotud sõnalise jõustumise korral, märgib jõustumisajaks jõustumisnormis sätestatud sündmuse. Selliste jõustumiste korral selgitab Riigi Teataja väljaandja enne jõustumisaja saabumist välja konkreetse kuupäeva ja teeb juba avaldatud tervikteksti sõnastuses ise vastavad muudatused, enne teada saadud konkreetset jõustumiskuupäeva.

8) kui ühe akti alusel on vaja koostada mitu tulevikus eri aegadel jõustuvat tervikteksti sõnastust, siis alustab varasemast ja viib järgnevad muudatused sisse juba ajalise jõustumisjärjekorra alusel ja koostab ka järgnevad tervikteksti sõnastused;

9) Koostatud terviktekstis peavad jääma tööle kõik lingid sh lisatud muutmismärgetelt lingid muutvatele aktidele;

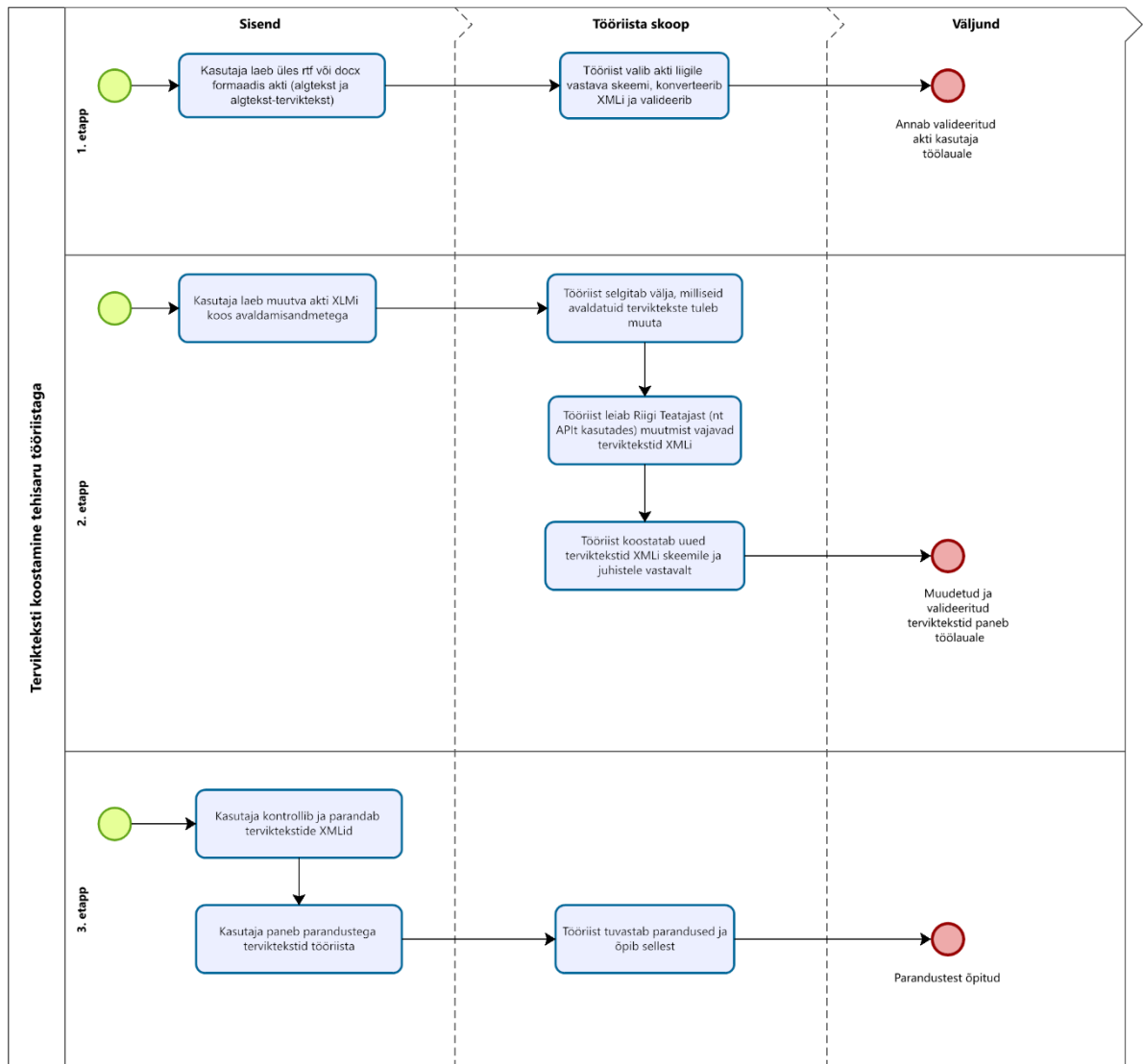
10) saadab koostatud terviktekstide sõnastused vastavale töölauale ülesande andjale või laeb need üles ettenähtud kohta – see täpsustub analüüsi käigus;

11) tehisaru poolt koostatud terviktekstis on tehtud muudatused märgistatud värviliselt. Tekst, mis on eemaldatud on läbi kriipsutatud.

