

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Kristin Salu 233145IABB
Karoliine Tarjus 233148IABB

**AI-kompetentside suhetepõhise
kaardistussüsteemi analüüs ja kasutajaliidese
disain Tallinna Tehnikaülikooli näitel**

Bakalaureusetöö

Juhendajad: Ago Luberg
Ph.D.

Sven Nõmm
Ph.D.

Tallinn 2026

Autorideklaratsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Kristin Salu, Karoliine Tarjus

17.05.2026

Lühikokkuvõte

Käesolev bakalaureusetöö analüüsib ning disainib kasutajaliidese AI-kompetentside suhete põhise kaardistussüsteemi Tallinna Tehnikaülikooli näitel. Töö lähtub probleemist, et TalTechi teadlaste, projektide, uurimisgruppide ja publikatsioonidega seotud informatsioon on hajutatult mitme andmeallika vahel, mistõttu puudub terviklik ülevaade ülikooli AI-kompetentsidest.

Töö teoreetilises osas analüüsitakse olemasolevaid andmeallikaid ning nende tugevusi ja puuduseid. Võrdlusanalüüs CRM-süsteemidega viis järelduseni, et suhete modelleerimise, ajaloo haldamise ning rollipõhise ligipääsu halduse põhimõtteid saab loodavale süsteemile üle kanda. Lisaks uuriti graafianalüüsi meetodeid.

Kasutajavajaduste kaardistamiseks ning täpsustamiseks viidi läbi kolm intervjuud TalTech AI esindajate ning Tehisintellekti- ja robotikakeskuse ehk AIRE-ga. Intervjuude põhjal kitsendati kasutajarollid kahele: TalTechi juhtkond ja AIRE.

Töö sünteesi osas loodi loogiline andmemudel, mis modelleerib järgnevaid põhiobjekte koos ajaliste seostega: isik, projekt, publikatsioon, uurimisgrupp, märksõna, asutus, instituut. Otsinguloogikaks valiti SBERT-põhise semantilise vektoriseerimise ja BM25 tekstipõhise otsingu hübriid. Sellele lisati sobivusprotsendi arvutamise mudel, mis võtab arvesse sisulist sarnasust, teadusalast kogemust ja informatsiooni värskust. Kasutajaliidese prototüüp valmis Figma Design keskkonnas ning sisaldab otsingu eelvaadet, graafivaadet, loendivaadet, märksõnade mulldiagrammi koos ajajoonega ning objektitüüpide detailvaateid.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 69 leheküljel, 6 peatükki, 18 joonist, 4 tabelit.

Abstract

Analysis and user interface design for an AI competency relationship mapping system:
A case study of Tallinn University of Technology

This bachelor's thesis analyses and designs the user interface of a relationship-based AI competence mapping system on the example of Tallinn University of Technology. The work addresses the problem that information about TalTech's researchers, projects, research groups, and publications is fragmented across multiple data sources, resulting in a lack of a comprehensive overview of the university's AI competences.

The theoretical part analyses existing data sources and evaluates their strengths and limitations. A comparative analysis with CRM systems led to the conclusion that the principles of relationship modelling, history management, and role-based access control are transferable to the proposed system. Graph analysis methods were also examined.

Three interviews were conducted with TalTech AI representatives and the AI & Robotics Estonia (AIRE) representatives to map and refine user needs. Based on the interviews, user roles were narrowed to two: TalTech management and AIRE.

The synthesis part presents a logical data model covering the following core objects with temporal relationships: person, project, publication, research group, keyword, institution, and institute. A hybrid search approach combining SBERT-based semantic vectorisation and BM25 text-based search was selected, supplemented by a relevance score model accounting for semantic similarity, research experience, and information recency. The user interface prototype was created in Figma Design and includes a search overview, graph view, list view, a keyword bubble chart with a timeline, and detail views for each object type.

The thesis is written in Estonian and contains 69 pages of text, 6 chapters, 18 figures, 4 tables.

Lühendite ja mõistete sõnastik

AI	<i>Artificial Intelligence</i> , tehisintellekt
AIRE	Tehisintellekti- ja robotikakeskus
API	<i>Application Programming Interface</i>
BM25	<i>Best Matching 25</i>
CERCS	<i>Common European Research Classification Scheme</i>
CRM	<i>Customer Relationship Management</i>
CSV	<i>Comma-Separated Values</i>
DOI	<i>Digital Object Identifier</i>
ESTER	E-kataloog, Eesti suuremate raamatukogude ühiskataloog
ETIS	Eesti Teadusinfosüsteem
GDPR	<i>General Data Protection Regulation</i>
ISO	<i>International Organization for Standardization</i>
IT	Infotehnoloogia
JSON	<i>JavaScript Object Notation</i>
LLM	<i>Large Language Model</i>
NIST	<i>National Institute of Standards and Technology</i>
NIST SP	<i>NIST Special Publication</i>
OPT	<i>Open Pre-trained Transformer</i> (märksõna prototüübis)
RAG	<i>Retrieval-Augmented Generation</i> (märksõna prototüübis)
RBAC	<i>Role-Based Access Control</i>
SBERT	<i>Sentence Bidirectional Encoder Representations from Transformers</i>
SCD	<i>Slowly Changing Dimensions</i>
TalTech	Tallinna Tehnikaülikool
UI	<i>User Interface</i>
UML	<i>Unified Modeling Language</i>
UX	<i>User Experience</i>

XML	<i>Extensible Markup Language</i>
XLS	<i>Excel Spreadsheet</i>

Sisukord

1	Sissejuhatus	13
1.1	Probleem.....	13
1.2	Eesmärk	14
1.3	Töö struktuur	15
2	Taustülevaade ja probleemianalüüs.....	16
2.1	Olemasolevad andmeallikad.....	16
2.1.1	ETIS.....	16
2.1.2	Tallinna Tehnikaülikooli veebileht.....	18
2.1.3	Tallinna Tehnikaülikooli raamatukogu digikogu	18
2.1.4	Muud allikad.....	19
2.2	Võrdlusanalüüs	20
2.2.1	Mis on CRM?	20
2.2.2	Kuidas CRM-id modelleerivad suhteid?	21
2.2.3	Kuidas nad haldavad ajalugu?	22
2.2.4	Kuidas nad lahendavad õiguseid?.....	24
2.3	Andmete kogumise viisid ja juriidilised piirangud	24
2.4	Graafianalüüs.....	25
2.4.1	Võrgustiku tihedus.....	26
2.4.2	Koostöö klastrid.....	26
2.4.3	Keskusmõõdikud	27
2.4.4	Isolatsioonid	27
3	Metoodika.....	29

3.1	Uurimismeetodid ja tööetapid	29
3.2	Tööriistad.....	30
3.3	Intervjuud	30
3.3.1	Valimi koostamine.....	31
3.3.2	Intervjuude läbiviimine.....	31
3.3.3	Intervjuude ülevaade.....	32
4	Tulemused ja analüüs	35
4.1	Andmeallikate hindamise tulemused.....	35
4.1.1	Süsteemis kasutatavad andmeväljad.....	36
4.2	CRM-süsteemi põhimõtete rakendamine	37
4.3	Intervjuude põhjal kujunenud süsteeminõuded	37
4.4	Kasutajavajaduste analüüs	38
4.5	Masinõppe rakendamine.....	39
5	Süntees.....	40
5.1	Kasutajarollid ja õigused	40
5.1.1	Rollid	40
5.1.2	Õigused.....	40
5.2	Süsteemi struktuur	41
5.2.1	Objektid	41
5.2.2	Atribuudid.....	42
5.2.3	Ajaloo ja ajajoone modelleerimine.....	43
5.2.4	Loogiline andmemudel	44
5.2.5	Andmete muutmise strateegia	51
5.3	Otsing	53
5.3.1	Otsingu põhimõtted	53
5.3.2	Sobivusprotsent	54

5.3.3	Graafivaade.....	56
5.3.4	Sorteerimine	57
5.4	Süsteemi disain ja kavandamine.....	57
5.4.1	Nõuded	57
5.4.2	Kasutajalood	59
5.5	Kasutajaliidese disaini põhimõtted.....	62
5.5.1	Kognitiivne koormus	63
5.5.2	Otsingupõhine vs visuaalipõhine lähenemine	63
5.5.3	Graaf vs loend.....	63
5.5.4	Filtreerimine ja navigatsioon	64
5.5.5	Värvikoodid ja visuaalne eristamine	64
5.5.6	Vaikimisi vaade rolliti	64
5.6	Kasutajaliidese vaated ja navigatsioon	64
5.6.1	Märksõnade koondvaade	65
5.6.2	Märksõna valik ja ajajoon.....	66
5.6.3	Otsingutulemuste graafivaade	67
5.6.4	Graafivaate filtreerimine ja serval klikkimine.....	68
5.6.5	Otsingutulemuste loendivaade.....	70
5.6.6	Kategooriaspetsiifilised nimekirjad	72
5.6.7	Projekti detailvaade	73
5.6.8	Inimese detailvaade	74
5.6.9	Uurimisgrupi detailvaade	75
5.6.10	Publikatsiooni detailvaade	76
5.7	Projekti tulevik ja edasised sammud	77
6	Kokkuvõte	79

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks	87
Lisa 2 – Poolstruktureeritud intervjuu küsimused	88
Lisa 3 – Kristin Salu eneseanalüüs	89
Lisa 4 – Karoliine Tarjus eneseanalüüs	91

Jooniste loetelu

Joonis 1. Asutuse hierarhia.....	45
Joonis 2. Instituudi seosed.....	46
Joonis 3. Uurimisgrupi seosed.....	47
Joonis 4. Isiku seosed.....	48
Joonis 5. Publikatsioonide seosed.....	49
Joonis 6. Projekti seosed.....	50
Joonis 7. Märksõna seosed.....	51
Joonis 8. Vaheleht "Tegevusülevaade".	65
Joonis 9. Märksõna mulldiagramm.....	66
Joonis 10. Märksõna mulldiagrammi filtreeritud otsingutulemus.....	67
Joonis 11. Otsingutulemus graafil.....	68
Joonis 12. Graafi filtreerimine.....	69
Joonis 13. Otsingu tulemus loendina.....	71
Joonis 14. Inimeste nimekiri.....	73
Joonis 15. Projekti detailvaade.....	74
Joonis 16. Inimese detailvaade.....	75
Joonis 17. Uurimisgrupi detailvaade.....	76
Joonis 18. Publikatsiooni detailvaade.....	77

Tabelite loetelu

Tabel 1. Intervjuude sidusrühmad ja nende esindajad.....	31
Tabel 2. Andmemudeli objektid ja nende definitsioonid.	41
Tabel 3. Andmemudeli objektid, nende atribuudid ja näited.	42
Tabel 4. Andmemudeli mitmesused.	44

1 Sissejuhatus

Tehisintellekti kasutamine on viimastel aastakümnetel muutunud üha olulisemaks osaks kaasaegsest teadustööst. Masinõpe, loomuliku keele töötlus ja andmemahukad analüüsimeetodid on muutnud viisi, kuidas teadlased andmeid koguvad, analüüsivad ja tõlgendavad. Ülikoolidel on selles arengus oluline roll, sest akadeemiline keskkond ühendab erinevate valdkondade eksperte, uurimisgruppe ja rahastusprojekte, mille koostöös sünnivad teadusalased avastused.

Kompetentside nähtavus on aga sageli probleem. Isegi ühe ülikooli piires võib tehisintellektiga tegeleva teadlase töö olla hajutatud mitme andmeallika vahel (teadusinfosüsteemid, instituutide veebilehed, digikogu), millest ükski ei paku terviklikku, seostepõhist ülevaadet. See tähendab, et nii ülikoolisisesed koostöövõimalused kui ka väliste partnerite jaoks olulised kompetentsid jäävad leidmata.

Käesolev töö käsitleb seda probleemi Tallinna Tehnikaülikooli (TalTech) kontekstis. Eesmärk on analüüsida olemasolevaid andmeallikaid, võrrelda neid sarnaste süsteemidega (CRM-lahendused) ning kavandada suhetepõhine AI-kompetentside kaardistussüsteem, mis koondab teadlaste, projektide, uurimisgruppide ja märksõnade vahelised seosed ühtsesse graafipõhisesse mudelisse. Töö praktilise väljundina valmib interaktiivne kasutajaliidese prototüüp Figma keskkonnas, mis on loodud TalTech AI veebilehe laiendusena.

1.1 Probleem

Eesti teadustöö toetamiseks ning info säilitamiseks on olemas erinevaid andmeallikaid, põhilisteks neist on teadusaruanded, ETIS ning uurimisgruppide ja ülikoolide veebilehed. Info nimetatud allikates on kvaliteetne, kuid killustunud ning kasutajale ebamugav leida. Puudub terviklik ülevaade teadlaste, projektide, uurimisgruppide ja kompetentside vahelistest seostest. Iga allikas võimaldab küll otsida üksikuid objekte, kuid ei toeta

suhete- ja võrgustikupõhist analüüsi. Killustatuse tõttu on teadlastel ja ettevõtetel raske tuvastada koostööpartnereid ning juhtkond ei saa teha piisavalt teadust puudutavaid andmepõhiseid otsuseid valdkonna kattuvuse või võimalike tühimike kohta.

Tervikliku ja suhetepõhise kaardistussüsteemi olemasolu lihtsustaks koostöövõimaluste tuvastamist, muudaks TalTechi teadusmaastiku kompetentsid nähtavamaks ülikoolisiseselt, toetaks juhtkonna strateegilisi otsuseid ning aitaks tuvastada valdkondlikke tühimikke.

1.2 Eesmärk

Käesoleva töö eesmärk on analüüsida ja disainida suhetepõhine AI-kompetentside kaardistussüsteem TalTechi näitel. Süsteem modelleerib teadlaste, projektide, uurimisgruppide ja märksõnade vahelised seosed graafipõhise andmemudelina, mis võimaldab analüüsida koostöövõrgustikke ja kompetentside ajalist dünaamikat.

Töö raames analüüsitakse võimalikke andmeallikaid, sealhulgas Eesti Teadusinfosüsteemi ehk ETIS-t ja muid avalikke allikaid ning nende integreerimise viise. Samuti käsitletakse rollipõhist haldus- ja andmeuunduse mudelit, võrreldes automaatset, hübriid- ja manuaalset lähenemist. Lisaks antakse ülevaade võrgustikuanalüüsi võimalustest, nagu tiheduse mõõtmine, tsentraalsuse analüüs ja koostööklastrite tuvastamine.

Eesmärkide saavutamiseks viiakse läbi intervjuud sidusrühmadega ning teostatakse olemasolevate lahenduste võrdlusanalüüs, mis annavad aluse nii andmemudeli kui ka kasutajaliidese disainiotsustele.

Töö praktilise väljundina valmib kontseptuaalne UI/UX prototüüp Figma keskkonnas, mis toetab otsingut, graafivaadet ja analüütilist ülevaadet. Prototüüp on loodud ühilduma TalTechi AI veebilehega ning seda saab kasutada aluseks süsteemi edasisele arendamisele.

1.3 Töö struktuur

Käesolev lõputöö koosneb kuuest peatükist. Teine peatükk annab ülevaate töö teoreetilistest alustest: olemasolevate andmeallikate analüüsist, CRM-süsteemide võrdlusanalüüsist, graafianalüüsi meetoditest ning andmete kogumise viisidest koos juriidiliste aspektidega. Kolmas peatükk kirjeldab kasutatud uurimismeetodeid, tööriistu ning sidusrühmadega läbi viidud intervjuusid. Neljas peatükk annab ülevaate andmeallikate analüüsi tulemustest, CRM-süsteemi põhimõtete rakendamisest, intervjuude järeldest ning selle põhjal loodud kasutajavajadustest. Lisaks on käsitletud masinõppe rakendamise võimalusi. Viies peatükk on töö süntees, mis hõlmab kasutajarollide ja õiguste defineerimist, süsteemi struktuuri ning andmemudelit ja loogilist andmemudelit, otsinguloogikat koos sobivusprotsendi mudeliga, kasutajaliidese prototüübi kirjeldust ning projekti edasiste sammude kirjeldust.

2 Taustülevaade ja probleemianalüüs

Järgnev peatükk jaguneb kolmeks osaks. Esimeses osas antakse ülevaade olemasolevatest andmeallikatest ning millised on käesoleva töö jaoks kõige sobilikumad. Teises osas kirjeldatakse CRM-süsteemi, võrreldakse seda loodava AI-kompetentside kaardistussüsteemiga ning analüüsitakse, mida on võimalik CRM-süsteemidelt üle võtta. Kolmas osa keskendub graafianalüüsile ja selle kasutusvõimalustele.

2.1 Olemasolevad andmeallikad

Järgnevas peatükis antakse ülevaade Eestis leiduvatest võimalikest andmeallikatest. Analüüsitakse potentsiaalseid variante ning nende sobivust töö eesmärgi täitmiseks. Olemasolevate allikate puuduste põhjal selgitatakse, miks on vaja uut terviklikku lahendust.

2.1.1 ETIS

Eesti Teadusinfosüsteem (ETIS) on platvorm, mis koondab teavet teadus- ja arendusasutuste, teadlaste, teadusprojektide ning teadustegevuse väljundite kohta. Lisaks toimib ETIS taotluste esitamise, menetlemise ja kinnitamise kanalina [1].

ETIS-e veebilehel on võimalik otsida infot isikute, asutuste, projektide ja teadustegevuse kohta. Teadustegevus jaguneb omakorda alamotsingumootoriteks, kus on võimalik leida infot publikatsioonide, juhendamiste, tööstusomandite, kollektsoonide ja kogude, teadusaparatuuride, teadustaristute, toodete ja teenuste, klassifikaatorite ja eetika kooskõlastuste kohta.

Isikute otsing annab vastuseks otsisõnaga seotud inimesed. Tuleb teha valik, kas otsisõna kohaldatakse isiku ees- ja perenimele, CV-väljadele või seotud asutusele. Asutuse abil inimesi otsides on võimalik määrata veel, kas töösuhe on kehtiv või lõppenud. Lisaks on võimalik filtreerida isikuid üle ETIS ja CERCS valdkondade, mille jaoks on rippmenüüs

valikud ette antud. Inimese kohta on võimalik lugeda tema kontaktandmete, tegutsemisvaldkonna, teenistuskäigu, teaduskraadide, haridustee, tunnustuste, õppetöö, teadusalase tegevuse, projektide, publikatsioonide ning seotud väitekirjade kohta.

Oluliseks piiranguks on, et pole võimalik otsisõnaga otsing üle isiku teadustegevusega seotud kirjade. Saab otsida eraldi igast teadustegevuse alamotsingumootorist, kuid nende üleselt mitte. Seega kui inimese CV-s pole kirjas, et ta tegeleb teatud valdkonnaga, aga ta on tööstusomandi autor antud teemal, siis teda otsingu vastetes pole. Näiteks otsides projektide alt märksõnaga “tehisintellekt” on üheks vasteks projekt “AI andmete rekonstrueerimiseks kõrge energia füüsika eksperimentides”, mille vastutav täitja on Laurits Tani. Kui aga isikute alt sama otsisõnaga otsida, siis teda vastetes pole. See näitab, et isikute otsingu põhjal jäävad sageli leidmata just need interdistsiplinaarsed koostööpartnerid, kelle AI-alased kompetentsid avalduvad muudes valdkondades, nagu eelnevas näites kõrge energia füüsikas.

Samuti ei ole AI alamvaldkonnad (näiteks masinõpe, arvutinägemine) ETIS-es ega CERCS-is eraldi esindatud. Seega pole võimalik isegi valdkondade järgi filtreerides saada täpset pilti AI-kompetentsidest. Märksõnapõhine otsing annaks lisandväärtust, võimaldades teadlasi ja nende töid siduda konkreetsete teemade ja tehnoloogiatega, mis jäävad praeguste üldiste valdkonnakategooriate vahele.

Kuigi ETIS kasutab hägust otsingut, mis aitab kirjavigade korral, ei lahenda see sünonüümide probleemi [2]. Kui otsida isikuid otsisõnaga “tehisaru”, tuleb vastuseks 25 kirjet, aga otsides sõnaga “tehisintellekt” tuleb 158 kirjet. Boole'i otsing pakub küll lahendust vastete laiendamiseks, kuid eeldab kasutajalt efektiivseks otsinguks erinevate sünonüümide defineerimist, mis võib kujuneda aeganõudvaks ning ei pruugi siiski anda kõiki soovitud vasteid [3]. Seetõttu ei sobi see meetod mugavaks ning kiireks kompetentside kaardistamiseks.

Teiste kategooriate alamotsingud töötavad sarnaselt. On võimalik erinevate kirjeldavate väljade abil filtreerida (aasta, pealkiri, väljaanne, asutus, valdkonnad jne). Ka nendes otsingutes esinevad sarnased piirangud. Näiteks publikatsioonide puhul toimub otsisõna otsing üle pealkirjade ja kui pealkirjas pole otsitavat sõna, jääb see publikatsioon tulemustest välja, kuigi sisu poolest võiks see siiski oluline olla.

Kirjeldatud infootsimise korralduse nõrkuseks on laialivalgusus. Iga kategooria (isikud, projektid jne) otsimiseks on vaja navigeerida eraldi vahelehele ja ühtset terviklikku pilti ei teki. Lisaks ei kuvata ETIS-es magistri- ega doktoritöid ning puudub info uurimisgruppide kohta. See piirab eriti AI-kompetentside tervikpildi saamist teadlaste, projektide ja uurimisteemade vahelistest seostest. ETIS toimib hästi spetsiifiliste vastete leidmiseks, kuid arusaama valdkonnas või asutuses toimuvast on keeruline saada.

2.1.2 Tallinna Tehnikaülikooli veebileht

Tallinna Tehnikaülikooli veebilehelt leiab infot erinevate teaduskondade uurimisgruppide kohta instituutide lõikes. Kui ETIS sisaldab põhjalikku infot teadlaste, projektide ja publikatsioonide kohta, siis uurimisgruppe ja nendesse kuulumist seal kirjas pole. Selle info jaoks on TalTechi enda veebileht sobivam allikas. Samuti on seal põhjalikum kontaktinfo koostöök. Kui ETIS-es on võimalik leida kindla isiku kontaktandmeid, siis Tallinna Tehnikaülikooli veebilehelt on võimalik saada infot uurimisgrupi kui terviku ning nende pakutava väärtuse kohta. Lisaks leiab sealt viiteid uurimisgrupi veebilehtedele ning tehtud projektidele [4].

2.1.3 Tallinna Tehnikaülikooli raamatukogu digikogu

Tallinna Tehnikaülikooli raamatukogu digikogust leiab ülikooli doktori-, magistri-, bakalaureuse- ja rakenduskõrghariduse diplomitöid; õpikute ja õppematerjalide e-versioone ning ülikooli ajalooga seotud publikatsioone [5].

Digikogus on võimalik teostada põhjalikku otsingut, leitavad doktori- ja magistritööd on märksõnastatud ning otsing töötab efektiivselt. Puudusteks on, et publikatsioonidel pole märksõnu küljes ning pole võimalik leida rahastusega projekte, erinevalt ETIS-est. Lisaks puudub info uurimisgruppide kohta, inimesi pole võimalik otsida ning magistri- ja doktoritöödel puudvad kokkuvõtted. Samuti on publikatsioone digikogus ETIS-ega võrreldes veidi vähem. Kui digikogus valida kõik kategooriad ning perioodiks valida 2020-2025, siis leidub 10288 vastet, ETIS-est samal perioodil publikatsioonide alt otsides ning ülikooliks „Tallinna Tehnikaülikool“ valides, leidub 10480 vastet.

Tallinna Tehnikaülikooli raamatukogu digikogu on heaks andmeallikaks doktori- ning magistritööde otsimisel. Märksõnade olemasolu on samuti kasulik, sest kokkuvõtteid digikogus ei ole ning töö sisu saab lugeda vaid PDF-failist.

2.1.4 Muud allikad

Lisaks eelnevalt välja toodud allikatele tutvuti ka teadusagentuuri grandiprojektide, Google Scholari ning e-varamuga. Järgnevalt põhjendatakse, miks antud allikad välja jäid.

Eesti Teadusagentuur rahastab erinevaid uurimustöid ning nende tulemustest tehakse iga aasta grandiprojektide kogumik "Teadusrikas Eesti". Kogumik annab ülevaate Eestis tehtavatest teadusuuringutest, nende eesmärkidest ja tulemuste olulisusest [6]. Otsimisfunktsiooni kui sellist kogumikusiseselt pole, aga ülevaade nimetatud aasta kõige olulisematest teadustöödest on põhjalik. Neid teadustöid on leitavad ka ETIS-est, mis sisaldab lisaks ka isikute, asutuste ja muud seonduvat infot. Seega ei ole antud kogumiku andmeallikana kasutamine oluline.

Google Scholariga on võimalik teostada teaduskirjanduse laiaulatuslikku otsimist [7]. Olemas on laiendatud otsing, kus on võimalik määrata otsitavad sõnad ja fraasid. On võimalik määrata, kas sõnad peaksid esinema ükskõik millises kohas artiklis või ainult pealkirjas. Lisaks on võimalik määrata autor, kus väljandes see publikatsioon avaldati, millises valdkonnas ja mis ajavahemikul. Google Scholar on hea tööriist teadustööga tegelevale inimesele valdkonnaalaste allikate leidmiseks, kuid Tallinna Tehnikaülikooli publikatsioonidest ülevaate saamine on raskendatud. Näiteks kui otsida fraasiga „Tallinn University of Technology“ ning jätta välja viited, siis on üheks vastetest „Legal design and access to justice: a case study on child support payments in Chile“. Selle autorid on Tšiili Adolfo Ibáñez ülikoolist [8]. Tallinna Tehnikaülikooliga on see seotud vaid sel määral, et publikatsioon on avaldatud teadusajakirjas TalTech Journal of European Studies. Et Google Scholarit allikana kasutada, peaks seal olema võimalik filtreerida institutsioonide põhjal. Hetkel seda võimalust pole ning leitud vasteid tuleks kohalikest andmeallikatest üle kontrollida, sest potentsiaalselt võib vastete hulka sattuda ka töid, mis mainivad Tallinna Tehnikaülikooli, kuid mille autorid pole ise ülikooliga seotud.

E-kataloog ESTER on veebikataloog. ESTER kajastab ELNET Konsortsiumiga liitunud raamatukogude, arhiivide ja muuseumide kogusid. E-kataloog on pidevalt täienev [9]. ESTER kajastab hästi raamatukogudes sisalduvaid teoseid ja trükiseid, kuid autori organisatsiooni järgi pole võimalik filtreerimist teha. ESTER on sobiv allikas teoste asukoha leidmiseks, kuid hetkel pole võimalik filtreerida tulemusi nii, et oleks kuvatud vaid publikatsioonid, mis on avaldatud Tallinna Tehnikaülikooli alt.

E-varamu portaal on keskkond, mis koondab Eesti raamatukogude, muuseumide ja arhiivide infosüsteeme, pakkudes keskset juurdepääsu erinevates andmekogudes sisalduvale teabele. E-varamu peamine eelis on võime otsida üheaegselt nii ETIS-e, ESTERi kui ka ülikoolide digikogude andmetest, mis teoreetiliselt võiks pakkuda kõige laiemat vaadet Eesti teadusmaastikule [10]. Käesoleva töö raames jäi e-varamu andmeallikana siiski teisejärguliseks sarnastel põhjustel nagu ESTER. Kuigi portaal võimaldab laiaulatuslikku otsingut, on tulemuste filtreerimine konkreetselt Tallinna Tehnikaülikooli autorite ja üksuste lõikes ebatäpne ning tekitab suurt andmete dubleerimist teiste, spetsiifilisemate allikatega (nt ETIS ja Digikogu). AI-kompetentside kaardistamiseks vajalikku struktureeritud informatsiooni on otstarbekam pärida otse algallikatest.

2.2 Võrdlusanalüüs

Loodav süsteem kuulub suhetepõhiste organisatsiooniliste infosüsteemide klassi. Sarnaselt CRM-süsteemidele modelleerib see objekte ja nendevahelisi seoseid ajas, võimaldades analüütilist ülevaadet koostöömustrist. Järgnevad alapeatükid annavad ülevaate CRM-i olemusest, suhete modelleerimisest, ajaloo ja õiguste haldamisest. Lisaks analüüsitakse loodava süsteemi kattuvusi CRM-süsteemiga.

2.2.1 Mis on CRM?

CRM ehk kliendisuhete juhtimine on terviklik juhtimis- ja ärifilosoofia, mis ühendab turunduse, tehnoloogia ja strateegilise juhtimise, eesmärgiga luua ja arendada pikaajalisi kliendisuheteid. CRM ei piirdu üksnes tehnoloogiliste lahendustega, vaid kujutab endast laiemat lähenemist, mis keskendub kliendi väärtuse suurendamisele ning ettevõtte ja kliendi vahelise vastastikku kasuliku suhte loomisele [11].

Varasemates käsitlustes rõhutatakse CRM-i rolli kliendiandmete kogumisel ja kasutamisel, et parandada otsustusprotsesse, suurendada kliendirahulolu ning toetada ettevõtte tulemuslikkust. CRM võimaldab ettevõtetel paremini mõista klientide vajadusi ja käitumist, mis omakorda aitab kujundada personaalsemaid ja tõhusamaid turundus- ning teenindusstrateegiaid [12], [13].

Adrian Payne käsitlese kohaselt on CRM eelkõige terviklik strateegiline lähenemine, mille eesmärk on juhtida kliendisuhteid viisil, mis suurendab ettevõtte väärtust ja kasumlikkust. CRM ei ole pelgalt IT-süsteem, vaid organisatsiooniülene protsess, mis hõlmab strateegia kujundamist, väärtuse loomist, suhtlust ja kliendiinfo juhtimist [14].

CRM-i definitsioonil on ka praktiline ja tehnoloogiline perspektiiv. Rõhutatakse, et CRM-süsteemide, eriti IT-põhiste lahenduste kasutuselevõtt võimaldab ettevõtetel integreerida erinevaid kliendiga seotud protsesse ning parandada müügi- ja teenindustegevuse efektiivsust. Samas sõltub CRM-i edukus mitte ainult tehnoloogiast, vaid ka organisatsioonilistest teguritest, nagu töötajate oskused, ettevõtte valmisolek ja juhtimispraktikad [15].

See viitab sellele, et CRM-i tulemuslikkus on tugevalt seotud nii tehniliste kui ka inimfaktoritega. Edukas CRM-i rakendamine eeldab organisatsiooniülest koostööd, protsesside kohandamist ning kliendikeskse mõtteviisi omaksvõttu. Seega võib CRM-i käsitleda integreeritud süsteemina, mis ühendab tehnoloogia, inimesed ja äriprotsessid.

Kokkuvõttes võib CRM-i defineerida kui strateegilist ja tehnoloogilist lähenemist, mille eesmärk on luua väärtust nii kliendile kui ka ettevõttele, arendades pikaajalisi ja kasumlikke kliendisuhteid ning toetades ettevõtte jätkusuutlikku arengut [16].

2.2.2 Kuidas CRM-id modelleerivad suhteid?

CRM-süsteemid modelleerivad suhteid peamiselt andmebaasipõhiste struktuuride kaudu, kus keskseteks objektideks on kliendid (nt kontaktid ja organisatsioonid) ning nendevahelised ja ettevõttega seotud seosed. Suhete modelleerimine tugineb enamasti relatsioonilisele andmemudelile, milles kasutatakse olemeid, atribuute ja seoseid [17].

Esiteks luuakse CRM-is põhiobjektid, nagu *kontakt* (üksikisik) ja *konto* (ettevõtte). Nende objektide vahel määratletakse seosed, näiteks „töötab ettevõttes“ või „on seotud projektiga“. Sellised seosed võimaldavad süsteemil kirjeldada nii otseseid kui ka kaudseid ärisuhteid [17].

Teiseks kasutatakse suhete kirjeldamiseks erinevaid kardinaalsusi: üks-ühele (1:1), üks-mitmele (1:N) ja mitme-mitmele (M:N). Näiteks võib ühel ettevõttel olla mitu kontakti (1:N), samas kui üks kontakt võib olla seotud mitme projektiga (M:N). Viimast tüüpi seoste puhul kasutatakse sageli vahetabeleid (*junction tables*), mis võimaldavad seoseid täpsemalt kirjeldada [15].

Kolmandaks sisaldavad CRM-id ajadimensiooni, mis võimaldab jälgida suhete arengut ajas. See hõlmab tegevuslogisid (nt e-kirjad, kõned, kohtumised), müügivõimalusi ja klienditeekonna etappe [14]. Selline ajaline modelleerimine on oluline, et mõista kliendi elutsüklit ja optimeerida suhtlusstrateegiaid.

Lisaks rikastatakse suhteid metaandmetega, nagu rollid (nt otsustaja, mõjutaja), prioriteetid ja suhtluse ajalugu. See võimaldab CRM-il mitte ainult salvestada, vaid ka analüüsida suhete kvaliteeti ja tugevust [14], [15].

Kaasaegsed CRM-süsteemid kasutavad sageli ka graafipõhiseid lähenemisi või laiendusi, kus kliendid ja nendevahelised seosed esitatakse sõlmede ja servadena. Selline lähenemine võimaldab paremini analüüsida keerukaid võrgustikke, näiteks soovitusseid või organisatsioonisisene mõju [18].

Kokkuvõttes modelleerivad CRM-id suhteid mitmetasandiliste andmestruktuuride abil, kombineerides relatsioonilisi seoseid, ajalugu ja kontekstuaalseid atribuute, et toetada nii operatiivset kliendihaldust kui ka strateegilist otsustamist.

2.2.3 Kuidas nad haldavad ajalugu?

Ajadimensioon vajab täpsemat käsitlust, kuna suhete ajaline jälgimine on CRM-süsteemides keeruline. Süsteem peab aru saama, milline oli tegelikkus antud ajahetkel ja mida süsteem sel ajahetkel tegelikult teadis [19]. Muudatuste haldamiseks ei tohiks sisestada kogu informatsiooni faktitabelisse, samuti ei ole praktiline käsitleda kõiki

dimensioone ajas muutuvatena. Kuna enamik dimensioone muutub ajas harva, saab andmemudeli struktuuri säilitada vaid väikeste kohandustega. Kimball ja Ross tähistavad selliseid ajas aeglaselt muutuvaid dimensioone mõistega *slowly changing dimensions*, mida tähistatakse lühendiga SCD [20].

Iga dimensioonitabeli atribuudi jaoks tuleb määrata strateegia muutuste käsitlemiseks. Strateegia peab kirjeldama, kuidas me reageerime muutustele dimensioonimaailmas, kui atribuudi väärtus muutub operatsioonimaailmas. On kolm põhilist tehnikat, et atribuutide muutumisega toime tulla [20]:

1. Väärtuse ülekirjutamine

Vana väärtus kirjutatakse dimensioonireal uue väärtusega üle. Seega atribuudil on alati kõige hilisemalt määratud väärtus. Faktitabel jääb muutumatuks. See on lihtsaim lähenemisviis atribuutide muudatuste haldamiseks. Probleemiks on, et kaob kogu atribuutide muudatuste ajalugu. Seega see variant pole sobilik, kui muudatuste ajalugu peab puutumatult säilima. Näiteks kui teadlane vahetab töökohta, siis kirjutatakse vana töökoht uuega üle ning varasem kuuluvus kaob.

2. Dimensiooni rea lisamine

See on peamine tehnika, mis võimaldab aeglaselt muutuvate dimensioonide atribuutide jälgimise. Atribuudi muutumisel lisatakse dimensioonitabelisse uus rida ja vana säilib. Ühe objekti kohta tekib mitu rida, milles igaüks kirjeldab mingit ajaperioodi. Ridadele saab ka lisada kuupäevad, mis võimaldab täpselt määrata, millal atribuut kehtis. Selle variandi eeliseks on ajaloo täielik ja täpne säilimine. Küll aga võib probleemiks osutuda iga muutusega kasvav dimensioonitabel.

3. Dimensiooniveeru lisamine

Selle tehnika puhul lisatakse dimensioonitabelisse uus veerg eelmise väärtuse salvestamiseks. Seega on tabelis vana ja uus väärtus kõrvuti ja see võimaldab andmeid analüüsida nii vana kui ka uue väärtuse alusel. Seda nimetatakse ka alternatiivse reaalsuse vaateks. See variant sobib, kui muutusi vähe. Paljude muudatuste puhul osutub see ebapraktiliseks ning levinud kasutusviisiks on 2. tehnika.

Keerulisemates süsteemides eristatakse kahte ajatelge. Üheks ajateljeks on see, millal muutus tegelikkuses toimus. Teiseks ajateljeks on süsteemi sisestamise aeg. Seda nimetatakse bi-temporaalseks ajalooks [19]. Käesolevas töös me keerulisemaid süsteeme ei käsitle.

2.2.4 Kuidas nad lahendavad õiguseid?

NIST (National Institute of Standards and Technology) ehk eesti keeles USA Riiklik Standardi- ja Tehnikainstituut näeb ette, et organisatsioonidel peaks olema rollipõhine juurdepääsu kontroll. NIST SP 800-53 (Rev. 5) standard rõhutab, et õiguste haldamine rollipõhise skeemi alusel toetab vähemate õiguste printsiipi, mis aitab piirata kasutajate ligipääsu vaid hädavajaliku infoni ja vähendada seeläbi süsteemi komprometterimisohtu [21].

Praktikas tähendab see, et CRM-süsteemides seotakse konkreetsed õigused (nt info vaatamine, muutmine või kustutamine) mitte üksikute kasutajate, vaid rollidega. See muudab süsteemi haldamise märkimisväärselt lihtsamaks, kuna administraatoril piisab kasutajale sobiva rolli määramisest, et tagada talle tööülesanneteks vajalik, kuid samas turvaliselt piiratud ligipääs süsteemi funktsioonidele. Rollipõhise juurdepääsu süsteemi puhul määratakse kasutajatele rollid, rollidele õigused ning sessioonide ajal aktiveerib kasutaja osa talle määratud rollidest. Rollid võivad moodustada hierarhia, kus kõrgema taseme roll pärib madalama taseme rolli õigused. Selline struktuur järgib ka organisatsioonisisest võimustruktuuri ning lihtsustab õiguste haldamist suurtes süsteemides, vähendades iga individuaalse kasutaja õiguste ülevaatamise vajadust [22].

2.3 Andmete kogumise viisid ja juriidilised piirangud

Tehnilise lahenduse kontekstis on kaks peamist võimalikku andmete kogumise viisi: API-d ja veebikraapimine. Eelanalüüsi raames hinnati mõlema meetodi teostatavust ja piiranguid.

API-d pakuvad struktureeritud ja ametlikult lubatud juurdepääsu andmetele [23] ning on tehnilise lahenduse seisukohalt eelistatum valik mitmel põhjusel. Esiteks on API kaudu saadavad andmed juba struktureeritud formaadis (nt JSON või XML), mis lihtsustab

nende töötlemist ja integreerimist otsingusüsteemi. Juriidiliselt on API kasutamine selge, kui teenusepakkuja on API avalikustanud, tähendab see üldjuhul, et andmete kasutamine on lubatud ja kooskõlas kasutustingimustega. Lisaks on API-põhine lahendus pikemas perspektiivis stabiilsem ja hooldatavam: veebilehe visuaalse kujunduse muutumisel läheb kraapimisloogika tõenäoliselt katki, samas kui API struktuur püsib üldjuhul muutumatuna. Samuti võimaldavad paljud API-d päringuid filtreerida märksõnade järgi, mis vähendab töötlemist vajavate andmete hulka ja muudab süsteemi efektiivsemaks.

Veebikraapimine on alternatiiv allikate puhul, millel puudub avalik API, kuid sellega kaasnevad olulised piirangud. Paljud veebilehed keelavad oma kasutustingimustes automaatse andmekogumise, mistõttu tuleb iga allika puhul tingimused eelnevalt hoolikalt üle vaadata. Isikuandmete töötlemist reguleerib Euroopa Liidus GDPR, mis seab ranged nõuded sellele, kuidas ja mis eesmärgil avalikustatud isikuandmeid võib kasutada. Isegi kui andmed on veebis vabalt nähtavad, ei tähenda see automaatselt, et nende masstöötlamine on lubatud. Lisaks on veebikraapimine tehniliselt hapram lahendus: lehtede struktuur võib muutuda etteteatamata, mis nõuab pidevat hooldust. Seetõttu on soovitatav veebikraapimist kasutada vaid viimase võimalusena, kui API või muu ametlik andmejuurdepääs puudub. Eelanalüüsi tulemusena saab arendusmeeskond teha informeeritud otsuse, millised allikad ja meetodid on realistlikult kasutatavad [24].

2.4 Graafianalüüs

Graafianalüüs on meetodite kogum, mida kasutatakse keerukate süsteemide ja nende elementide vaheliste seoste uurimiseks [25]. Sellistes süsteemides esitatakse objektid sõlmedena ning nendevahelised suhted servadena, moodustades võrgustiku. Graafipõhine lähenemine võimaldab visualiseerida ja analüüsida mitte ainult üksikuid objekte, vaid ka nendevahelisi seoseid ja struktuure tervikuna.

Käesoleva töö kontekstis on graafipõhine lähenemine valitud otsese vastusena peatükis 2.1 kirjeldatud andmeallikate piirangule. Olemasolevad süsteemid nagu ETIS võimaldavad otsida üksikuid objekte, kuid ei toeta suhete- ja võrgustikupõhist analüüsi. Iga kategooria otsimiseks tuleb navigeerida eraldi vahelehele ning terviklikku pilti seostest ei teki. Graafimudel lahendab selle probleemi, kuna see võimaldab modelleerida

teadlaste, projektide, uurimisgruppide ja märksõnade vahelised mitmetasandilised seosed ning analüüsida neid tervikuna, mitte eraldiseisvate kirjetena [18]. Järgnevalt kirjeldatakse graafianalüüsi meetodeid ning nende rakenduslikku väärtust AI-kompetentside kaardistussüsteemi kontekstis.

2.4.1 Võrgustiku tihedus

Võrgustiku tihedus kirjeldab, kui suur osa võimalikest seostest võrgustikus tegelikult eksisteerib [26]. Seda kasutatakse, et hinnata, kui tihedalt on süsteemi elemendid omavahel seotud.

TalTechi AI-valdkonna kontekstis võimaldab võrgustiku tiheduse analüüs hinnata, kui aktiivselt teadlased ja uurimisgrupid omavahel koostööd teevad. Kõrge tihedus viitab aktiivsele koostöökultuuriga keskkonnale, madal tihedus aga sellele, et teadlased töötavad pigem eraldiseisvalt ning koostöövõimalused on kasutamata [27]. Juhtkonnale on see oluline strateegiline näitaja. Kui tihedus on madal, võib see viidata vajadusele koostööd aktiivsemalt soodustada näiteks ühisprojektide või interdistsiplinaarsete uurimisgruppide kaudu.

2.4.2 Koostöö klastrid

Koostöö klastrid on võrgustiku alamhulgad, kus elemendid on omavahel tihedamalt seotud kui ülejäänud võrgustikuga. Neid kasutatakse, et tuvastada loomulikke gruppe või kogukondi süsteemis [28].

Käesolevas süsteemis võivad klastrid paljastada seoseid, mis ei ole ametliku organisatsioonistruktuuri kaudu nähtavad. Näiteks võib selguda, et masinõppe ja arvutinägemise uurimisgrupid teevad omavahel tihedalt koostööd ning moodustavad ühtse kogukonna, samas kui loomuliku keele töötlusega tegelevad teadlased on ülejäänud võrgustikust eraldiseisvamad. Selline info annab juhtkonnale sisendi, millised koostöövormid on orgaaniliselt tekkinud ning kus võiks olla potentsiaali uute seoste loomiseks.

2.4.3 Keskusmõõdikud

Keskusmõõdikud kirjeldavad, kui oluline või mõjukas on konkreetne element võrgustikus. Levinumad mõõdikud on näiteks *degree centrality*, mis mõõdab otseste seoste arvu, *betweenness centrality*, mis näitab kui sageli läbivad lühimad teed selle sõlme, ning *closeness centrality*, mis kirjeldab, kui lähedal on sõlm kõigile teistele võrgustikus [29].