1.7.4 Tehisaru koostab terviktekstid muutvas aktis sisalduvate muutmismärgistite alusel. Muutmise reeglid on sätestatud HÕNTE's vt [Hea õigusloome ja normitehnika eeskiri–Riigi Teataja](#), kuid arvestama peab, et praktikas võib HÕNTEst olla erisusi ja seetõttu tuleb analüüsi aluseks võtta ka Riigi Teatajas avaldatud muutvaid akte ja nende alusel koostatud terviktekstide sõnastusi.

1.7.5 Tervikteksti koostamiseks on ka Riigi Teataja väljaandja sisemine juhend, mida tuleb analüüsis arvestada võimalike tehisarust tulenevate erisustega ja milles on täpsemad reeglid muudatuste tegemiseks ja kehtivusaegade märkimiseks.

1.7.6 Üldine skeem, mida töövahend peab tegema:



2. Üldised nõuded:

loodav lahendus peab olema:

2.1 ainult selle ülesande täitmiseks;

2.2 Riigi Teataja väljaandja kontrolli all ja tema otsusel teistele levitatav;

2.3 edasi arendatav ja õpetatav , et vältida võimalike tehtud vigu edaspidi;

2.4 analüüsi alusel loodav prototüüp on esimene versioon töötavast lahendusest, milles enamik nõudeid rakendatud.

2.5 Prototüübi piirid

2.5.1 MVP hõlmab kõikide akti liikide konverteerimist skeemikohasesse vormi, seaduste ja määruste terviktekstide koostamist, samuti valideerimist, kasutaja kontrollvaadet ja vigade parandamist ning paranduste edaspidist arvestamist.

2.6 Tehnilised nõuded prototüübile

2.6.1 RTF või DOCX sisend peab olema Microsoft Word 2007+ formaadis või LibreOffice DOCX standardis.

2.6.2 XML peab vastama RT ametlikule XSD dokumendiliigi skeemile, mille vastu peab olema valideeritav ja näitama ka vigu näitama vigu XML-s reavaates koos selgituste ja parandussoovitustega.

2.6.2 Toetama pildimaterjali vorminguid ja nende genereerimist, nt PDF lisad Base64Binary kodeeringusse.

2.6.3 Lubama käsitsi XML-i muutmist (advanced mode).

2.6.4 Peab võimaldama kontrollida tervikteksti sõnastusse tehtud muudatusi ja eelvaadet muudatustest ning näitama visuaalselt erinevusi ("enne ja pärast"). Muuhulgas:

- 1) salvestamisel kontrollima, kas kõik muudatused on üle vaadatud ja kinnitatud.
- 2) peab toetama kasutajapõhist kinnitamisvoogu (ei saa liikuda edasi, kui kõik muudatused pole kinnitatud).
- 3) peab salvestama muudatuste logi (audit trail).
- 4) peab lubama kasutajal kustutada ebavajalikud muutmismärked ja parandama keelevigu.
- 5) peab toetama hiirepõhist ja klaviatuuri abil navigeerimist.
- 6) madal õppimiskõver – sobilik juristidele ja toimetajatele ilma tehnilise taustata.

2.6.5 Peab suutma töödelda suuri dokumente (1000+ paragrahvi) ilma märgatava viivitusega.

2.7 . Töömudel ja arhitektuuripõhimõtted

2.7.1 Lahenduse töömudel on automaatne töötlusahel koos inimese kontrollvaatega: tehisarv koostab terviktekstide sõnastused valmis, tervikteksti koostaja kontrollib tulemust võrdlusvaates (enne/pärast) ja kinnitab või lükkab tagasi. Täisväärtuslikku interaktiivset XML-redaktorit ei ehitata.

2.7.2 Muudatuste rakendamine peab olema deterministlik: keelemudel tõlgendab muutmismuudatuse struktureeritud muudatusoperatsiooniks (nt asenda / täiendada / tunnista kehtetuks + täpne sihtkoht), kuid operatsiooni rakendab XML-ile deterministlik programmikood. Keelemudel ei tohi tervikteksti tervikuna ümber kirjutada.