Loodavas süsteemis on kõrge tsentraalsusega teadlane selline, kellel on seosed paljude erinevate objektidega: projektide, uurimisgruppide ja märksõnadega. Sellised teadlased on AIRE-le eriti väärtuslikud koostööpartnerid, kuna nad on juba tõestanud võimet töötada interdistsiplinaarsetes kontekstides ning nende kompetentsid ulatuvad üle mitme valdkonna. *Betweenness centrality* mõttes on kõrge väärtusega teadlane sild kahe muidu eraldiseisva uurimissuuna vahel, mille tuvastamine võib olla strateegiliselt oluline nii juhtkonnale kui ka uute koostöövõimaluste otsimisel [29].

2.4.4 Isolatsioonid

Võrgustiku isolatsioonid avalduvad mitmel kujul, millest igaüks annab erineva signaali süsteemi struktuuri kohta.

Isolated nodes ehk isoleeritud sõlmed on elemendid, millel puuduvad seosed teiste võrgustiku elementidega [25]. Antud töö kontekstis võivad isoleeritud sõlmed tähendada teadlasi või projekte, mis ei ole seotud teiste AI-valdkonna tegevustega. Oluline on eristada kahte võimalikku põhjust: kas andmed on puudulikud ning tegelikud seosed on lihtsalt süsteemi kandmata, või on tegemist tegeliku eraldatusega. Mõlemal juhul on tegemist administraatorile suunatud signaaliga, mis vajab tähelepanu.

Structural holes viitavad olukordadele, kus võrgustikus puuduvad ühendused kahe muidu eraldiseisva grupi vahel, kuigi temaatiline lähedus oleks olemas [30]. Süsteemi kontekstis võib selline strukturealne auk esineda näiteks kahe AI alamvaldkonna vahel, mille uurimisgrupid töötavad paralleelselt, kuid pole omavahel seni koostööd teinud. Selliste kohtade tuvastamine on strateegiliselt väärtuslik, kuna uue seose loomine just sinna võib tuua mõlemale osapoolle lisaväärtust.

Disconnected components on võrgustiku osad, mis ei ole omavahel üldse ühendatud ning moodustavad eraldiseisvaid alamvõrgustikke [26]. Käesolevas süsteemis võivad sellised komponendid viidata uurimisvaldkondadele või gruppidele, mis toimivad iseseisvalt ega ole seotud ülejäänud AI-ökosüsteemiga TalTechis.

Network sparsity ehk võrgustiku hõredus viitab üldisemalt puuduvatele või alaarenenud ühendustele võrgustikus [25]. Erinevalt *structural holes* mõistest, mis tähistab konkreetset lünka kahe tuvastatud grupi vahel, viitab *network sparsity* laiemalt katmata aladele, kus vajalikud kompetentsid või koostöövormid puuduvad. Juhtkonnale on see oluline sisend strateegiliste otsuste tegemisel, näiteks millistes AI alamvaldkondades on TalTechis kompetentsid nõrgad ning kus oleks vaja uusi inimesi või projekte juurde.

Graafianalüüsi meetodite rakendamine võimaldab muuta hajutatud teadusandmed struktureeritud teadmisteks, pakkudes ülevaadet nii olemasolevatest seostest kui ka puudujääkidest. Kirjeldatud meetodid moodustavad teoreetilise aluse, millele tugineb käesoleva töö andmemudel ning kasutajaliidese graafivaade, mida on täpsemalt kirjeldatud peatükkides 5.2 ja 5.5.2.

3 Metoodika

Selles peatükis kirjeldatakse uurimismeetodeid ja tööetappe, lisaks analüüsitakse peatükis 2.1 välja valitud andmeallikaid ning käsitletakse nendest andmete kogumise juriidilisi piiranguid. Antakse ülevaade läbiviidud intervjuudest ning nende põhjal pannakse paika kasutajavajadused.

3.1 Uurimismeetodid ja tööetapid

Käesolev töö ühendab kolm peamist uurimismeetodit: kvalitatiivse analüüsi, võrdleva analüüsi ning kasutajaliidese prototüüpimise. Kuna töö teema pärineb Protsessorist ning tegemist on kliendi tellimusel valminud tööga, kujunes uurimisprotsess interaktiivseks. Iga etapp tugines eelmise tulemustele ning täpsustas järgmise suunda.

Töö koostamine algas kvalitatiivse analüüsiga, mille peamiseks andmekogumise meetodiks oli poolstruktureeritud ning valideerimisintervjuud sidusrühmadega, mida on täpsemalt kirjeldatud peatükis 3.3. Esimesest intervjuust saadud sisendi põhjal hakati kujundama laienduse võimalikku struktuuri ning töösuunda ja paralleelselt alustati prototüübi esmase versiooni loomisega Figma, et järgnevatel intervjuudel seda esitleda ning tagasisidet saada.

Võrdlevat analüüsi rakendati kahel tasandil. Esiteks analüüsiti olemasolevaid andmeallikaid, ETIS-t, TalTechi veebilehte ning digikogu, et kaardistada nende tugevused, piirangud ja kattuvused, mis on kirjeldatud peatükis 2.1. Teiseks viidi läbi CRM-süsteemide võrdlusanalüüs ning graafianalüüsi meetodite uuring peatükkides 2.2 ja 2.3, et tuvastada, milliseid suhete modelleerimise, ajaloo ja õiguste haldamise ning võrgustikuanalüüsi põhimõtteid on võimalik loodavasse süsteemi üle kanda. Intervjuude käigus kogutud mõtteid analüüsiti samuti võrdleva meetodi abil, sidudes need ajakohaste teoreetiliste käsitlustega.

Kasutajaliidese prototüüpimine oli töö praktiline väljund ning kulges läbivalt kogu uurimisprotsessi vältel, mitte üksnes lõppfaasis. Samuti tegeleti ka pidevalt paralleelselt andmemudeli loomisega. Prototüüp loodi Figma Design keskkonnas, andmemudel Enterprise Architectis. Kuna eesmärgiks oli olemasoleva TalTechi tehisintellekti veebilehe laiendamine, kasutati Figma pluginat `html.to.design` [31], mis genereeris olemasoleva veebilehe põhjal visuaalse struktuuri, millele sai hakata uut lisandit juurde looma. Prototüüpi esitleti igal intervjuul ning saadud tagasiside põhjal täiustati seda järjepidevalt, mille kulg on kirjeldatud peatükis 3.3.3.

Töö etapid järgnesid üksteisele järgmiselt. Esmalt viidi läbi andmeallikate, CRM-süsteemide ja graafianalüüs teoreetilise aluse loomiseks. Seejärel toimusid intervjuud sidusrühmadega, mille käigus kaardistati kasutajavajadused ja täpsustati süsteemi skoopi. Intervjuude vahepeal ja järgselt arendati prototüüpi edasi, tuginedes nii intervjuudest saadud tagasisidele kui ka analüüsitulemustele. Töö kulmineerus andmemudeli, otsinguloogika ning kasutajaliidese disaini sünteesiga, mis on esitatud peatükis 5.

3.2 Tööriistad

Käesolev töö on kirjutatud Microsoft Wordis. Kasutajaliidese prototüübi loomiseks on kasutatud Figma toodet Figma Design [32]. Andmemudelite loomiseks on kasutatud Enterprise Architecti [33]. Intervjuude läbiviimiseks kasutati Microsoft Teamsi, intervjuujärgseteks mõttevahetusteks kasutati Microsoft Outlooki. Õigekirja kontrolliks ning sobivusprotsendi genereerimise loogika arendamiseks kasutati tehisintellekti mudeleid Gemini 3 Flash [34] ja Claude Sonnet 4.6 [35].

3.3 Intervjuud

Süsteemi vajaduste kaardistamiseks ja skoobi täpsustamiseks viidi läbi intervjuud potentsiaalsete sidusrühmade ning lahenduse tellijaga. Töö raames toimus kokku kolm põhilist kohtumist: kaks intervjuud TalTech AI esindajatega, üks kohtumine Tehisintellekti- ja robotikakeskuse töötajatega (edaspidi AIRE).

3.3.1 Valimi koostamine

Valim koostati potentsiaalsete kasutajate põhjal. Algselt oli tulevase süsteemi olemasolust huvitatud TalTech AI, kes soovis seda pakkuda ka TalTechi juhtkonnale analüüsi eesmärgil. Lisaks kaasati AIRE, mis ühendab Eesti ülikoole ning teadlasi ettevõtjatega [36]. Teaduspartnerite leidmine toimub AIRE ja TalTechi koostöös valminud Adapteris digiplatvormi ja kliendihaldurite abil [37]. Loodav lahendus oleks Adapterile sobivate vastete leidmisel kasulikuks tööriistaks, automatiseerides teaduspartnerite leidmist. Viimases etapis kaasati ka juhtkonna huve esindavad osapooled, et testida prototüübi sobivust. Sidusrühmad ja intervjuueeritavad on välja toodud tabelis 1.

Tabel 1. Intervjuude sidusrühmad ja nende esindajad.

Sidusrühm	Intervjuueeritavad
Tellija – TalTech AI	Sven Nõmm, Arianna Jater, Julija Mõnnakmäe
AIRE	Evelin Ebruk, Katre Eljas, Jaanus Vahesalu, Dmitri Derevjanko
Juhtkonna esindaja – TalTech AI	Julija Mõnnakmäe, Kristel Kriisa

3.3.2 Intervjuude läbiviimine

Intervjuud viidi läbi eesti keeles videokõnede vahendusel MS Teamsi keskkonnas vahemikus 17. märts – 20. aprill. Kohtumistel osalesid mõlemad autorid. Need kestsid ligikaudu 30 minutit kuni üks tund ning need salvestati osapoolte nõusolekul, et töö autorid saaksid hiljem vestlust üle kuulata ja analüüsida.

Läbi viidi kahte tüüpi intervjuusid. Esimene neist oli poolstruktureeritud formaadis intervjuu [38], mis võimaldas vestlust juhtida struktureeritult varem ette valmistatud küsimustega, samas jättes intervjuueeritavatele vabaduse valida, kui süvitsi soovitakse teemat käsitleda (vt Lisa 2). Teise tüübina viidi läbi valideerimisintervjuud [39], mille eesmärk oli kinnitada, kas valitud kasutajarollide jaoks on pakutud lahendus kasulik.

3.3.3 Intervjuude ülevaade

Esimene intervjuu viidi läbi TalTech AI-ga, kes on projekti tellija. Esindajateks olid Sven Nõmm, kellest sai ka töö kaasjuhendaja, Arianna Jater ja Julija Mõnnakmäe. Selles vestluses selgitas Arianna põhjalikumalt, milline on nende nägemus loodavale lahendusele.

Selles intervjuus sõnastati probleem, milleks oli kasutajasõbraliku allika puudumine, kust saaks kätte kogu informatsiooni inimeste, projektide ja nendega seonduva kohta. Laiema probleemina toodi välja, et puudus on teaduskondade vahelisest tihedamast koostööst. Mainiti, et ülikoolisiselt ei ole inimesed teadlikud, millist teadustööd TalTechis tehakse ning seetõttu pole võimalik koostöö potentsiaali täielikult kasutada.

Omalt poolt olid intervjuueeritavad kokku kogunud võimalikud killustunud infoga andmeallikad, milleks on ETIS, ülikoolide- ja uurimisgruppide veebilehed ja teadusaruanded. Töö suunaks sai AI valdkonna kaardistamine, mida hiljem saaks laiendada ka teistele valdkondadele. Avaldati soovi andmete eksportimisvõimaluseks ning nõustuti ideega kuvada seoseid graafil. Sihtgruppidenähti ülikoolisiseseid kasutajaid, kes saaksid uurida seniseid teadlaste vahelisi koostöid ning neil oleks olemas ülevaade, millised kompetentsid on ülikoolis kaetud. Samuti oleks üheks kasutajarolliks välispartnerid ja lihtsalt huvilised, kes otsivad endale koostööpartnereid või meelepärasel teemal lugemist. Lõpplahendusena sooviti TalTech AI veebilehele ai.taltech.ee integreeritud otsingumootorit, mis uuendaks oma andmeid üks kuni kaks korda kuus. Seejuures kaardistati koos algsed süsteemi kasutajad:

- rektoraat või teadusosakond, kellel peaks olema võimalik näha tulemusi visualiseeritult aga neid ka raportina eksportida,
- teadlane, koostööpartneri otsija, kellel oleks võimalik näha kompetentse, välja antud artikleid ja muid teadustöid.

Need kasutajarollid olid esialgsed ning täpsustusid järgnevate intervjuude käigus. Kasutajate muutmisõiguste osas mainiti, et kuna loodav lahendus koondab andmeid erinevatelt lehtedelt kokku, siis sellist võimalust ei tule, et teadlased või uurimisgrupid enda kohta käivat infot muuta saaks. Selleks tuleb pöörduda algallika poole.

Teine intervjuu toimus Tehisintellekti- ja robotikakeskuse ehk AIRE-ga, kellest sai esimese intervjuu järgselt kaasjuhendaja Sven Nõmme suunamisel otsingufunktsiooni uus potentsiaalne kasutaja. Vestluses osalesid Evelin Ebruk, Katre Eljas, Jaanus Vahesalu ning Dmitri Derevjanko. Neile esitleti esimeste nõuete põhjal loodud prototüüpi, kuhu oli parima tulemuse põhjendamiseks kirjetele lisatud ka sobivusprotsent, mis annaks neile lisandväärtust sobivate teadlaste otsingul.

AIRE peamiseks soovitusel oli sihtrühma/kasutaja kindlaks määramine, et oleks läbimõeldud, kellele sobivusprotsent enim lisandväärtust loob. Samuti tõstatati info dubleerimise ohtu ning pöörati tähelepanu Adapterile, et ei tekiks kahte keskkonda, kust on võimalik partnereid leida. Kui eelnevalt olid kasutajateks määratud teadlased, juhtkond ja AIRE, siis vestluse järgselt jõuti arusaamale, et lahendus peaks olema algselt kooli siseveebis. Samuti tutvustas AIRE põhjalikumalt koostöövõrgustikku Adapter ning kuidas loodav lahendus potentsiaalselt Adapteri tööd lihtsustaks ja klientide ning päringute hulka suurendaks. Lisaks tõsteti esile probleemi, et TalTechi teenused on märksõnastamata, mis raskendab Adapteril teenuste filtreerimist. Valesti märksõnastatud teenused tingivad otsingul liigselt andmemüra. Tehnilise poole pealt pakkus AIRE abi embedding-mudelite ja rerouting-lahenduste rakendamisega seotud küsimustega.

Kolmas intervjuu toimus TalTech AI-st Julija Mõnnakmäe ja Kristel Kriisaga, kes esindasid TalTechi juhtkonna huve. Kuna loodav süsteem on mõeldud ka juhtkonnale, nähti vajadust nende poolseks tagasisideks. Neile esitleti edasi arendatud prototüüpi, mille järgselt peeti meilivestlust koosoleku järelduste ja edasiste soovitude osas. Soovitusel oli, et sihtrühm oleks kitsas, et testida ning hinnata, kas ja kuidas on võimalik lahendust hiljem edasi laiendada. Otsingufunktsiooni puhul juhiti tähelepanu näidismärksõnade valikule, mille kohta uuriti, miks on neid prototüübis kuvatud ainult viis. Samuti anti nõu graafivaadete kohta ning informeeriti, et need peaks jääma ainult juhtkonnale ning AIRE-le nähtavaks. Nende poolse tagasisidena soovitati töötavat süsteemi kasutada päris teadlaste peal, et mõista, kas tulemused on arusaadavad, leitav info kasulik ja kas kasutajad jäävad rahule.

Prototüübi parendamiseks anti teemavälise kasutaja perspektiivi. Näiteks oli esialgses lahenduses menüüpunktiks „Kaardistamine“, mis tundus igati arusaadav, kuid kasutaja,

kes pole teemaga kursis vajaks paremat sõnastamist. Lisaks toodi välja, et ette antud otsingu variantides võiks olla laiem ja praktilisem valik märksõnu. Samuti soovitati prototüübis näited tuua päris võimalike tulemustega, kus oleks teststsenaariumi otsingu „LLM“ tulemuseks konkreetsed teadlased ja projektid, mitte nummerdatuna pseudoandmed „Isik 1“, „Isik 2“, „Projekt 1“, „Projekt 2“ jne, seeläbi luua ka näidisena isiku detailvaade, kus on toodud kontaktandmed ja ametikoht, et lihtsustada teadlastega kontakti leidmist, projekti detailvaade oli juba olemas.

4 Tulemused ja analüüs

See peatükk annab ülevaate andmeallikate analüüsi tulemustest, CRM-süsteemi põhimõtete rakendamisest, intervjuude järeldest ning selle põhjal loodud kasutajavajadustest. Samuti on käsitletud masinõppe rakendamise võimalusi.

4.1 Andmeallikate hindamise tulemused

Andmeallikate analüüsimisel peatükis 2.1 sai selgeks, et olemasolevad allikad ei paku ühtset ülevaadet AI-kompetentside olemasolust. ETIS võimaldab asutuse järgi filtreerida, kuid märksõnapõhine otsing puudub ning kui inimese CV või projekti pealkiri otsisõna ei sisalda, siis muidu sobilikku vastet tulemuses ei kuvata. Lisaks ei ole võimalik ETIS-es näha uurimisgruppe ning nendega seotud isikuid, seda infot on võimalik leida TalTechi veebilehelt. TalTechi digikogu annab hea ülevaate magistri- ja doktoritöödest, võimaldades ka märksõnapõhist otsingut, kuid projekte, inimesi ja uurimisgruppe seal ei kajastata.

Analüüsi tulemusena saadi kinnitust, et AI-kompetentsid on hajutatud erinevate süsteemide vahel ning olemasolevad lahendused ei võimalda näha terviklikku ülevaadet teadlaste, projektide, uurimisgruppide, publikatsioonide ja märksõnade vahelistest seostest. Teravalt puudutab see probleem AI-valdkonda, kus kompetentsid on hajutatud eri instituutide ja valdkondade vahele ning AI alamvaldkonnad ei ole ETIS-es ega CERCS-is eraldi esindatud.

Peatükis 2.1.1 toodud näite põhjal jäävad interdistsiplinaarsed AI-kompetentsid praeguste tööriistadega sageli leidmata ning AI-alane tegevus jääb seetõttu osaliselt nähtamatuks. Seega on vaja graafipõhist suhetesüsteemi, mis modelleerib teadlaste, projektide, uurimisgruppide ja märksõnade vahelised seosed ning võimaldab AI-kompetentside terviklikku kaardistamist.

4.1.1 Süsteemis kasutatavad andmeväljad

Loodava süsteemi eesmärk ei ole olemasolevate andmeallikate täielik dubleerimine, vaid AI-kompetentside kaardistamiseks vajalike andmete koondamine ja seostamine. Seetõttu kasutatakse igast andmeallikast ainult süsteemi funktsionaalsuse jaoks vajalikke andmevälju.

ETIS sisaldab detailset infot projektide, teadlaste ja publikatsioonide kohta. Sealt kogutakse üksnes esmaseks ülevaateks vajalikud andmeobjekti kirjeldavad andmeväljad. Põhjalikuma teabe leidmiseks on kasutajal võimalik pöörduda otse ETIS-esse.

Projektide puhul on asjakohasteks väljadeks pealkiri, kestusvahemik, eesmärk, projekti number, vastutav täitja, täitjad, teadus- ja arendusasutus, juhendaja ning seotud projektid. TalTechi kodulehelt uurimisgruppide alt leitava info põhjal on võimalik tuvastada, millised uurimisgrupid konkreetsete projektidega seotud on.

Isikute puhul on vajalikud väljad eesnimi, perenimi, sünniaeg, email, telefoninumber, teenistuskäik koos instituudi ja rolliga, seotud projektid ja publikatsioonid. Seotud magistri- ja doktoritööd pärinevad TalTechi digikogust. TalTechi veebilehelt uurimisgruppide alt leitava info põhjal saab määrata inimese rolli ja kuuluvuse uurimisgruppi.

Uurimisgruppide kohta leiab info TalTechi kodulehelt instituutide alt. Vajalikud andmeväljad on uurimisgrupi juht, liikmed, instituut, kirjeldus, kompetentsid ning seotud projektid ja publikatsioonid.

Publikatsioonide jaoks on digikogust olulised pealkiri, autor(id), märksõnad, kirjeldus, kaitsmiskuupäev, instituut, publikatsiooni liik (magistritöö, doktoritöö, artikkel jne), väljaandja, juhendaja, DOI. ETIS-es esinevad sama sisuga andmeväljad osaliselt erinevate nimetuste all.

Nende andmete koondamisel tekib süsteemis terviklik ülevaade TalTechi AI-valdkonna teadustegevusest ning erinevate objektide vahelistest seostest.

4.2 CRM-süsteemi põhimõtete rakendamine

Peatükis 2.2 väljatoodud info põhjal selgub, et CRM-süsteemide põhimõtted on loodavale süsteemile osaliselt ülekantavad. CRM-i suhete modelleerimine kattub hästi loodava süsteemi vajadustega.

AI-kompetentside suhete põhises kaardistussüsteemis on vaja säilitada ja kuvada suurt hulka erinevaid olemite vahelisi suhteid ja osaliselt neid ka graafil näidata. Lisaks on olulisel kohal suhete arengu jälgimine ajas, mis on tihedalt seotud ajadimensiooniga. Loodavas süsteemis on vaja säilitada infot inimese haridustee, AI-kompetentside arengu, inimesega seotud projektide jpm kohta, mistõttu tuleb tähelepanu pöörata ajaloo haldamisele ja atribuutide muutumisele.

Alapeatükis 2.2.3 väljatoodud atribuutide muutumise tehnikatest on antud lahendusele sobivaim dimensiooni rea lisamise tehnika. See jätab alles kogu vajaliku ajaloo ning võimaldab ka kompetentside arengut ja lisandumisi jälgida. Mitmetasandiline ajalugu ei ole vajalik, sest andmesisestuse ja tegeliku sündmuse vaheline ajaline nihe ei anna lisandväärtust ega pole analüütiliselt oluline.

Õiguste osas sobib AI-kompetentside kaardistussüsteemile rollipõhine ligipääsuhoodus, mis seob konkreetsed õigused rollidega. Rollipõhist ligipääsuhoodust soovitab ka NIST. See lihtsustab oluliselt õiguste haldamist, kuna kasutajale piisab sobiva rolli määramisest, et tagada talle vajalik, kuid piiratud ligipääs süsteemi funktsioonidele. Loodavas süsteemis on rollideks administraator, TalTechi juhtkond ja AIRE, kes peavad nägema eri vaateid süsteemist ning omama erinevaid õigusi.

4.3 Intervjuude põhjal kujunenud süsteeminõuded

Intervjuude ülevaatest alapeatükis 3.3.3 sai esimese intervjuu järgselt selgeks probleem, esialgsed kasutajad ja prototüübi suund. Samuti kinnitati idee esitada kasutajaliideses andmed kasutajale graafina, kuid graafivaate nägemisõigus täpsustus alles intervjuude järel.

Pärast teist intervjuud selgitati töö skoopi ning jõuti järeldusele, et teadlased kasutajatena jäävad selle töö raames välja. Esile tõstetud märksõnastamise probleem on süsteemi terviklikkuse seisukohalt oluline, kuid jääb käesoleva töö fookusest välja ning vajab tähelepanu süsteemi edasiarenduses. Kuna antud töö keskendub analüüsile ja kasutajaliidese disainile, siis edastati AIRE digitaliseerimise ja andmehalduse eksperdi Dmitri Derevjanko kontaktid süsteemi arendavale üliõpilasele.

Kolmas intervjuu kinnitas, et pilootprojekti sihtrühm on kitsas, seega otsustati teadlased sihtrühmast välja jätta. Samuti kitsendati graafivaate kasutusõigust AIRE-le, jättes selle üksnes juhtkonnale, järgides vähemate õiguste printsiipi, mis on põhjalikumalt kirjeldatud peatükis 2.2.4. Sarnaselt teise intervjuu järeldusele, jäi ka teadlaste peal töötava lahenduse testimise soovitus tulevikuks süsteemi loojatele. Näidismärksõnade arvu osas muudatusi ei tehtud, kuna viie kuni seitsme märksõna kuvamine on teadlikult valitud kognitiivse minimeerimise põhimõtetest, mida tutvustatakse lähemalt peatükis 5.5.1. Küll aga mõisteti segadust vahelehe nimetamise osas ning prototüübis muudeti „Kaardistamine“ ümber „Tegevusülevaateks“.

Läbiviidud intervjuudest selgus, et loodav pilootprojekt peab olema kasutatav algselt eelkõige juhtkonnale ning AIRE-le. Väliseid kasutajaid selle töö raames ei kaasata. Juhtkonna sooviks on lihtne ülevaade ning kiire informatsiooni kättesaadavus, AIRE sooviks on leida kõige sobivamaid partnereid sisestatud projektikirjeldustele. Lisaks peab Figma prototüüp sisaldama näidisandmeid ning kõik detailvaated peavad olema prototüübis näha.

4.4 Kasutajavajaduste analüüs

Intervjuude järgselt kaardistati kasutajavajadused: juhtkond soovib teadustööst ülevaate saamiseks võimalikult efektiivset otsingumootorit ja AIRE soovib leida parimaid koostööpartnereid oma klientidele.

Sõltuvalt juhtkonna ja AIRE erinevatest vajadustest, luuakse neile ka erinevad vaated. Seega tuleb süsteemi kaks kasutajarolli. Mõlemad rollid saavad kasutada otsingut ning näha tulemusi nii nimekirjana kui graafil. Lisaks kuvatakse mõlemale objekti detailvaates sobivusprotsenti.

Juhtkonnal on eraldi õigus näha märksõnade populaarsust mulldiagrammil, et saada kiireim ülevaade märksõnade kasutusest projektides. Märksõna kasutuse täpsemaks analüüsiks on juhtkonnale nähtav ka ajajoon, millel on välja toodud perioodide kaupa valitud märksõna kasutuse koguarv projektides ja publikatsioonides. Statistika eesmärgil luuakse ka CSV- ja XLS-andmeformaadis raporti allalaadimisvõimalus.

4.5 Masinõppe rakendamine

Peatükis 2.1 kirjeldatud ETIS-e näide illustreeris, et ainult märksõnapõhine otsinguloogika ei ole piisav sünonüümide ja sisulise konteksti tuvastamiseks. Selleks, et käesoleva töö raames loodav süsteem suudaks mõista päringu konteksti, on vaja masinõppel põhinevat semantilist vektoriseerimist. Semantiline vektoriseerimine võimaldab tuvastada sisulisi seoseid ka juhul, kui kasutaja päring ja andmebaasis olev tekst ei kattu täpselt (nt "tehisaru" vs "tehisintellekt"). Antud lähenemine tugineb Sentence-BERT (edaspidi SBERT) raamistikule, mille abil teisendatakse tekstid vektorkujule. See võimaldab arvutada tekstide vahelist sarnasust kiiresti ja suure täpsusega [40]. Otsingu matemaatiline loogika ja rakendatud valemid kirjeldatakse põhjalikumalt peatükis 5.3.

5 Süntees

Järgnevas peatükis antakse ülevaade süsteemi kasutajarollidest ja nende õigustest. Selgitatakse süsteemi struktuuri, andmemudelit ning süsteemi disaini koos kasutajaliidesega. Samuti kirjeldatakse otsingufunktsiooni matemaatilist mudelit ning antakse ülevaade projekti edasistest sammudest.

5.1 Kasutajarollid ja õigused

Selles peatükis selgitatakse süsteemis olevaid rolle ning nende õiguseid. Lisaks defineeritakse andmete muutmise strateegia.

5.1.1 Rollid

Pilootsüsteemis on esialgu kolm rolli: administraator, TalTechi juhtkond ja AIRE. Hiljem on võimalik seda laiendada ja rolle juurde lisada. Rollid on TalTechi siseveebis, seega süsteemivälisel kasutajal ei ole TalTech AI veebileheküljel ligipääsu „Tegevusülevaate“ vahelehele, kus asub otsingumootor.

5.1.2 Õigused

1. Kasutajal on olemas vaid vaatamisõigus, kuid kuna kasutajaid on kahte tüüpi, siis on ka vaatamisõiguseid kahte tüüpi:
 - a. Juhtkond näeb graafi-, detail- ja loendivaadet. Lisaks on nähtav ka märksõna populaarsust illustreeriv diagramm ning sellega kaasnev märksõnaspetsiifiline ajajoon. Juhtkonnal on ka õigus loendivaate tulemusi raportina alla laadida.
 - b. AIRE näeb tulemuste loendivaadet ja detailvaateid.
2. Administraator saab andmeid vaadata, muuta, kustutada, lisada. Tema poole jõuavad muutmissoovid, kui keegi kasutajatest leiab, et tema kohta käiv

informatsioon on ebakorrektne. Sellisel juhul on administraatoril võimalik probleemiga tutvuda ja vajadusel soovitud muudatused veebilehel ära parandada.

5.2 Süsteemi struktuur

Kontseptuaalne andmemudel põhineb eeldusel, et uuritav valdkond koosneb eristatavatest objektidest ehk olemitest ning nendevahelistest seostest. Selline lähenemine aitab analüütikul keskenduda sellele, milliseid andmeid on vaja kirjeldada, mitte sellele, kuidas neid tehniliselt salvestatakse [41].

Mudeli kolm põhilist komponenti on:

- objektid ehk olemid (*entities*),
- seosed (*relationships*),
- atribuudid (*attributes*).

5.2.1 Objektid

Objekt tähistab reaalse või abstraktse maailma nähtust, mille kohta süsteem andmeid säilitab. See võib olla füüsiline objekt, nt inimene või seade, või kontseptuaalne üksus, nt sündmus või teenus [41]. Antud töös kasutatavad objektid koos definitsioonidega on tabelis 2.

Tabel 2. Andmemudeli objektid ja nende definitsioonid.

isik	Teadlane, õppejõud, doktorant või muu füüsiline isik
projekt	Harilikult taotletava rahastusega ja kindla eesmärgiga ettevõtmine, mis viiakse ellu etteantud ajavahemikul [42]
publikatsioon	Trükis, internetis või muul viisil avaldatud harilik teaduslik tekst [43]
märksõna	Teemavaldkonda tähistav silt
uurimisgrupp	Millegi teaduslikuks läbitöötamiseks või põhjalikuks analüüsimiseks moodustatud teadlaste või muude isikute rühm [44]
asutus	Väline organisatsioon, mis osaleb koostööprojektides
instituut	TalTechi organisatsiooniline üksus

5.2.2 Atribuudid

Atribuut on objekti või seose omadus, mis kirjeldab selle tunnuseid ja millele saab omistada väärtuseid. Atribuutide abil saab objekte detailsemalt iseloomustada ja ühtselt identifitseerida [41]. Objektide atribuudid on paika pandud tabelis 3.

Tabel 3. Andmemudeli objektid, nende atribuudid ja näited.

Objekt	Atribuudid	Näited
Isik	Eesnimi	Anna
	Perenimi	Saar
	Sünniaeg	12.03.1985
	Telefoni number	53467842
	Email	anna.saar@taltech.ee
	Töökoht	IT Kolledž
	Amet	Insener
Projekt	Pealkiri	„Andmeränne: sotsiaalse transformatsiooni raamistik (ANDMERÄNNE)“
	Kirjeldus	Projekt keskendub uuendusliku andmerände teoreetilis-metodoloogilise raamistiku arendamisele
	Algusaeg	01.01.2026
	Lõppaeg	31.12.2030
Publikatsioon	Pealkiri	„A hybrid deep learning and large language models framework for ship collision accident analysis“
	Autor	Jakub Jerzy Montewka
	DOI	https://doi.org/10.1016/j.res.2026.112333
Uurimisgrupp	Nimi	Andmelabor
	Kirjeldus	Uurimisgrupis uurime andmete ja ühiskonna süvitsi põimumist ning analüüsime inimfaktori rolli selles võrrandis.
Asutus	Nimi	Tallinna Tehnikaülikool

Tabel 3. Andmemudeli objektid, nende atribuudid ja näited.

Objekt	Atribuudid	Näited
Instituut	Nimi	Arvutisüsteemide instituut
Märksõna	Nimetus	GPT
Andmeallikas	Nimetus	TalTechi veebileht
	Link	https://taltech.ee/en/datalab

5.2.3 Ajaloo ja ajajoone modelleerimine

Lähtudes peatükis 2.2.3 kirjeldatud SCD dimensiooni rea lisamise tehnikast, on süsteemis ajalistel seostel atribuudid *kehtib_alates* ja *kehtib_kuni*. Kui seos kehtib praeguseni, on *kehtib_kuni* määramata. Graafimudelis realiseeritakse see ajalise omaduste graafi (*temporal property graph*) põhimõtete järgi servade atribuutidena [45]. See võimaldab määrata seosele selge algus- ja lõppaja.

Järgnevalt loetletakse süsteemis eksisteerivad ajalised seosed:

- isik-instituut
- isik-uurimisgrupp
- isik-märksõna
- isik-projekt
- uurimisgrupp-projekt
- uurimisgrupp-märksõna

Isik võib vahetada instituuti, kuid info selle kohta tuleb säilitada. Isik ja uurimisgrupp võivad olla projektiga seotud osaliselt, näiteks kui on tegemist pika projektiga. Isik võib kuuluda elu jooksul erinevatesse uurimisgruppidesse. Isik ning uurimisgrupp võivad muuta enda fookust ning olla ajas erinevate märksõnadega seotud. Ajaliste seoste säilitamine võimaldab analüüsida ja jälgida kompetentside kujunemist ning koostöömustreid.

5.2.4 Loogiline andmemudel

Loogiline andmemudel koosneb järgmistest põhiobjektidest: isik, uurimisgrupp, projekt, publikatsioon, märksõna. Kõrvalobjektideks on asutus, teaduskond, instituut, andmeallikas. Atribuutidel on järgnevad andmetüübid:

4. String – tekstiline andmetüüp
5. Integer – numbriline andmetüüp
6. Date – kuupäeva andmetüüp
7. Year – aasta andmetüüp

Diagrammides on kasutatud UML 2.5.1 [46] standardile vastavaid andmetüüpe String ja Integer. Kuupäevade ja aastate esitamiseks on kasutatud tüüpe Date ja Year, mis järgivad rahvusvahelist ISO 8601 [47] standardit.

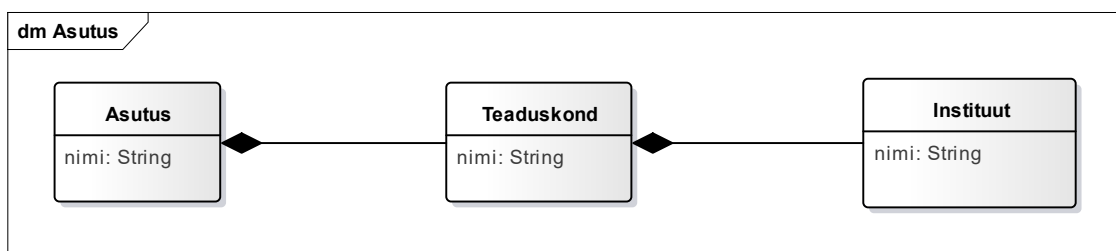
Diagrammidel on seoste kujutamiseks kasutatud UML 2.5 klassidiagrammi notatsiooni. Must romb noole otsas tähistab kompositsiooni, mis tähendab, et osa ei saa tervikuta eksisteerida. Tavaline joon tärnidega kujutab assotsiatsiooni koos mitmesusega, mis näitab mitu objekti võib seoses osaleda. Vaheklassid on kujutatud kriipsjoonega seotud eraldi klassidena, mis tähistavad assotsiatiivklasse. Assotsiatiivklasse kasutatakse, kui seosel on atribuudid (nt periood, roll), mida ei saa kummalegi objektile omistada [48].

Kasutatud mitmesused on välja toodud tabelis 4.

Tabel 4. Andmemudeli mitmesused.

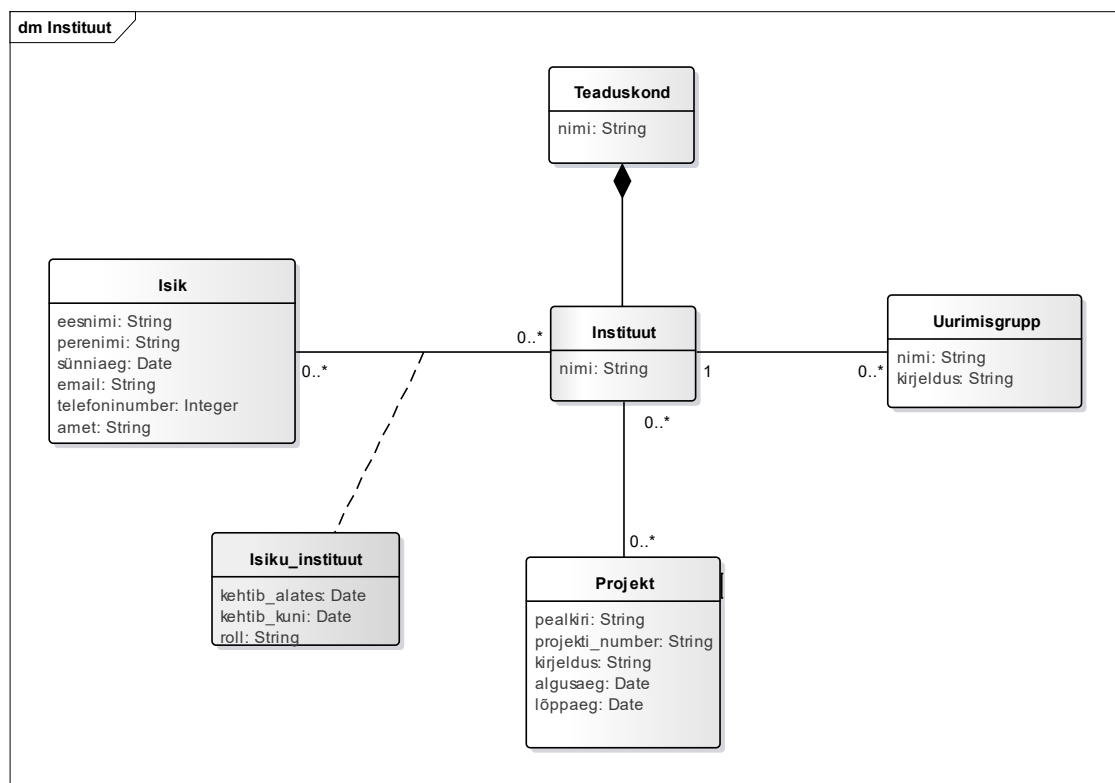
Mitmesus	Tähendus
1	täpselt üks
0..1	null või üks
0..*	null kuni mitu
1..*	vähemalt üks

Asutus on teadusorganisatsiooni struktuuris kõrgeimal tasemel. Asutuse sisse kuulub mitmeid teaduskondi, mille sisse omakorda erinevad instituudid. Nende vahel on range suhe, seega instituut ei saa eksisteerida ilma teaduskonnata ning teaduskond ei saa eksisteerida ilma asutusega. Küll aga võib asutus eksisteerida ilma teaduskonnata, sest edasisel arendamisel võib süsteemi lisanduda muid asutusi, mis ei pruugi olla ülikoolide struktuuriga. Asutusel, teaduskonnal ja instituudil on kirjeldavaks väljaks vastava struktuurüksuse nimi (vt joonis 1).



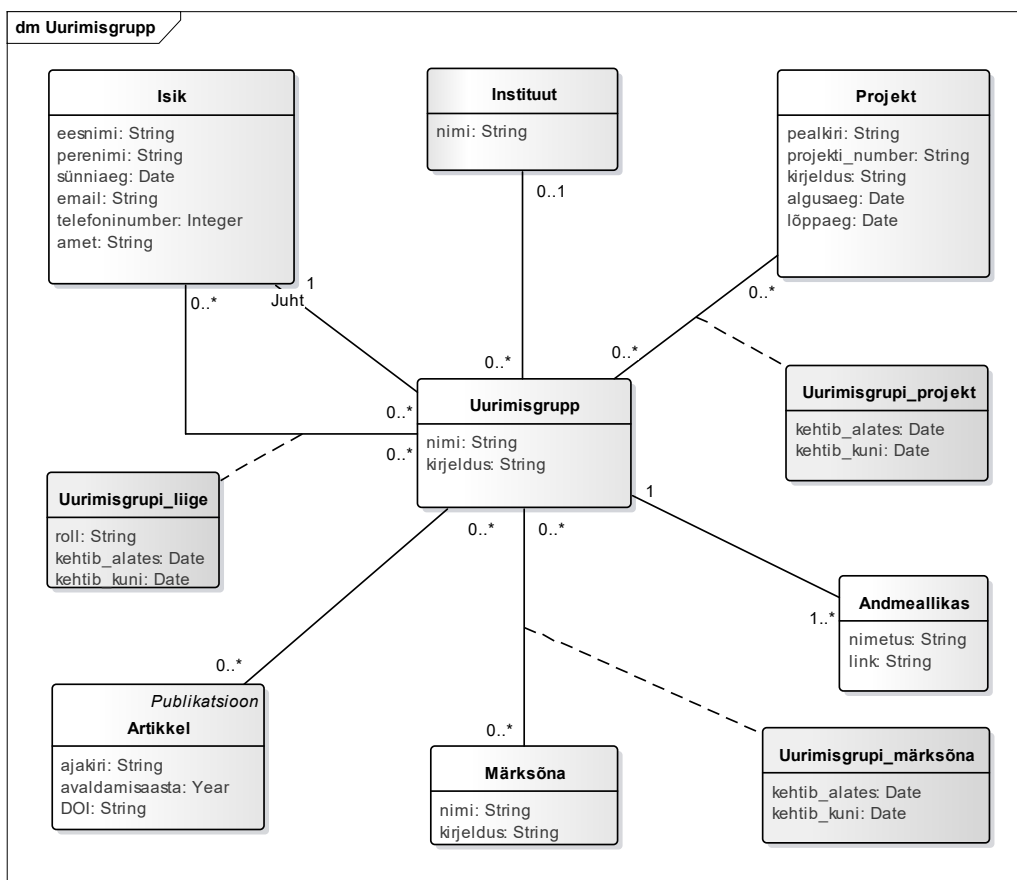
Joonis 1. Asutuse hierarhia.

Instituudiga on seotud mitmed isikud, uurimisgrupid ja projektid. Isiku ja instituudi vahel on ajaline seos, mis väljendab *kehtib_alates* ja *kehtib_kuni* väljade kaudu, kui kaua inimene antud instituudis on töötanud. Lisaks on seosel väli *roll*, mis väljendab, milline amet, ülesanne inimesel instituudis sel perioodil on (vt joonis 2).



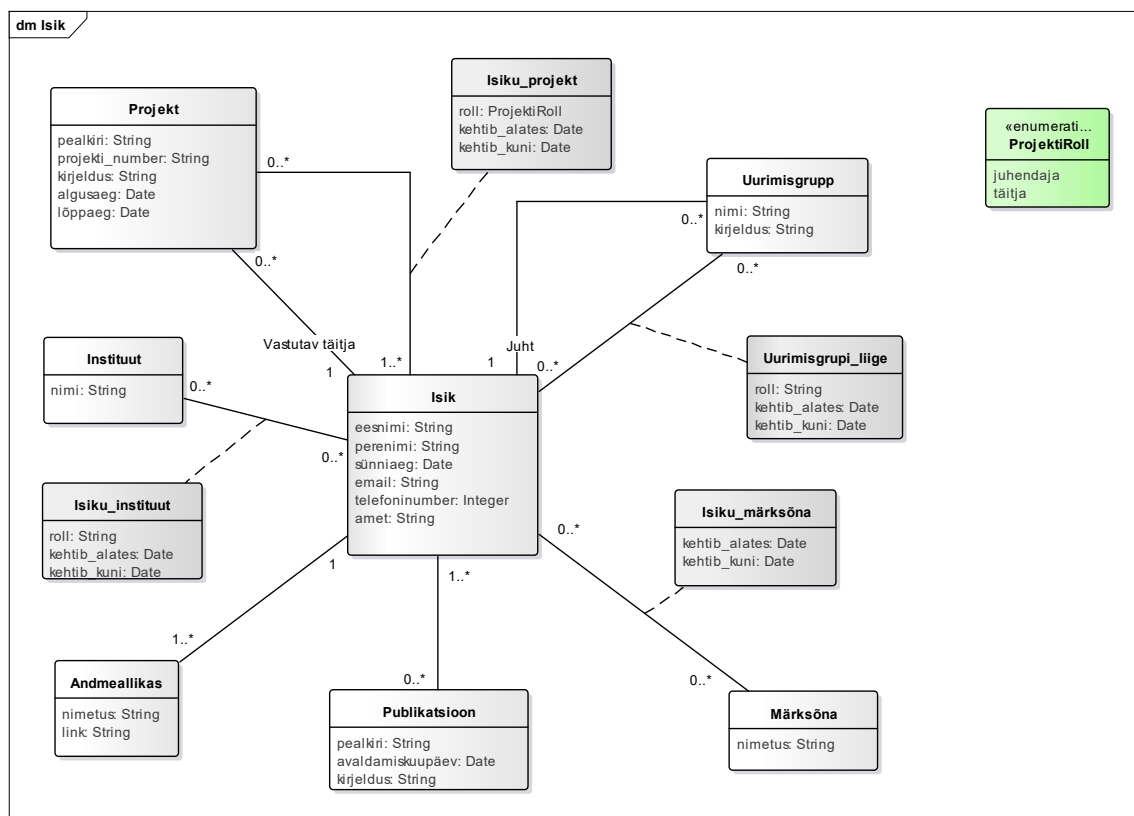
Joonis 2. Instituudi seosed.