2.7.3 Kohustuslik kvaliteedinõue: tervikteksti osad, mida muutev akt ei puuduta, peavad jääma lähtetekstiga täht-tähele identseks. Lahendus peab seda iga koostatud tervikteksti puhul automaatselt kontrollima (diff-kontroll) ja rikkumise korral töö

peatama. Lubatud muudatuste arv väljaspool muutmismuutisvormelitest hõlmatud kohti on null.

2.7.4 Ebakindluse käsitus: kui süsteem ei suuda muutmismuutisvormelit, sihtkohta või jõustumisaega üheselt tuvastada, peab ta peatuma ja märgistama probleemi inimesele, mitte rakendama tõenäosuslikku pakkumist vaikimisi. Iga muudatuse juurde lisatakse kindlustunde hinnang.

2.7.5 Konverteerimise ja tervikteksti koostamise vahele luuakse masinloetav struktureeritud muudatusoperatsioonide kiht (eraldi dokumenteeritud vorming). See on lahenduse keskne liides, mis võimaldab operatsioone kontrollida, logida ja taasesitada.

3. Näidisfailide analüüsist tulenevad täiendavad nõuded ja riskikohad

3.1 Muutmismuutisvormelite parsimine. Muutva akti XML-is on muudatuste sisu HTML-konteineris (CDATA) lame HTML, st muutmismuutisvormeleid ei struktureerita XML-is. Tervikteksti koostamine eeldab vormelite parsimist tekstist. Muudetav akt on tuvastatav ainult pealkirjast (nt „Väärteomenetluse seadustikus tehakse järgmised muudatused”) – vaja on aktinime, sh lühendite ja ümber nimetatud aktide, sidumist Riigi Teataja akti ID-ga.

3.2 Lame HTML → sügav struktuur. Kui muutev akt lisab terve peatüki või paragrahvid (näidises lisatakse 7.² peatükk paragrahvidega), tuleb tsiteeritud lame HTML teisendada tüviakti skeemi sügavaks struktuuriks (peatükk → paragrahv → loige → alampunkt) koos ülaindeks-atribuutide, kuvatavNr-vorminduse ja õige paigutusega ülaindeksitega nummerduse järjestuses (7¹ ja 8 vahele). See on tehniliselt kõige nõudlikum teisendus ja peab olema analüüsis eraldi käsitletud.

3.2 Elementide ID-de konventsioon ja püsivus. Terviktekstides on loetavad ID-d (nt para30lg1p4, para31b2, kus b2 tähistab ülaindeksit 2). Nõuded: olemasolevate elementide ID-d ei tohi tervikteksti koostamisel muutuda (püsilinkide toimimine); uued elemendid saavad konventsioonikohased ID-d; ID-de ebaühtlane pärand (osal elementidel GUID-id, osal ID puudub) vajab analüüsis fikseeritud reeglit.

3.3 Tsitaadimärgid ja malli stiilid. Sisend-DOCX kasutab tsitaatide mõlemas otsas märki ” (avaldatud XML-is korrektseks „...”), mistõttu tsiteeritud uue sõnastuse piire ei saa tuvastada jutumärkide paardumise järgi – tuvastus peab tuginema Wordi stiilidele (muudetavtekst). Hankes tuleb fikseerida: kas malli stiilide kasutus on sisendis kohustuslik ja kuidas käsitletakse stiilideta või võõra vormistusega dokumente (varumeetod, inimese sekkumine).

3.4 Tüpopraafia normaliseerimine. Avaldatud XML sisaldab sisutühikuid (), järjestikuste DOCX-lõikude liitmist üheks lõiguks reavahetustega (
), joondusatribuute jm. Muutmata osa identsuse nõude jaoks tuleb defineerida kanooniline normaliseerimisreeglistik (mida loetakse samaväärseks), muidu muutub automaatne kontroll kasutuskõlbmatuks.

3.5 Sõnataseme muudatused. Vormelid kujul „punktist 4 jäetakse välja sõna «alaealine»” ja „täiendatakse pärast sõna «Keskkonnaamet,» sõnaga ...” nõuavad täpset stringivastavust sätte tekstis, reeglit sõna mitmekordse esinemise puhuks, kirjavahemärkide kuuluvuse käsitlust (koma tsitaadi sees) ning suurtähestuse korrigeerimist, kui eemaldatakse või lisatakse lause esimene sõna.

3.6 Muutmismärgete paigutus ja asendamine. Muutmismärke lisatakse väikseima muudetud üksuse juurde (alampunkt, lõige, paragrahv) ja akti alguse muutmismärgete koondloendisse. Reeglid, millal varasem märges asendatakse ja millal säilitatakse, sisalduvad väljaandja sisejuhendis – juhend tuleb lisada hanke lisana.