Uurimisgrupp on seotud mitmete projektide, artiklite, märksõnade, isikute ja andmeallikatega. Uurimisgrupp sisaldab kahte atribuuti: nime, mis identifitseerib gruppi, ning kirjeldust, mis annab ülevaate grupi tegevusest ja fookusest. Uurimisgrupil on üks kindel instituut, aga on arvestatud ka võimalusega, et võib tekkida instituutide üleseid või ajutisi uurimisgruppe, seega võib instituut ka puududa. Uurimisgrupil on isikuga kaks seost: uurimisgrupil peab olema juht ja uurimisgrupil saavad olla liikmed. Liikmeseos on ajaline ning *Kehtib_alates* ja *Kehtib_kuni* väljade kaudu saab määrata, kui kaua inimene uurimisgrupiga seotud on. Projektidega on uurimisgrupil ajaline seos, et määrata, kui kaua uurimisgrupp projektiga seotud on. Uurimisgrupil on ajaline seos ka märksõnadega, et kindlaks määrata, kui kaua üks uurimisgrupp ühe AI-valdkonnaga seotud on. Lisaks on uurimisgrupp seotud vähemalt ühe andmeallikaga, mis viitab, kust andmed on hangitud (vt joonis 3).



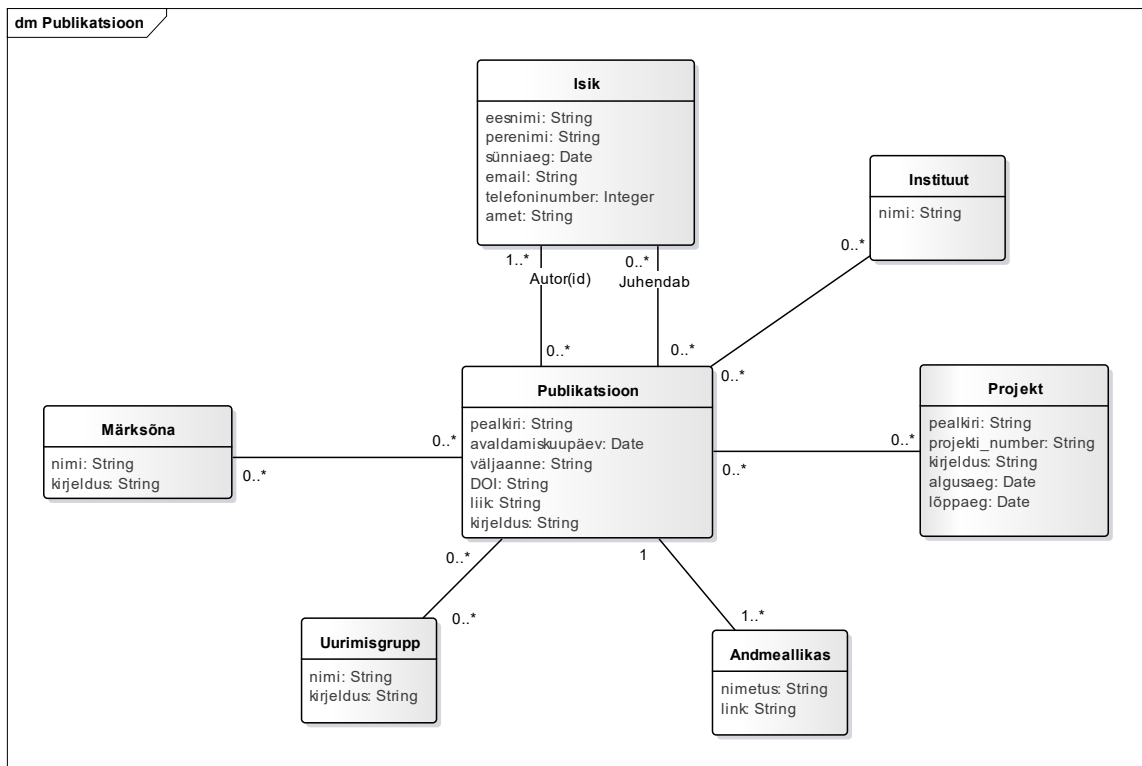
Joonis 3. Uurimisgrupi seosed.

Isikul on 6 kirjeldavat välja: eesnimi, perenimi, sünniaeg, email, telefoninumber, amet. Isik saab olla uurimisgrupis juht ja projektis vastutav täitja. Samuti võib isik uurimisgruppi, projekti ja lisaks instituuti liikmena kuuluda. Liikmeseos nendes on ajaline, et määrata, mis perioodil isik uurimisgrupis, projektis, instituudis liige on ning lisaks milline roll/amet isikul uurimisgrupis, projektis, instituudis sel perioodil on. Projektis on võimalikud kolm rolli: juhendaja, täitja ning vastutav täitja. Isik võib olla lisaks projektidele seotud publikatsioonidega ja märksõnadega. Märksõnadega seosel on periood, et kuvada, kui kaua üks isik ühe AI-valdkonnaga seotud on. Lisaks on isik seotud vähemalt ühe andmeallikaga, et kuvada, kust info pärineb (vt joonis 4).



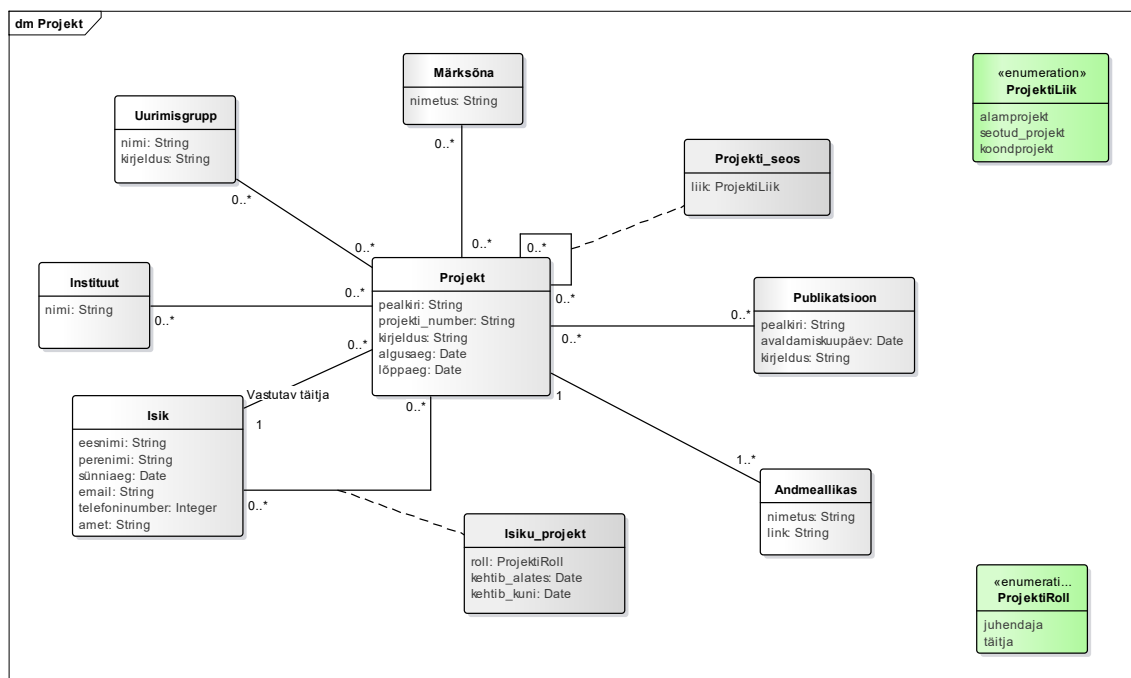
Joonis 4. Isiku seosed.

Publikatsioonil on pealkiri, avaldamiskuupäev, väljaanne, DOI, liik ja kirjeldus. Liik väljendab, kas tegemist on näiteks magistritöö, doktoritöö, artikli, raamatuga vms. Publikatsioon peab olema seotud vähemalt ühe isikuga, kes on autor ning vähemalt ühe isikuga, kes on juhendaja. Publikatsiooniga peab olema seotud vähemalt üks andmeallikas, kust info on hangitud. Lisaks võib publikatsioon olla seotud erinevate märksõnade, uurimisgruppide ning projektidega (vt joonis 5).



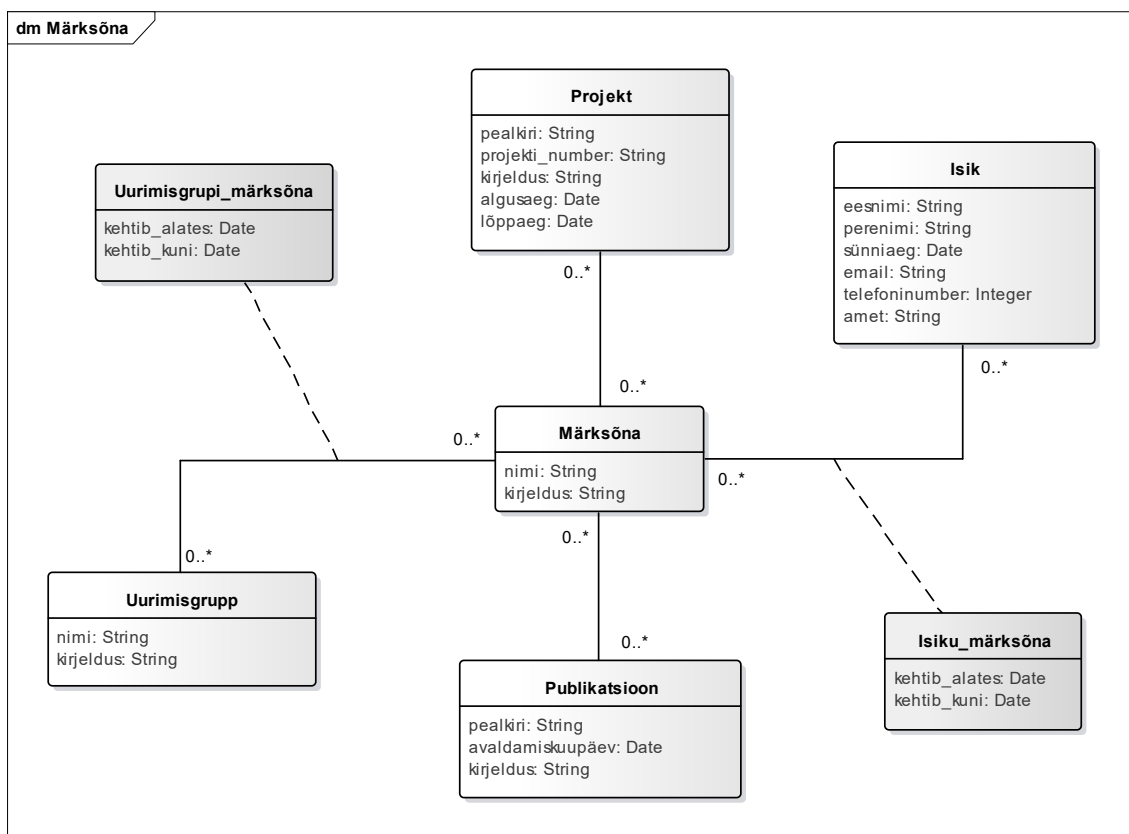
Joonis 5. Publikatsioonide seosed.

Projektil on pealkiri, projekti_number, kirjeldus, algusaeg, lõppaeg. Projektiga võib olla seotud mitu inimest, instituuti, uurimisgruppi, märksõna, projekti, publikatsiooni. Vaheklassil Projekti_seos on väli projektiLiik, et määrata, kas tegemist on alamprojekti, seotud projekti või koondprojektiga. Vaheklass Isiku_projekt kirjeldab, millist rolli isik projektis omab (juhendaja või täitja). Lisaks saab selles vaheklassis määrata seose kehtivuse perioodi. Projektil peab olema üks vastutav täitja ning vähemalt üks andmeallikas, kust info pärineb (vt joonis 6).



Joonis 6. Projekti seosed

Märksõnal on kaks atribuuti: üks kirjeldab nime ja teine märksõna sisu. Märksõna võib olla seotud projektide, publikatsioonide, isikute ja uurimisgruppidega. Isiku_märksõna ja Uurimisgrupi_märksõna vahetabelite abil saab määrata perioodi, kui kaua märksõna isiku või uurimisgrupiga seotud on (vt joonis 7).



Joonis 7. Märksõna seosed.

Kirjeldatud loogiline mudel on loodud vastenduma graafandmebaasi mudeliga. Objektid vastavad sõlmedele, seosed servadele ning vaheklassid on servade omadused. Märksõnad on mudelis kujutatud eraldi sõlmedena, mitte sõlmede siltide ega omadustena. Kuna märksõnadel on oma atribuudid ning nendega on seotud ajaliselt kehtivad seosed isikute ja uurimisgruppidega, on nende modelleerimine eraldi sõlmedena põhjendatum, sest see võimaldab märksõna kaudu navigeerida kõigi seotud objektideni ning analüüsida, kuidas AI-valdkondade katvus ajas muutub.

5.2.5 Andmete muutmise strateegia

Andmete ajakohasuse tagamiseks tuleb süsteemis määrata strateegia, kuidas andmeid uuendatakse ning kes ja kuidas muudatusi valideerib. Kaaluti kolme võimalikku lähenemist: täisautomaatset, manuaalset ja hübriidset.

Täisautomaatse lähenemise puhul imporditakse andmed allikatest süsteemi ilma inimese sekkumiseta etteantud ajavahemike järel. Eeliseks on minimaalne halduskoormus ning andmete pidev ajakohasus, kuid kolme erineva allika kombineerimisel on konfliktide risk

liiga kõrge, et sellele täielikult toetuda. Näiteks võivad sama teadlase ametinimetus ETIS-es ja TalTechi veebilehel erineda ning süsteem ei suuda sellist vastuolu iseseisvalt lahendada. Manuaalse lähenemise puhul sisestab ja uuendab andmeid administraator käsitsi, mis tagab täieliku kontrolli andmekvaliteedi üle, kuid arvestades TalTech AI-valdkonna teadlaste, projektide ja publikatsioonide mahtu, ei ole see pikemas perspektiivis jätkusuutlik.

Seetõttu valiti süsteemi aluseks hübriidlähenemine, mis ühendab automaatse andmeimpordi administraatori poolse valideerimisega. Andmed imporditakse allikatest automaatselt, kuid konfliktide või ebaselgete kirjete korral edastatakse need administraatorile kontrollimiseks. See tasakaalustab andmete ajakohasuse vajaduse ning andmekvaliteedi tagamise ja on kooskõlas kasutajaloos 7 kirjeldatud nõuetega. Konfliktide edastamise täpne viis jääb käesoleva töö skoobist välja ning vajab edasist lahendamist süsteemi arendamisel.

Kõik andmemuudatused logitakse süsteemis koos kolme atribuudiga: muudatuse kuupäev, muudatuse allikas ehk kas muudatuse tegi süsteem andmeimpordi käigus või administraator käsitsi ning muudetud välja eelmine ja uus väärtus. Logimine võimaldab administraatoril muudatused vajadusel tagasi võtta, taastades eelmise andmeväärtuse logikirje põhjal, nagu kirjeldatud kasutajaloos 7.2. Lisaks toetab see peatükis 2.2.3 kirjeldatud SCD dimensiooni rea lisamise tehnikat, mille kohaselt säilitatakse kõik ajalised muudatused ning ühtegi varasemat väärtust ei kirjutata lihtsalt üle.

Administraator on süsteemi ainuke roll, kellel on õigus andmeid lisada, muuta ja kustutada, nagu kirjeldatud peatükis 5.1.2. Automaatimport tuvastab konfliktid väljad ning märgib need administraatorile ülevaatamise järjekorda, kes otsustab, milline allikas on usaldusväärne, tuginedes 4.1 kirjeldatud allikate usaldusväärsuse hierarhiale. Lisaks on kasutajatel võimalik teavitada administraatorit ebakorrektest infost, misjärel administraator saab tutvuda probleemiga ning teha vajalikud parandused. Andmeid uuendatakse süsteemis üks kuni kaks korda kuus, mis tagab, et info on piisavalt ajakohane ilma, et administraatori koormus muutuks liiga suureks.

5.3 Otsing

Otsingutehnoloogiate valimisel lähtuti kasutajate vajadustest, mis intervjuu käigus kaardistati. AIRE soovib otsingukasti sisestatud potentsiaalselt pika projektikirjelduse põhjal leida sobivaid teadlasi ja uurimisgruppe. Selline kasutusjuht eeldab semantilist otsingut, sest päringu ja vastete vahel ei pruugi esineda täpset sõnade kattuvust. Juhtkonna eesmärk on leida võimalikult kiiresti ja konkreetseid vastuseid, seega selleks on täpne tekstiotsing tõhusam. Mõlema kasutusjuhuga arvestamiseks valiti hübriidlähenemine. Käesolev peatükk kirjeldab otsingu põhimõtteid, sobivusprotsendi leidmise meetodeid ning filtreerimist ja sorteerimist. Käesoleva peatüki loomisel on kasutatud tehisintellekti mudeleid Gemini 3 Flash ja Claude Sonnet 4.6, et arendada sobivusprotsendi loogikat.

5.3.1 Otsingu põhimõtted

Semantilise otsingu teostamiseks valiti SBERT raamistik. Mudeli valikul lähtuti AIRE soovitusel kasutada *embedding*-põhist lähenemist ning asjaolust, et SBERT on avatud lähtekoodiga, tasuta kasutatav ja võimalik kasutada enda serveris.

Semantiline vektoriseerimine võimaldab tuvastada sisulisi seoseid ka juhul, kui kasutatud märksõnad ei kattu täpselt. SBERT raamistiku abil teisendatakse tekstid vektorkujule. Seejärel mõõdetakse otsingufraasi ja kirje vahelist sarnasust kosinus-meetodil, mis väljendab tekstide tähenduslikku kattuvust skaalal 0–1 [40].

Tekstipõhise otsingu teostamiseks valiti BM25 algoritm. See hindab vastete asjakohasust päringu sõnade esinemissageduse alusel ning järjestab vasted asjakohasuse tõenäosuse järgi kahanevas järjekorras. Lisaks arvestab algoritm päringu sõna esinemissagedusega vastes ning jõuab ühel hetkel küllastuspunkti ehk asümptootilise limiidini. Seega ühe sõna panus vaste skoorile ei saa lõpmatuseni kasvada, mis väldib olukorda, et näiteks pikad artiklid saavad tugeva eelise lühemate ees [49].

Hübriidotsingu loogika rakendamiseks tuleb skoorid normaliseerida ning leida kaalutud summa [49].

$$\text{hübriidskoor} = \alpha \cdot \text{koosinussarnasus} + (1 - \alpha) \cdot \text{normaliseeritud(BM25)} \quad (1)$$

Valemis tähistab α kaalutegurit, mis määrab semantilise ja tekstipõhise osa tasakaalu. 0.7 on sobilik α -väärtus, sest sel viisil on semantilisus prioriteedis, kuid arvestatakse ka märksõnade kattuvuse mõju lõpptulemusele. On arvestatud sellega, et AIRE peamiseks kasutusjuhiks on erineva pikkusega kirjelduste sisestamine, mis eeldab semantilist mõtlemist ning tingib seetõttu suurema osakaalu.

5.3.2 Sobivusprotsent

Kliendipoolseks sisendiks oli probleem, et paari märksõna sidumisel konkreetse profiiliga, jääb süsteemidesse märk, justkui oleks see isik antud valdkonnas ekspert. Näiteks on teadlane juhendaja teadustöös, mille teema pakub talle huvi, kuid tema teadmised on algtasemel. Kuna teema seostatakse tema nimega, hakkab tulema pakkumisi valdkonna konverentsidele või teadustööde kaitsmistele, mille jaoks vastav pädevus tegelikult puudub. Lisaks on problemaatiline ka info ajakohasus, sest ka otsinguga hästi sarnanev vaste kaotab oma väärtuse, kui see pärineb aastakümnete tagant.

Selleks, et nimetatud probleeme lahendada, kuvatakse vastete juures sobivusprotsent. Sobivusprotsent kasutab peatükis 5.3.1 kirjeldatud skoori, mis väljendab maksimaalset asjakohasust nii pikkade kirjelduste kui ka täpsete terminite puhul. Lisaks võetakse arvesse teadlase kogemust valdkonnas kui ka informatsiooni värskust, et vältida vananenud informatsiooni ja teadmust.

Sobivusindeksi matemaatilise mudeli väljatöötamisel kasutati abistavalt tehisintellekti mudeleid Gemini 3 Flash [34] ja Claude Sonnet 4.6 [35], eesmärgiga luua kaalutud pingerida, mis arvestab nii semantilist sarnasust, ajakohasust kui ka teaduslikku mõju.

Eksperdi kogemuse leidmiseks antud valdkonnas tuginetakse publikatsioonide arvule ja Hirsch'i poolt defineeritud teadustöö mõjukuse indeksile (h-indeks) [50]. Selleks, et vältida staažikamate uurijate ebaõiglast domineerimist pingereas, rakendatakse andmete logaritmilist normaliseerimist valemis (2) ja (3). Antud lähenemine järgib Craswell'i jt. kirjeldatud hindamismudelit [51], kus väärtuse kasvades selle lisanduv mõju koondskoorile järk-järgult väheneb. Selline kombineeritud lähenemine tagab tasakaalustatud ja objektiivse võrdlusmomendi erineva staažiga ekspertide vahel.

$$\text{sügavusskoor} = \log(1 + \text{publikatsioonide_arv}) \cdot \text{tsiteerimiste_kaal} \quad (2)$$

$$\text{tsiteerimiste_kaal} = \log(1 + \text{h_indeks}) / \log(1 + \text{max_h_indeks}) \quad (3)$$

Hetkel olemasolevate allikate põhjal pole teadustöö mõjukuse indeksi kasutamine võimalik, kuna valitud andmeallikatest pole see kättesaadav. Google Scholar on üks allikatest, mis seda indeksit kuvab, kuid seda ainult teadlastele, kellel on profiil olemas. Kuna peatükis 2.1.4 välja toodud põhjuste tõttu Google Scholarit käesolevas töös andmeallikana ei kasutata, jäetakse valem (3) edasiseks arenduseks ning pilootlahenduses rakendatakse lihtsustatud valemit (4):

$$\text{sügavusskoor} = \log(1 + \text{publikatsioonide_arv}) \quad (4)$$

Teadustöö värskuse arvutamiseks rakendatakse eksponentsiaalset kahanemisfunktsiooni, kus λ on kahanemiskiiruse parameeter:

$$\text{uudsusskoor} = e^{(-\lambda \cdot \text{aastad_viimasest_publikatsioonist})} \quad (5)$$

Kuna AI-valdkond on kiiresti arenev, siis on λ väärtuseks valitud 0.15. Sellise väärtuse juures saab viis aastat vana publikatsioon väärtuseks ligikaudu 0.47, mis tähendab, et see on endiselt arvestatav, kuid selgelt madalama skooriga kui hiljutisem töö. See aitab vältida vananenud info ülehindamist kiiresti arenevas tehnoloogiasektoris. Mudelis rakendatud eksponentsiaalne kahanemisfunktsioon põhineb Longi ja Changi kirjeldatud valdkonnapõhiste otsingumootorite relevantsuse teoorial, kus info värskust käsitletakse kaalutud faktorina [52]. Selline lähenemine eelistab tulemustes aktiivseid eksperte, kuid jätab nähtavaks ka olulised varasemad teadustööd vastavalt nende vähenevale aktuaalsusele.

Eelnevalt kirjeldatud kolme komponendi kombineerimiseks rakendatakse kaalutud summa meetodit [49]. Otsingutulemuste asjakohasuse tagamiseks on suurim osakaal määratud sisulisele kattuvusele (50%), millele järgnevad eksperdi kogemus (30%) ja teadustöö värskus (20%). Nende tegurite koondamisel saadakse lõplik sobivusindeks järgneva valemi alusel:

$$\text{sobivusindeks} = 0.5 \cdot \text{hübriidskoor} + 0.3 \cdot \text{sügavusskoor} + 0.2 \cdot \text{uudsusskoor} \quad (6)$$

Valemi kaalude jagunemine on tingitud sihtgruppide vajadustest. Kuna AIRE peamiseks sooviks on leida sisestatud tekstile sobivaimat vastet, on valitud hübriidskoori kaaluks 0,5. Teadustöö värskus on AI-valdkonnas oluline, kuid mitte primaarne ning ei tohiks varjata kogemust. Seega on teadustöö värskuse kaaluks 0,2 ning valdkonna kogemuse kaaluks 0,3.

Lõpptulemuse paremaks tõlgendamiseks teisendatakse saadud indeks standardseks sobivusprotsendiks:

$$\text{sobivusprotsent} = \text{ümardatud}(\text{sobivusindeks} \cdot 100) \quad (7)$$

Kirjeldatud matemaatiline mudel on loodud sobivate inimeste leidmiseks, kuid on laiendatav ka teistele andmeobjektidele. Uurimisgruppide puhul koondatakse liikmete sügavuse ja värskuse näitajad ning arvutatakse grupi koondindeks liikmete näitajate aritmeetilise keskmise alusel. Projektide ja publikatsioonide puhul rakendatakse lihtsustatud mudelit, kus sobivusprotsent leitakse hübriidse sisulise sarnasuse ja ajakohasuse kaalutud summana. Selline lähenemine võimaldab võrrelda erinevat tüüpi vastete sobivust päringule ühtsel skaalal.

5.3.3 Graafivaade

Graafivaate eesmärk on pakkuda visuaalset ja interaktiivset liidest, mis koondab hajutatud teadusandmed ühtseks võrgustikuks. Kuna tegemist on disainprototüübiga, on fookus seatud informatsiooni kättesaadavusele ja seoste avastamise lihtsusele, mis toetab juhtkonna strateegilist vaadet. Graafil on erinevad objektitüübid eristatud erinevate värvidega, et saada kiire ülevaade tulemustest.

Erinevalt staatilisest tabelist on graafivaade loodud dünaamilisena. Peamine väärtus seisneb seoste sisulises avamises. Esiteks, kui graafis kuvatakse seos kahe objekti, näiteks kahe teadlase vahel, siis servale klikates genereeritakse infoaken, mis koondab kokku spetsiifilised andmed, mis antud seost õigustavad. Seal kuvatakse ühised publikatsioonid, koos läbi viidud projektid, samasse uurimisgruppi kuulumised ning juhendamissuhted.

Selline lahendus võimaldab kiiresti mõista koostöö olemust ilma süsteemis täiendavaid otsinguid tegemata.

Selleks, et vältida info üleküllust on prototüübis ette nähtud filtreerimisfunktsioon. Kasutaja saab kohandada graafi vastavalt huvipakkuvale valdkonnale või ajaraamile.

Disainilahendus rõhutab avastatavust, see tähendab, et isegi algelises vormis prototüüp demonstreerib, kuidas graafipõhine lähenemine toob esile teadlastevahelised nii-öelda peidetud võrgustikud, mis on sisendiks uute koostöövõimaluste loomisel või teemade arendamisel.

5.3.4 Sorteerimine

Sorteerimine on võimalik sobivusprotsendi, tähestiku ja uudsuse järgi. Vaikimisi on tulemused järjestatud sobivusprotsendi järgi, sobivamad eespool, sest loendivaate peamine kasutaja on AIRE, kelle eesmärgiks on leida sisendtekstile kõige asjakohasem vaste. Sobivusprotsendi leidmine on kirjeldatud alapeatükis 5.3.2.

Tähestiku järgi sorteerimine on lisatud funktsioonina juhuks, kui peaks tekkima vajadus leida loendist konkreetset nime või nimetust. Uudsuse järgi sorteerimine võimaldab leida kõige hiljutisemalt avaldatud projekte ja publikatsioone. Inimesed ja uurimisgrupid reastatakse uudsuse järgi nende seotud tegevuste kaudu: eespool kuvatakse isikud ja grupid, kellel on süsteemis enim värskeid (ajaliselt hilisemaid) publikatsioone ja projekte.

5.4 Süsteemi disain ja kavandamine

Selles peatükis tutvustatakse lühidalt, mis on nõuded ja kasutuslood ning kirjeldatakse nende rakendamist loodavas süsteemis.

5.4.1 Nõuded

Nõuded kirjeldavad süsteemi oodatavat käitumist, omadusi ja piiranguid, luues aluse edasisele arendusprotsessile. Selgelt sõnastatud nõuded suunavad disainivalikuid ning tagavad, et loodav toode on kooskõlas sidusrühmade vajaduste ja ootustega [53].

TalTechi juhtkonnal on võimalus kasutada süsteemi järgmiste funktsioonide jaoks:

- Teostada otsingut märksõna või sisestatud teksti põhiselt
 - Näha loendivaadet
 - Näha tulemuse sobivusprotsenti vastavalt sisestatud tekstile
 - Sorteerida otsingutulemusi sobivuse, uudsuse ja tähestiku järgi.
 - Laadida alla XLS/CSV formaadis aruande
 - Avada projekti, asutuse, isiku ja uurimisgrupi detailvaate
 - Näha allikaid, kust keskkonnast info päriti
 - Näha graafi
 - Näha graafil objektide vaheliste seoste kirjeldust
 - Filtreerida graafi objektide põhiselt
 - Määrata graafil perioodi
- Näha märksõnade kasutust
 - Näha märksõnade kasutussagedust diagrammil
 - Näha üksikute märksõnade arengut ajajoonel

AIRE kasutajatel on võimalik:

- Teostada otsingut märksõna või sisestatud teksti põhiselt
 - Näha loendivaadet
 - Näha tulemuse sobivusprotsenti vastavalt sisestatud tekstile
 - Sorteerida otsingutulemusi sobivuse, uudsuse ja tähestiku järgi.
 - Avada projekti, asutuse, isiku ja uurimisgrupi detailvaate
 - Näha allikaid, kust keskkonnast info päriti

Mittefunktsionaalsed nõuded:

- AI-kompetentside tegevusülevaate vahelehe värvid ja disain peavad sobima TalTech AI veebilehele.
- AI-kompetentside tegevusülevaate vahelehte näevad ainult selleks volitatud kasutajad.
- Administraatoril on õigus veebilehel olevat informatsiooni vaadata, muuta, kustutada, lisada
- Administraator valideerib automatiseeritud andmeimporti.

- Veebilehel uuendatakse informatsiooni 1-2 korda kuus
- Isikute ja uurimisgruppide detailvaates on kuvatud tulemise sobivusprotsent vastavalt otsingukasti sisestatud tekstile.
- Märksõnade populaarsust kujutaval diagrammil värvub märksõnale vajutamisel või märksõna otsimisel sellele vastav objekt.
- Graafil kuvatakse erinevaid objekte erinevate värvidega, värvide legend on kuvatud graafi all paremas servas.
- Graafi filtreerimisel eristatakse seose tugevust objektide vahel visuaalselt paksema servaga.

5.4.2 Kasutajalood

Kasutajalugu on funktsionaalsuse kompaktne kirjeldus lõppkasutaja perspektiivist. Kasutajalood pakuvad kasutajakesksemat lähenemisviisi, võimaldades paremini mõista kasutajate ootusi ning tagada nende kooskõla arendustegevusega [54]. Järgnevalt antakse ülevaade AI-kompetentside kaardistussüsteemi kasutajalugudest.

1. Projektide ja kompetentside ülevaade

Roll: AIRE

Kasutajalugu: Kasutajana soovin leida kliendi nõuetele vastavaid partnereid, et viia klient sobiva teadlasega kokku.

Aktsepteerimiskriteeriumid:

1.1 Süsteem võimaldab otsingumootoris sisestada projektikirjeldusi.

1.2 Süsteem kuvab sobivaid vasteid sisestatud tekstile, näidates ka otsingusse sisestatud tekstiga kattuvusprotsenti.

2. Graafivaade

Roll: TalTechi juhtkond

Kasutajalugu: Kasutajana soovin näha teadlaste vahelisi seoseid graafivaates, et mõista koostöövõrgustikku.

Aktsepteerimiskriteeriumid:

- 2.1 Graafis kuvatakse teadlased sõlmedena (nodes).
- 2.2 Graafis kuvatakse seosed teadlaste vahel servadena.
- 2.3 Kasutaja saab graafivaates serval klikkida ja näha seose detailinfot (mis seos on sõlmede vahel).
- 2.4 Graaf on filtreeritav.
- 2.5 Graaf on navigeeritav.

3. AI valdkonna teadusala areng

Roll: Taltechi juhtkond

Kasutajalugu: Kasutajana soovin näha, millised AI-kompetentsid on kaetud, et edendada valdkonna arengut.

Aktsepteerimiskriteeriumid:

- 3.1 Märksõnade diagrammil kuvatakse märksõnu mullidena.
- 3.2 Kasutaja saab mulldiagrammil klikkida ja näha üksiku märksõna kasutust ajas.
- 3.3 Märksõnal kuval saab määrata perioodi, mis kuvatakse diagrammil ja ajajoonel.
- 3.4 Märksõna mulli suurus sõltub perioodis selle kasutusest perioodi jooksul projektidest ja publikatsioonides.
- 3.5 Vaikimisi on kuvatud perioodiks viimased viis aastat.

4. Otsingutulemuste sorteerimine

Roll: TalTechi juhtkond, AIRE

Kasutajalugu: Kasutajana soovin kategooriapõhiselt otsingutulemusi loendivaates sorteerida, et leida kõige sobivamad vasted võimalikult kiiresti.

Aktsepteerimiskriteeriumid:

4.1 Süsteem kuvab otsingutulemused kategooriate kaupa loendina, kust kasutaja saab avada konkreetse kategooria tulemused vajutades nuppu „Vaata kõiki“.

4.2 Kasutaja saab tulemusi sorteerida sobivuse, uudsuse ja tähestiku järgi.

4.3 Vaikimisi on tulemused sorteeritud sobivuse järgi, sobivamad eespool.

5. Detailvaate avamine

Roll: TalTechi juhtkond, AIRE

Kasutajalugu: Kasutajana soovin avada otsingutulemuse detailvaate, et saada objekti kohta rohkem infot ning leida kontaktandmeid.

Aktsepteerimiskriteeriumid:

5.1 Kasutaja saab avada projekti, isiku, uurimisgrupi ja publikatsiooni detailvaate.

5.2 Detailvaates kuvatakse objektiga seotud märksõnad ning sobivusprotsent sisestatud otsingutekstile.

5.3 Detailvaates kuvatakse allikad, kust info on pärit, koos otselinkidega.

5.4 Detailvaates kuvatakse iga objektitüübi kohta sellele vastav informatsioon.

5.4.1 Isiku ja uurimisgrupi detailvaates kuvatakse kontaktandmed.

6. Tulemuste eksportimine

Roll: TalTechi juhtkond

Kasutajalugu: Kasutajana soovin eksportida otsingutulemusi aruandena, et teha andmepõhiseid otsuseid väljaspool süsteemi.

Aktsepteerimiskriteeriumid:

6.1 Kasutaja saab eksportida kategooriaspetsiifilise nimekirja tulemusi CSV või XLS formaadis.

6.2 Eksportitud fail sisaldab samu kirjeid, mis on kuvatud loendivaates.

7. Andmete haldamine

Roll: Administraator

Kasutajalugu: Administraatorina soovin hallata süsteemis olevaid andmeid, et tagada nende õigsus ja terviklus.

Aktsepteerimiskriteeriumid:

7.1 Administraator saab lisada uusi objekte (isikud, projektid, uurimisgrupid, publikatsioonid, märksõnad).

7.2 Administraator saab muuta olemasolevate objektide andmeid.

7.3 Administraator saab kustutada objekte.

7.4 Kasutaja saab teavitada administraatorit ebakorrektest infost ning administraator saab probleemiga tutvuda ja vajalikud muudatused teha.

5.5 Kasutajaliidese disaini põhimõtted

Kasutajaliidese prototüübi loomiseks kaaluti mitmeid variante ja platvorme, nende hulgas kombinatsiooni AI-st ja Lovable'ist, päris koodi kirjutamist või Figmas klikitava, kuid ilma tehnilise taustata, näite disainimist. Kuna antud töö skoobiks sai siiski vaid analüüs ja disain, siis koodi kirjutamine välistati pea kohe. Lovable'iga tehti mõned katsetused, kuid ka sellest variandist otsustati loobuda. Seega jäi alles vaid Figma, millega ühel autoritest oli juba varasem kogemus ning platvorm ei olnud täiesti võõras.

Figma on koostööl põhinev veebirakendus lihtsal viisil kasutajaliidese loomiseks. Koostööl põhinevaks teeb rakenduse võimalus töötada samal ajal samal töölehel mitmel tiimiliikmel korraga. Platvorm annab kasutajaliidese loomiseks oma toodetega mitmeid võimalusi: Figma Design laseb prototüüpi luua disainimise teel, Figma Make promptimisega, Dev Mode loob disainist koodi. Erinevaid loomisvõimalusi on seal veelgi ning muidugi on erinevate toodete kasutamiseks ja limiitide suurendamiseks tarvis toodete eest ka maksta [32]. Kaardistussüsteemi prototüübi loomiseks kasutati Figma Design toodet, milles oli väga lihtne luua ai.taltech.ee veebilehe põhja, koos olemasolevate komponentidega, ja sinna juurde lisada loodava laienduse kuvad. Lisaks

olid selle toote tasuta võimalused disaini loomiseks täiesti piisavad, seega lõputöö raames ei pidanud eraldi tellimust tegema.

Süsteemi kasutajaliidese disainimisel on lähtutud eesmärgist pakkuda juhtkonnale ja AIRE-le kiiret ning selget ülevaadet, minimeerides sealjuures kasutaja kognitiivset koormust. Prototüübi loomisel on rakendatud järgmisi põhimõtteid.

5.5.1 Kognitiivne koormus

Läbivaiks disainiprintsiibiks on kognitiivse koormuse minimeerimine. Inimese lühimälu suudab korraga hoida ligikaudu 7 ± 2 infoüksust [55], mistõttu näiteks süsteem pakub vaikimisi välja 5-7 kõige populaarsemat otsingumärksõna, mis jääb soovitud piiridesse. Lisaks on prototüübis välditud liigseid dekoratiivseid elemente, et kasutaja tähelepanu püsiks sisulistel seoste, mitte liidesel endal.

5.5.2 Otsingupõhine vs visuaalipõhine lähenemine

Disainis on kombineeritud kaks peamist infotarbimise meetodit. Otsingupõhine lähenemine sobib olukordades, kus kasutajal on konkreetne küsimus, näiteks soovib AIRE leida teadlasi kindla projektikirjelduse põhjal. Visuaalipõhine ehk avastuslik lähenemine on suunatud mustrite leidmisele, näiteks, kuidas on AI-kompetentsid aastate lõikes muutunud või millised uurimisgrupid on omavahel seotud. See on eelistatud juhtkonna vaates, kuna võimaldab tuvastada seoseid, mida kasutaja ei oska otsingusõnana sõnastada. Prototüübis on juhtkonnale mõlemad lähenemised koos: kasutaja võib alustada märksõnaotsingust ning seejärel liikuda graafivaatesse, et avastada seoseid laiemalt.

5.5.3 Graaf vs loend

Valik graafivaate ja loendi vahel on vaid juhtkonnal ning nende vahel otsustamine sõltub analüüsi eesmärgist. Loend on efektiivne siis, kui on vaja täpseid detaile või järjestusi, näiteks tulemuste sorteerimisel sobivuse järgi. Graaf on asendamatu relatsiooniliste andmemudelite puhul, kus oluline ei ole mitte objekti üksikud omadused, vaid objektide vahelised seosed. Prototüübis on leitud tasakaal, kus graaf pakub üldist struktuurset ülevaadet ning servadel klikkides esitatakse seoste detailid.

5.5.4 Filtreerimine ja navigatsioon

Filtreerimisvõimalus asub vasakul külgpaneelis, mis on veebirakenduste puhul levinud disainivalik, sarnast paigutust kasutavad näiteks Google Scholar ja e-poed. Selline konventsioon vähendab kasutaja õppimiskõverat, kuna liidese loogika on juba tuttav. Perioodi filter võimaldab lisada ajalise mõõtme, realiseerides andmemudeli *kehtiv_alates* ja *kehtiv_kuni* komponendid kasutajaliideses.

5.5.5 Värvikoodid ja visuaalne eristamine

Graafil on erinevad objektitüübid eristatud värvide abil, lähtudes põhimõttest, et värv on üks kiiremini töödeldavaid visuaalseid tunnuseid ning võimaldab kasutajal objektitüüpe eristada ilma teadliku pingutuseta. Värvide legend asub graafi all paremas servas, et mitte segada põhivaadet. Seose tugevust objektide vahel väljendab serva paksus, mis annab visuaalse signaali koostöö tihedusest ilma numbrilisi väärtusi kuvamata.

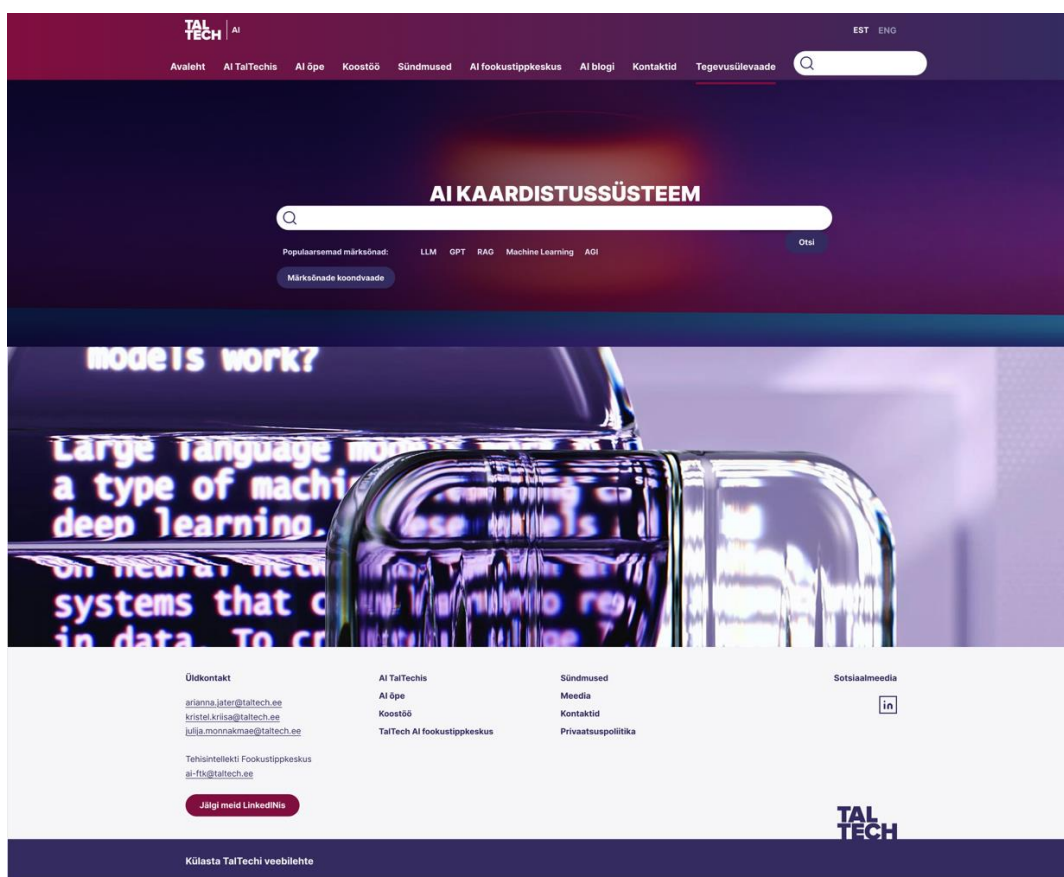
5.5.6 Vaikimisi vaade rolliti

Kasutajaliidese vaikimisi vaade on määratud kasutajarolli järgi, lähtudes intervjuudest saadud teadmistest. Juhtkonnale on vaikimisi kuvatud graafivaade, kuna nende peamine vajadus on mõista koostöövõrgustiku struktuuri tervikuna. AIRE-le on vaikimisi kuvatud loendivaade, kuna nende töövoog on suunatud konkreetsete sobivate vastete leidmisele sisestatud projektikirjeldusele. Selline rollipõhine kuvamine vähendab sammude arvu, mida kasutaja peab oma eesmärgi saavutamiseks tegema. Seda eelkõige juhtkonna kasutajale, kellel on valikus mitu kuva varianti.