3.7 Versioonimaatriks. Sihtakti võib muudatuse saabumisel olla juba avaldatud tulevikus jõustuvaid tervikteksti sõnastusi (näidis-tervitekstil on kehtivuse lõpp määratud, st tulevikuvärsioon eksisteerib). Muudatused tuleb sisse viia kõikidesse tuleviku sõnastustesse, mis on Riigi Teataja avaldatud.

3.8 Tuletamatud metaandmed. Osa avaldatava XML-i metaandmeid ei sisaldu sisend-DOCX-is ega ole tekstist tuletatavad (nt aktViide, globaalID, RT artikli number, terviktekstiGrupiID, metaandmete versiooninumbrid, akti lühend, väljakuulutamise andmed). Lähteülesandes tuleb defineerida sisendparameetrite kanal ja vorming, millega need andmed lahendusele ette antakse.

3.9 Jõustumisvormelite kataloog. Analüüsietapi väljund peab olema Riigi Teataja korpuse põhjal koostatud võimalikult täielik jõustumisvormelite kataloog koos igaühe arvutusreegliga.

4. Vastuvõtukriteeriumid

4.1 Analüüsietapi kohustuslik väljund on kuldstandard-testikorpus: vähemalt 50–100 komplekti kujul sisend-DOCX + avaldatud muutva akti XML + kõik selle alusel käsitsi koostatud ja avaldatud terviktekstide XML-id (analoogiliselt käesoleva hanke näidisfailidega). Korpus peab katma eri akti liigid, muutmisvormelite tüübid ja jõustumiserisused.

4.2 Prototüübi vastuvõtu mõõdikud (sihttasemed täpsustatakse analüüsietapis, arvestades lähtetasemeid: a) vähemalt 80% koostatud terviktekstidest on avaldatud võrdlustekstiga sisuliselt identsed ilma inimese paranduseta; b) 100% väljunditest valideeruvad vastava liigi XSD skeemi vastu; c) 0 muudatust väljaspool muutmisvormelitega hõlmatud kohti.

5. Lähtekoodi üleandmise, taristu ja keelemudelite nõuded

5.1 Pakkuja annab tellijale üle lahenduse täieliku lähtekoodi koos ehitusskriptide, konfiguratsiooni ja tehnilise dokumentatsiooniga. Litsentsitingimused peavad võimaldama tellijal lahendust iseseisvalt kasutada, muuta ja edasi arendada ning tellija otsusel teistele levitada.

5.2 Lahendus peab olema konteineriseeritud ja käivitav tellija Azure'i pilvekeskkonnas Kubernetese klastri (AKS). Üleandmine sisaldab Kubernetese manifeste või Helm charte, keskkonnanõuete kirjeldust (ressursid, võrgunõuded, salvestus, saladuste haldus) ja juurutusjuhendit.

5.3 Juurutuskonveieri (CI/CD pipeline) seadistab tellija ise oma taristul. Pakkuja annab üle kõik selleks vajaliku: konteinerite ehitusdefiniitsioonid, sõltuvuste loetelu ja ehitusjuhised. Lahendus ei tohi sõltuda pakkuja ehitus- ega juurutustaristust.

5.4 Keelemudeleid kasutatakse lahenduses API-de kaudu. Mudelivalik peab olema konfigureeritav pakkujaneutraalse abstraktsioonikihi kaudu: tellija otsustab, millist mudelit ja millise teenusepakkuja kaudu kasutatakse, ning saab mudelit vahetada ilma lähtekoodi muutmata või minimaalse konfiguratsioonimuudatusega.

5.5 Kui lahenduse osaks on eraldi treenitud või peenhäälestatud mudel, peab tellija olema võimeline seda ise hostima: pakkuja annab üle mudeli kaalud, treening- ja peenhäälestuskoodi ning kasutatud andmestikud, ja mudel peab olema käivitav tellija taristul. Sõltuvus pakkuja hallatavast mudeliteenusest ei ole lubatud.

5.6 Lahendus ei tohi sõltuda pakkuja SaaS-teenustest ega suletud komponentidest, mida ei ole tellijale üle antud. Kolmandate osapoolte komponendid peavad olema avatud lähtekoodiga või tellijale litsentsitud viisil.

6. Töökorraldus

6.1 Lähteülesandes toodud kriteeriumeid, nõudeid ja tingimusi võib analüüsi tegemise käigus tellija ja täitja konsensuslikult täiendada või ka mõnest kriteeriumist loobuda, kui see ei ole otstarbekas ja sellest loobumine ei mõjuta lõpptulemust.

6.2 Tellija ja täitja moodustavad ühise töörühma ning lepivad kokku selle koostumise regulaarsuse.