5.6 Kasutajaliidese vaated ja navigatsioon

Esmalt loodi prototüübis TalTech AI veebilehele menüüsse uuele vahelehele navigeerimiseks menüünupp „Tegevusülevaade“, mille alla kogu antud analüüs koondub. „Tegevusülevaade“ on tööriist, mis lahendab kliendi poolt sisendiks antud probleemi, kus teadustööga seotud informatsioon on killustatud ja raskesti leitav, kuid nüüd on andmed kompaktsemalt kokku pandud ning visualiseeritud.

„Tegevusülevaade“ nupule vajutades avaneb kasutajale AI kaardistussüsteemi eelvaade, kus saab otsingut teostada, sisestades tekstikasti märksõna või tekstilõigu, või valida otsisõna hetkel populaarsete märksõnade seast (vt joonis 8). Populaarsed märksõnad on piiratud 5-7 valikule, mis jääb soovitatud kognitiivse koormuse piiridesse ning tagab, et kasutajal on alati käepärast praktilisi lähtepunkte ilma liigse valikuta. Juhtkonna kasutajal on veel eraldi kuvatud nupp „Märksõna koondvaade“, mis viib kasutaja Märksõnade vahelehele (vt joonis 9). AIRE-le seda nuppu ei kuvata, kuna märksõnade ajaline analüüs ei ole nende peamine kasutusjuht.

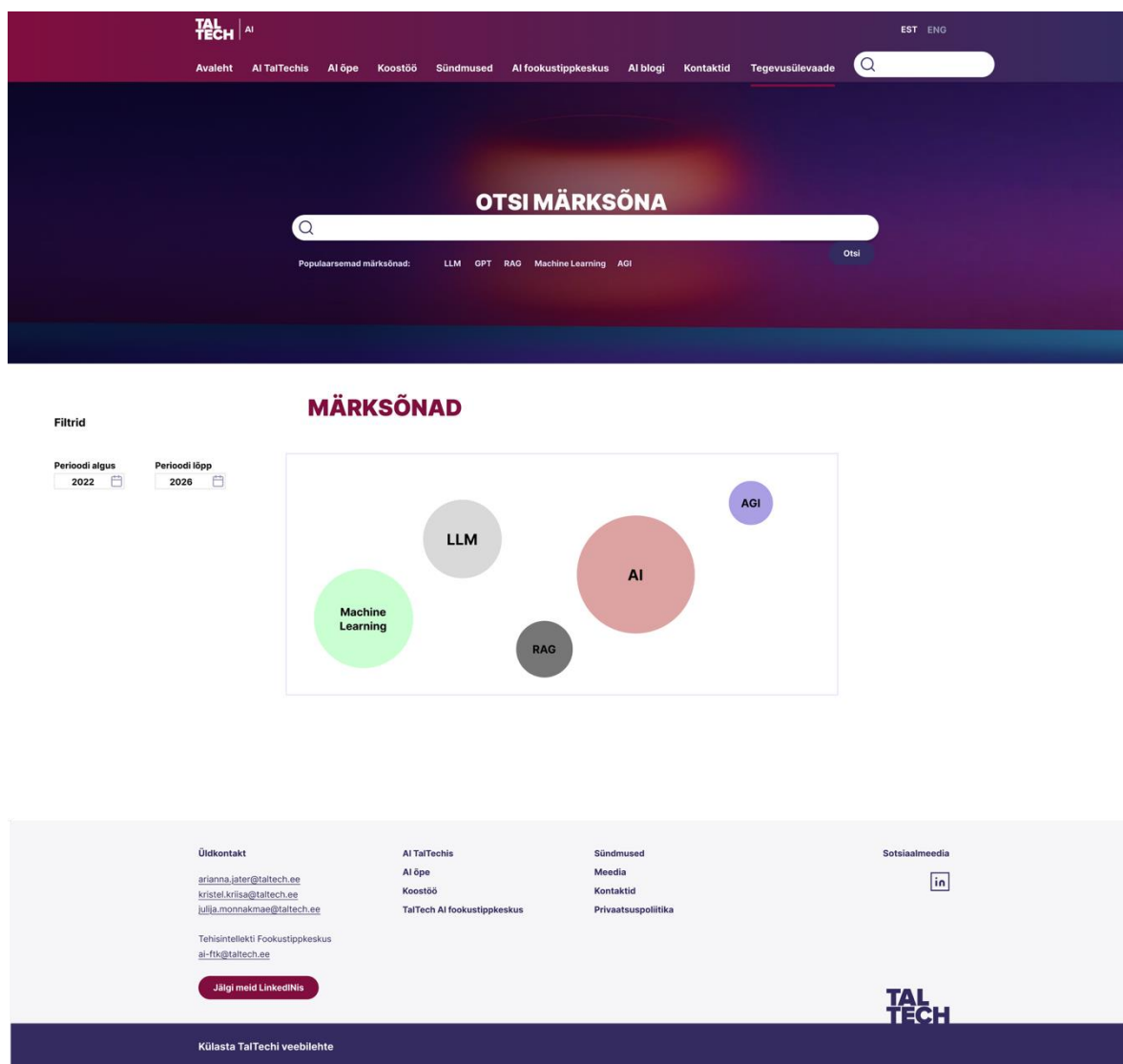


Joonis 8. Vaheleht "Tegevusülevaade".

5.6.1 Märksõnade koondvaade

Märksõnade koondvaates on kuvatud projektide ja publikatsioonide tulemuste koguarv mulldiagrammil (vt joonis 9). Mulldiagramm valiti, kuna see võimaldab korraga edastada kaks andmedimensiooni: märksõna populaarsus mulli suuruse kaudu ja märksõnade omavaheline temaatiline lähedus ruumilise paigutuse kaudu. Selline lahendus võimaldab juhtkonnale kiire visuaalse ülevaate ilma, et kasutaja peaks iga märksõna eraldi läbi

klikkima. Vaikimisi on tulemuseks filtreeritud viimase viie aasta projektid ja publikatsioonid, mis tagab, et kuvatud info oleks ajakohane ning vastab kiiresti areneva valdkonna vajadustele. Ilma mullidele klikkimata saab informatsiooni kätte visuaalselt nende suuruse ja paigutuse järgi.

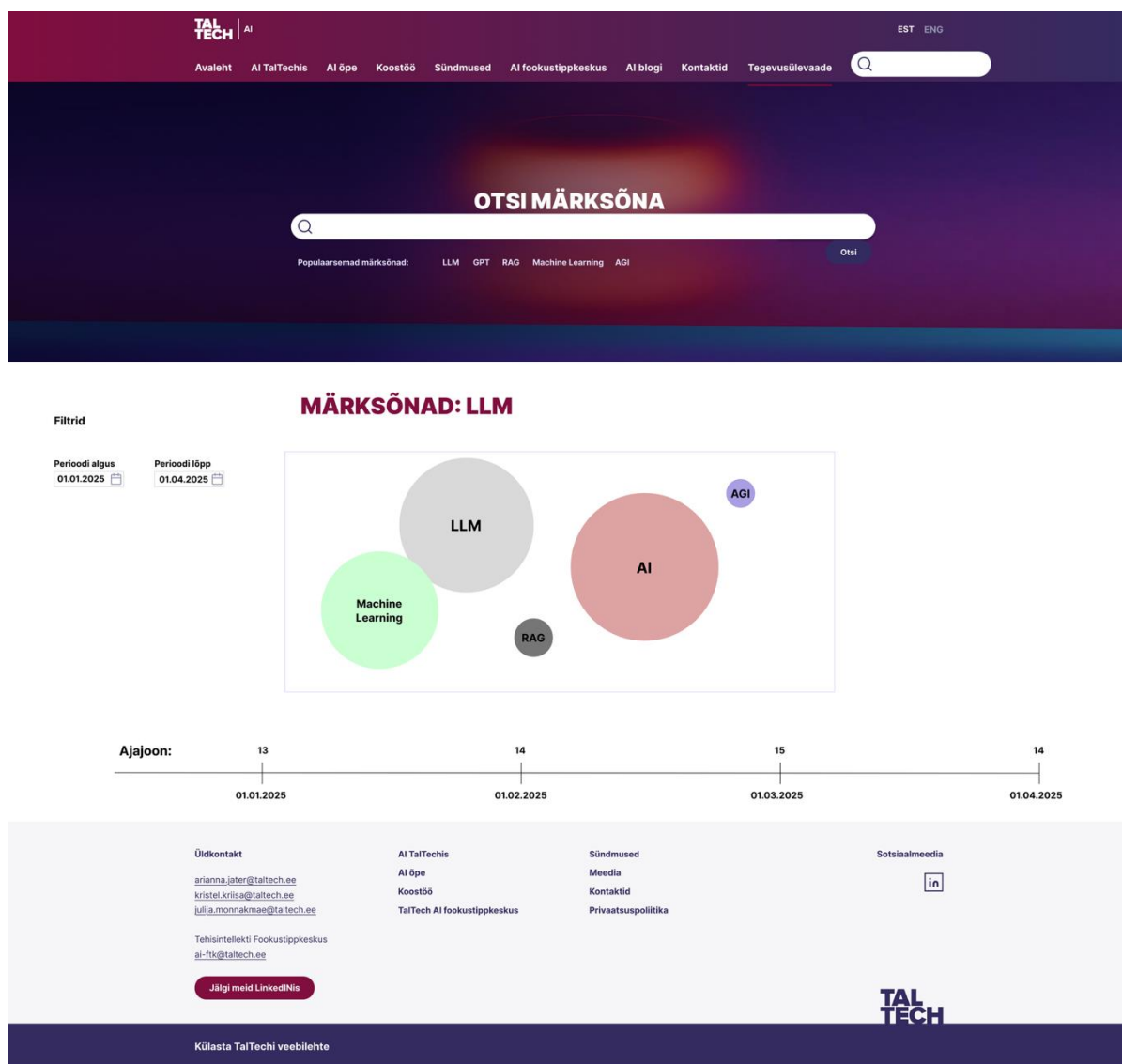


Joonis 9. Märksõna mulldiagramm.

5.6.2 Märksõna valik ja ajajoon

Samas saab ka mullidel teostada otsingut, vajutades ühe märksõna peale ning soovi korral perioodi muutes. Konkreetset märksõna mulli valides kuvatakse kasutajale märksõna kohta ajajoon, mis näitab perioodide kaupa ning arvuliselt, kui palju projekte ja publikatsioone selles vahemikus aktiivsed on (vt joonis 10). Numbrid on kahe kategooria

koondtulemus. Ajajoon diagrammi all võimaldab juhtkonnale kriitilise küsimuse vastamist: „kas mingi AI-valdkond on kasvamas või kahanemas?“. See toetab kasutajalugu 3 (AI valdkonna teaduslane areng) ning annab juhtkonnale sisendi strateegiliste otsuste tegemiseks, näiteks millistes valdkondades on vaja kompetentse juurde arendada või kus on tegevus märgatavalt vähenenud.

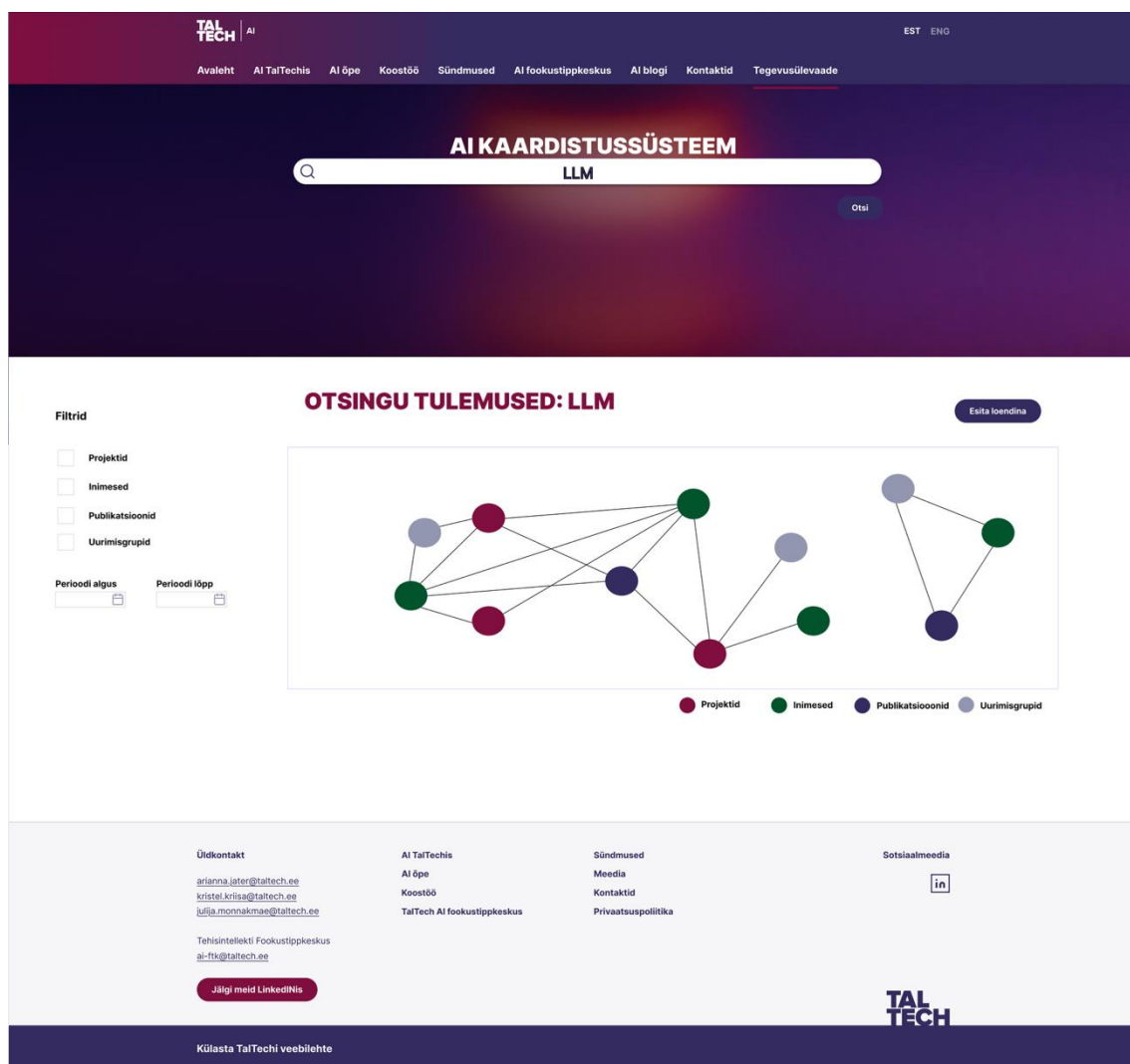


Joonis 10. Märksõna mulldiagrammi filtreeritud otsingutulemus.

5.6.3 Otsingutulemuste graafivaade

Teostades tavalist otsingut konkreetse märksõnaga, antud prototüübis „LLM“, avaneb juhtkonna kasutajale graafivaade (vt joonis 11). Graafivaade on vaikimisi vaade juhtkonnale, kuna see vastab nende peamisele vajadusele: mõista koostöövõrgustiku

struktuuri tervikuna, mitte ainult loetleda üksikuid tulemusi. See otsus tuleneb otseselt intervjuudest, kus juhtkond rõhutas kiire ja visuaalse ülevaate vajadust, ning realiseerib kasutajalugu 2 aktsepteerimiskriteeriumeid. Erinevad värvid tähistavad objektitüüpe (projektid, inimesed, publikatsioonid, uurimisgrupid) ja nende legend asub diagrammi all-paremal, et mitte segada põhivaadet. Seoste tugevust objektide vahel väljendab serva paksus, mis annab visuaalse signaali koostöö tihedusest ilma numbrilist väärtust kuvamata.



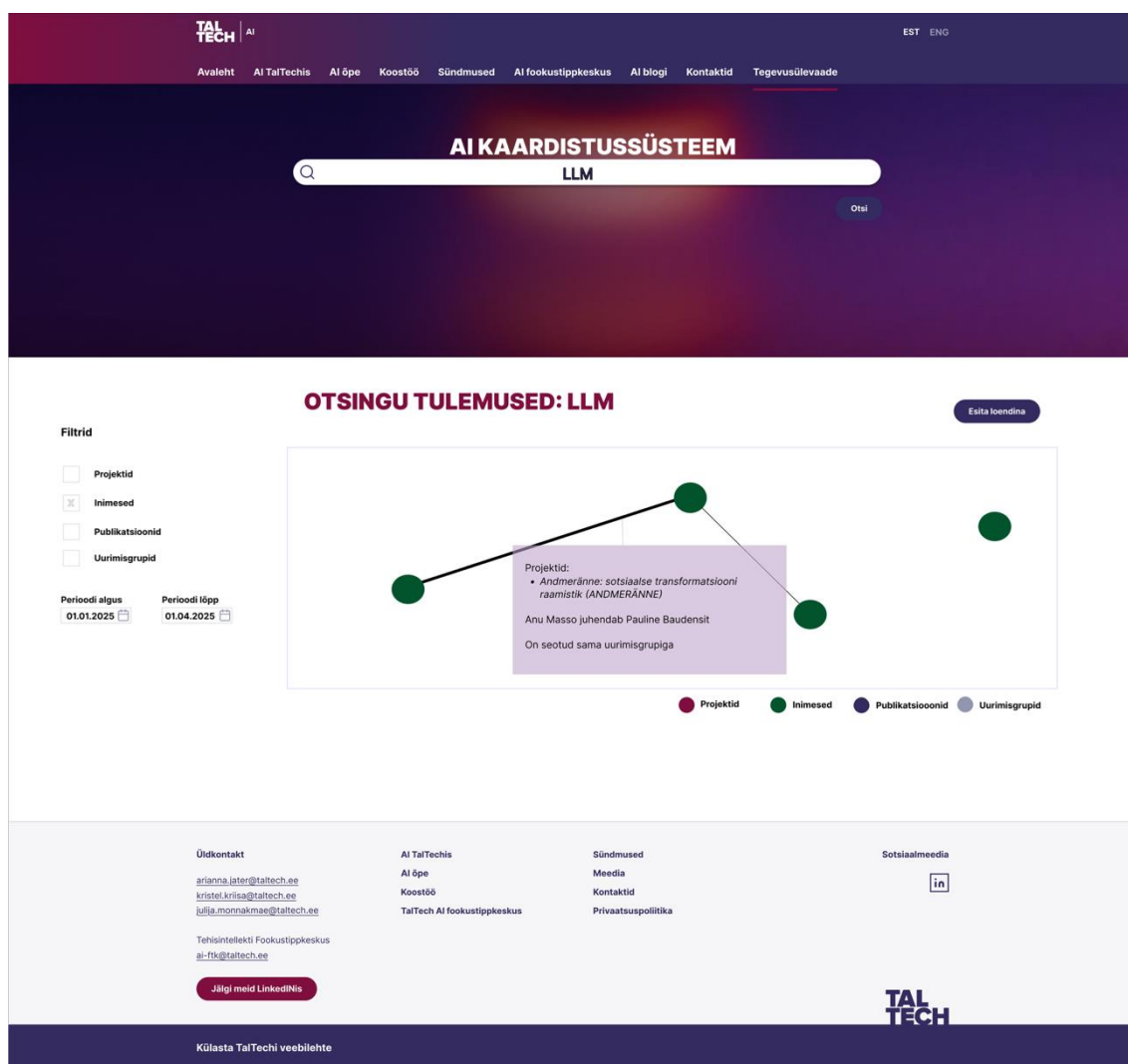
Joonis 11. Otsingutulemus graafil.

5.6.4 Graafivaate filtreerimine ja serval klikkimine

Kasutajal on võimalik tulemust filtreerida, valides tulemustest konkreetse kategooria või defineerides perioodi. Filtreerimisvõimalus asub vasakul külgpaneelis, mis on

veebirakenduste puhul levinud disainivalik vähendades õppimiskõverat. Perioodi filter võimaldab juhtkonnale ajalise mõõtmise lisamist, näha, millised seosed olid aktiivsed konkreetsel aastal, mis realiseerib andmemudeli ajalise komponendi (kehtib_alates, kehtib_kuni) kasutajaliideses.

Serval klikkimine avab seose kirjelduse, mis vastab kasutajaloole 2.3 (vt joonis 12). Kirjelduses on välja toodud kõikvõimalikud seosed, mis kahe objekti vahel on, näiteks juhendamised, koostööd ja ühte uurimisgruppi kuulumised. Selline lahendus võimaldab kasutajal kiiresti mõista koostöö olemust ilma süsteemis täiendavaid otsinguid tegemata, mis on eriti oluline juhtkonnale, kes vajab kiiret situatsioonipilti.



Joonis 12. Graafi filtreerimine.

5.6.5 Otsingutulemuste loendivaade

Soovi korral saab juhtkonna kasutaja navigeerida loendivaatesse (vt joonis 13), vajutades nuppu „Esita loendina“. Selle tulemusel avanev loend on filtreerimata, sest filtreerimisvõimalust loendis ei ole, kuvatud on kõige algelisemas vormis tulemused sisestatud otsisõnale.

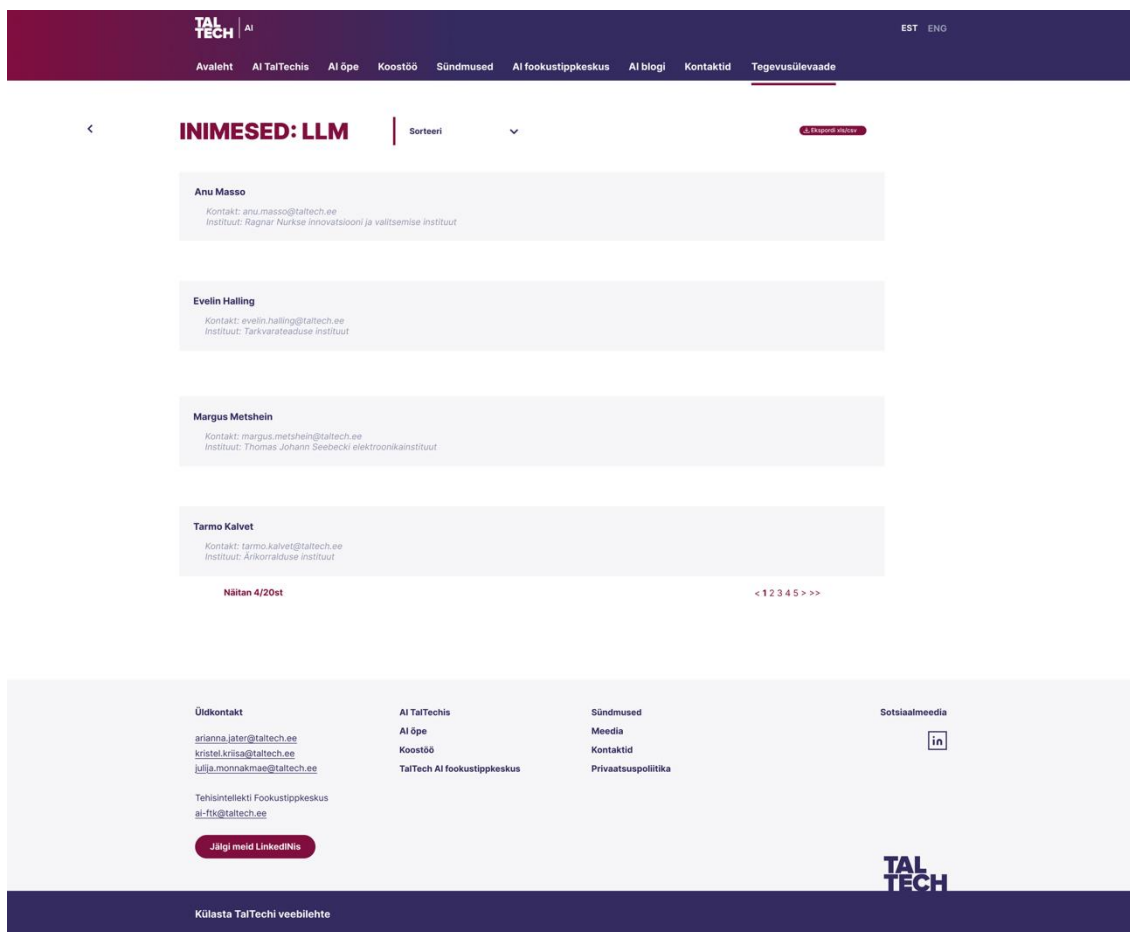
Loendivaade on AIRE-le vaikimisi vaade, kuna nende töövoog on suunatud konkreetsete sobivate vastete leidmisele sisestatud projektikirjeldusele ning loend toetab seda eesmärki paremini kui graaf. Otsingu tulemus on loendina kuvatud nii juhtkonnale kui AIRE-le. Loendis tulevad tulemused kategooriate kaupa, mida saab soovi korral lahti või kinni vajutada, prototüübis on kõik kategooriad avatud vormis. Igale kategooriale on kuvatud tulemuses selle nimi, nimetus või pealkiri (edaspidi kokkuvõtvalt nimi) ning tähtsam informatsioon. Siit saab avada soovitud tulemuse detailvaate, vajutades selle nime peale. Kui tulemusridu on rohkem, on kasutajale kuvatud nupp „Vaata kõiki“, mis viib vastava kategooria loendivaatesse. Tulemuste arvu piiramine ühel lehel toetab kognitiivse minimeerimise põhimõtet, mis on kirjeldatud peatükis 5.5.1.

Selline lahendus annab kasutajale võimaluse näha kõikide kategooriate tulemusi korraga ilma, et ta peaks iga kategooria jaoks sooritama eraldi vahelehel otsingu, nagu see toimib ETIS-es. See võimaldab tabada kõige teadlikumaid inimesi, kui nad ilmuvad loendites esimestena mitme kategooria alt, mis omakorda on suureks abiks AIRE-le, kes otsib oma klientidele potentsiaalseid koostöökaaslasi, toetades kasutajalugu 1.

5.6.6 Kategooriaspetsiifilised nimekirjad

Kategooriate nimekirjades (vt joonis 14), kuhu saab minna klikates „Vaata kõiki“ nupule otsingu tulemuste juures, on kuvatud tulemused samasuguse ülevaatliku informatsiooniga nagu on otsingust saadud loendis (vt joonis 13). Kategooriaspetsiifiliselt on võimalik tulemusi sorteerida sobivuse, uudsuse ning tähestiku järgi. Vaikimisi on tulemused sorteeritud sobivuse järgi, kõige sobivamad eespool, kuna AIRE peamine vajadus on leida sisestatud tekstile kõige asjakohasem vaste ning sobivuse järgi sorteerimine toetab seda eesmärki kõige otsesemalt. Tähestiku järgi sorteerimine on lisatud juhuks, kui kasutajal on vaja leida loendist konkreetset nime. Uudsuse järgi sorteerimine võimaldab leida kõige hiljutisemalt avaldatud projekte ja publikatsioone, mis on kasulik juhtkonnale, kes soovib ülevaadet aktiivsetest uurimissuundadest. Rohkemate tulemuste puhul on kasutajale kuvatud osa kõigist, mille kogus on näidatud loendi lõpus ning seal saab kuvamise arvukust vastavalt soovile muuta. Samuti saab lehtede vahel liikuda vajutades leheküljenumbritele või liikudes nooltega. Vastete arvu ühel lehel on piiratud, et ei tekiks info üleküllust.

Juhtkonnal on lisaks võimalus tulemused eksportida CSV või XLS formaadis vajutades nuppu „Ekspordi XLS/CSV“. Need formaadid valiti, kuna need on universaalselt ühilduvad Exceli ja teiste analüüsitööriistadega, mida juhtkond tõenäoliselt juba igapäevaselt kasutab, ning ei eelda kasutajalt täiendavate tarkvarade paigaldamist. See annab juhtkonnale võimaluse teha edasi analüüse juba valmis genereeritud andmefailide pealt, ilma et nad peaks vajalikku informatsiooni ise käsitsi kokku koguma, toetades kasutajalugu 6.



Joonis 14. Inimeste nimekiri.

Tulemuste kategooriaid on süsteemis neli, seega on sama palju ka detailvaateid: projekt (vt joonis 15), inimene (vt joonis 16), uurimisgrupp (vt joonis 17) ning publikatsioon (vt joonis 18). Läbivad väljad, mis igal detailvaatel on, on „Sobivus otsinguga“, „Seotud märksõnad“ ning „Allikad“. Mõistagi on sobivus ja seotud märksõnad olulised, et kuvada vastava objekti kattuvust otsingusse sisestatud tekstiga. Allikad on kuvatud, et kasutaja saaks soovi korral minna algallika juurde ning lisainformatsiooni otsida. Muud väljad on valitud spetsiifiliselt objektide kohta, mis võiks olla otsingu puhul olulised.

5.6.7 Projekti detailvaade

Projekti detailvaates (vt joonis 15) on kuvatud väljad, mis tulenesid otseselt ETIS-e andmeväljade analüüsist peatükis 4.1.1: vastutav täitja, täitjad, instituut, projekti number, periood, sobivus, seotud märksõnad, publikatsioonid ning allikad. Väljad on valitud nii, et kasutaja saaks kiire ülevaate projekti olemusest ja osapooltest, ilma et peaks selleks ETIS-esse navigeerima. Samas on allikate väljal kuvatud otselink ETIS-esse ette nähtud

juhuks, kui kasutaja soovib lisainfot. Süsteem ei kopeeri ETIS-t, vaid koondab olulisima ning suunab vajaduse korral algallika juurde. Sobivusprotsent arvutatakse projekti tasandil lihtsustatud mudeli järgi, mis arvestab hübriidset sisulist sarnasust ja informatsiooni ajakohasust vastavalt peatükis 5.4.2 kirjeldatud põhimõtetele, kuna projektil puudub isikupõhine teaduslik mõjukuse näitaja.

The screenshot shows the project detail page for "ANDMERÄNNE: SOTSIAALSE TRANSFORMATSIOONI RAAMISTIK (ANDMERÄNNE)". The page is part of the TalTech AI website, with a navigation bar at the top containing links like "Avalaht", "AI TalTechis", "AI õpe", "Koostöö", "Sündmused", "AI fookustippkeskus", "AI blogi", "Kontaktid", and "Tegevusülevaade".

The project details are organized into two columns:

- Vastutav täitja:** Anu Masso
- Täitjad:** Pauline Baudens, Tayfun Kasapoğlu, Ode Järvi (with a link "Vaata kõiki")
- Instituut:** Ragnar Nurkse innovatsiooni ja valitsemise instituut
- Projekti number:** PRG3205
- Periood:** 01.01.2026–31.12.2030
- Sobivus otsinguga:** Kattuvad märksõnad: LLM. Sisestatud tekst kattub isiku kirjeldusega 30%* osas.
- Seotud märksõnad:** LLM, GPT, RAG
- Publikatsioonid:** In the blind spot of AI-based mobility: eye-tracking visual attention to people with reduced mobility in autonomous vehicles and the ethics of inclusive design
- Allikad:** https://www.etis.ee/Portal/Projects/Display/1fa15dd9-1aab-42ee-bd4b-6c391005d2ef

A note at the bottom right states: "* % on leitud matemaatilise valemi järgi".

The footer section includes contact information for the AI fookustippkeskus, a LinkedIn link, and the TalTech logo.

Joonis 15. Projekti detailvaade.

5.6.8 Inimese detailvaade

Inimeste detailvaade (vt joonis 16) on detailvaadetest kõige informatiivsem, kuna isik on süsteemi keskne objekt, millega kõik teised objektid on seotud. Kuvatud on kontaktandmed, ametikoht, instituut uurimisgrupp, projektid, publikatsioonid, sobivus otsinguga, seotud märksõnad ning allikad. Selline koondvaade lahendab intervjuudes tuvastatud probleemi, kus teadlaste kontaktinfo asub TalTechi veebilehel ning teadustegevuse info ETIS-es, kuid ükski allikas ei paku mõlemat koos. Kontaktandmete

ja ametikoha lisamine detailvaatesse tulenes otseselt juhtkonna tagasisidest, kes rõhutas, et teadlastega ühenduse võtmine peab olema võimalikult lihtne. Sobivusprotsent arvutatakse täismudeli järgi, mis arvestab hübriidset sisulist sarnasust, teadusliku kogemuse sügavust läbi publikatsioonide arvu ja h-indeksi ning informatsiooni värskust, nagu kirjeldatud peatükis 5.4.2.

ANU MASSO

Kontakt, amet
 25. jaanuar 1977
 anu.massos@taltech.ee
 Sotsiaalteaduslike suurandmete kaasprofessorite uurimisrühm

Instituut
 Ragnar Nurkse innovatsiooni ja välismise instituut

Uurimisgrupp
 AndmeLabor

Projektid
 Andmeränne: sotsiaalse transformatsiooni raamistik (ANDMERÄNNE)
 FINEST TWINS: Targa linna tippkeskuse loomine
 "FinEst Targa Linna tippkeskuse tegevuste elluviimine" - UrbanSplash pilootprojekt
[Vaata kõiki](#)

Publikatsioonid
 Comparing Digital, Data, and AI Literacy: A Narrative Review
 In the blind spot of AI-based mobility: eye-tracking visual attention to people with reduced mobility in autonomous vehicles and the ethics of inclusive design
 Arvuliste ja suurte andmete analüüsi eetika
[Vaata kõiki](#)

Sobivus otsinguga
 Kattuvad märksõnad: LLM
 Siestatud tekst kattub isiku kirjeldusega 30%* osas

Seotud märksõnad
[LLM](#) [GPT](#) [RAG](#)

Allikad: <https://www.etis.ee/portal/persons/display/174fdada-7fcb-4128-a438-820cf653a261/est>

* % on leitud matemaatilise valemi järgi

Üldkontakt
 arianna.jater@taltech.ee
 kristel.kriisa@taltech.ee
 julia.monakmaa@taltech.ee
 Tehisintellekti Fookustippkeskus
 ai-fik@taltech.ee
[Jälgi meid LinkedInis](#)

AI TalTechis
 AI õpe
 Koostöö
 TalTech AI fookustippkeskus

Sündmused
 Meedia
 Kontaktid
 Privaatsuspoliitika

Sotsiaalmeedia
[in](#)

TAL TECH

Külasta TalTechi veebilehte

Joonis 16. Inimese detailvaade.

5.6.9 Uurimisgrupi detailvaade

Uurimisgrupp on süsteemis ainulaadne objekt selles mõttes, et see info puudub ETIS-es täielikult ning on saadaval üksnes TalTechi veebilehelt. Detailvaates (vt joonis 17) on kuvatud uurimisgrupi juht, liikmed, instituut, projektid, publikatsioonid, sobivus otsinguga, seotud märksõnad ning allikad. Juhi ja liikmete eristamine eraldi väljadena peegeldab andmemudeli Juht- ja Liige-seose erisust, mis on kirjeldatud 5.2.4 (vt joonis

3). Liikmete loetelu on piiratud, rohkemate kirjete korral on kuvatud nupp „Vaata kõiki“, et vältida info üleküllust. Sobivusprotsent arvutatakse grupi tasandil, koondades uurimisgrupi liikmete individuaalsed skoorid, nagu kirjeldatud peatükis 5.4.2.

TAL TECH AI EST ENG

Avalaht AI TalTechis AI õpe Koostöö Sündmused AI fookustippkeskus AI blogi Kontaktid Tegevusülevaade

ANDMELABOR

Uurimisgrupi juht Anu Masso

Instituut Ragnar Nurkse innovatsiooni ja valitsemise instituut

Projektid
 Andermänn: sotsiaalse transformatsiooni raamistik (ANDMERÄNNE)
 FINEST TWINS: Targa linna tippkeskuse loomine
 "FinEst Targa Linna tippkeskuse tegevuste elluviimine" - UrbanSplash pilotprojekt
[Vaata kõiki](#)

Publikatsioonid
 Comparing Digital, Data, and AI Literacy: A Narrative Review
 In the blind spot of AI-based mobility: eye-tracking visual attention to people with reduced mobility in autonomous vehicles and the ethics of inclusive design
 Arvuliste ja suurte andmete analüüsi eetika
[Vaata kõiki](#)

Uurimisgrupi liikmed
 Pauline Baudens
 Dr. Tayfun Kasapoglu
 Mergime Ibrahim
 Dr. Anniki Puura
 Jevgenia Gerassimenko
[Vaata kõiki](#)

Sobivus otsinguga
 Kattuvad märksõnad: LLM
 Sisestatud tekst kattub uurimisgrupi kirjeldusega 30%* osas

Seotud märksõnad
 LLM GPT RAG

Allikad: <https://taltech.ee/en/datalab>

* % on leitud matemaatilise valemi järgi

Üldkontakt
 arianna.jater@taltech.ee
 kristel.kriisa@taltech.ee
 julia.monnamae@taltech.ee
 Tehisintellekti Fookustippkeskus
 ai-ftk@taltech.ee
[Jälgi meid LinkedInis](#)

AI TalTechis
 AI õpe
 Koostöö
 TalTech AI fookustippkeskus

Sündmused
 Meedia
 Kontaktid
 Privaatsuspoliitika

Sotsiaalmeedia

TAL TECH

Külasta TalTechi veebilehte

Joonis 17. Uurimisgrupi detailvaade.

5.6.10 Publikatsiooni detailvaade

Publikatsiooni detailvaates (vt joonis 18) on kuvatud autorid, väljaanne, avaldamisaasta, DOI, sobivus otsinguga, seotud märksõnad ning allikad. DOI kuvamine otselingina annab kasutajale vahetu juurdepääsu publikatsiooni täistekstile, kuna süsteem ise tekste ei hoia. Autorite kuvamine linkidena võimaldab kasutajal navigeerida otse autori detailvaatele, luues nii loomuliku liikumistee objektide vahel ilma uut otsingut tegemata, kasutaja saab ühelt publikatsioonilt liikuda teadlase profiilile ning sealt edasi tema teistele töödele või uurimisgrupile. Sobivusprotsent arvutatakse lihtsustatud mudeli järgi, arvestades

hübriidset sisulist sarnasust ning informatsiooni ajakohasust, kuna publikatsiooni puudub isikupõhine teaduslik mõjukuse näitaja.

TAL TECH AI EST ENG

Avaleht AI TalTechis AI õpe Koostöö Sündmused AI fookustippkeskus AI blogi Kontaktid Tegevusülevaade

A HYBRID DEEP LEARNING AND LARGE LANGUAGE MODELS FRAMEWORK FOR SHIP COLLISION ACCIDENT ANALYSIS

Autorid	Jakub Jerzy Montewka, Xiyu Zhang, Langxiong Gan, Hailong Cui, Yaqing Shu, Zaili Yang	Väljaandisaasta	2026
Juhendaja	---	Liik	Artikkel
Väljaanne	Reliability Engineering & System Safety	DOI	https://doi.org/10.1016/j.ress.2026.112333
Sobivus otsinguga	Kattuvad märksõnad: LLM Sisestatud tekst kattub isiku kirjeldusega 30%* osas	Seotud märksõnad	Accident analysis Artificial intelligence Deep learning Large language model Ship collision accident
Allikad: https://www.etis.ee/Portal/Publications/Display/507d6ea5-2d6c-4577-8c44-51bead9f5ce1		* % on leitud matemaatilise valemiga järgi	

Üldkontakt: arianna.jater@taltech.ee, kristel.krisa@taltech.ee, julija.monakmaa@taltech.ee
Tehisintellekti fookustippkeskus: ai-fik@taltech.ee
Jälgige meid LinkedInis

AI TalTechis: AI õpe, Koostöö, TalTech AI fookustippkeskus

Sündmused: Meedia, Kontaktid, Privaatsooluspoliitika

Sotsiaalmeedia: in

Külalaste TalTechi veebilehte

TAL TECH

Joonis 18. Publikatsiooni detailvaade.

5.7 Projekti tulevik ja edasised sammud

Käesolev töö keskendus AI-kompetentside suhetepõhise kaardistussüsteemi analüüsile ja kasutajaliidese disainile. Kuna kasutajaliideses on eristatud rollipõhiseid vaateid, tuleb TalTech AI veebilehele luua autentimis- ja õiguste haldussüsteem, mis võimaldaks rakendada rollipõhist ligipääsuhaldust.

Tehnilise edasiarendusena tuleb realiseerida andmete kogumise ja uuendamise loogika. Võimalusel tuleks eelistada API-põhist andmevahetust, kuna see on veebikraapimisest stabiilsem ja paremini hooldatav. Samuti tuleb veebikraapimise puhul arvestada juriidiliste piirangutega.

Hetkel on lahendus loodud kitsale sihtrühmale vaid TalTechi AI-valdkonnas. Pilootprojekti edukuse korral on võimalik süsteemi laiendada ka teistele

teadusvaldkondadele ning täiendavatele kasutajagruppidele, näiteks TalTechi teadlastele ja tudengitele. Potentsiaalselt võiks süsteemi kasutada ka ülikoolidevahelise koostöö toetamiseks ning väliste partnerite teaduskoostöö lihtsustamiseks.

Edasise arenduse käigus oleks oluline süsteemi testida pärisandmete ja reaalse kasutusstsenaariumitega, et hinnata otsingutulemuste kvaliteeti, kasutusmugavust ja süsteemi praktilist väärtust kasutajate jaoks.

Lisaks avastati töö käigus probleem TalTechi teenuste märksõnastamisega, mis raskendab AIRE töövoogu. Antud probleem jäi töö skoobist välja, kuid vajab kindlasti tähelepanu.

6 Kokkuvõte

Käesolev bakalaureusetöö analüüsis ja kavandas suhetepõhise AI-kompetentside kaardistussüsteemi Tallinna Tehnikaülikooli näitel. Töö lähtus probleemist, et TalTechi teadlaste, projektide, uurimisgruppide ja märksõnadega seotud informatsioon on hajutatud mitme andmeallika (ETIS-e, ülikooli veebilehe ja raamatukogu digikogu) vahel, mille tõttu puudub terviklik, suhete- ja võrgustikupõhine ülevaade ülikooli AI-kompetentsidest.

Töö teoreetilises osas analüüsiti olemasolevaid andmeallikaid ning hinnati nende tugevusi ja piiranguid. Selgus, et ETIS sobib küll üksikute objektide otsimiseks, kuid ei toeta märksõna- ega suhete põhist analüüsi. TalTechi veebileht pakub infot uurimisgruppide kohta, mis ETIS-es puudub, ning digikogu võimaldab märksõnapõhist otsingut magistri- ja doktoritöödest. Ükski allikas eraldi ei suuda anda terviklikku ülevaadet teadlaste, projektide ja uurimisgruppide vahelistest seostest.

CRM-süsteemide võrdlusanalüüs näitas, et suhete modelleerimise, ajaloo haldamise ja rollipõhise ligipääsu halduse põhimõtted on loodavale süsteemile hästi ülekantavad. Sobivaima ajaloolahendusena valiti aeglaselt muutuvate dimensioonide rea lisamise tehnika (SCD Type 2), mis säilitab atribuutide muutumise ajaloo täielikult. Rollipõhise ligipääsu halduse aluseks võeti NIST SP 800-53 standard, mis toetab vähemate õiguste printsiipi.

Graafianalüüsi meetodite uurimine kinnitas, et võrgustiku tihedus, koostöö klastrid, keskusmõõdikud ja isolatsioonide tuvastamine pakuvad TalTechi juhtkonnale strateegilist lisaväärtust, mida olemasolevad süsteemid ei võimalda.

Kolme intervjuu põhjal sidusrühmadega (TalTech AI esindajad ja AIRE) täpsustati süsteemi skoopt ja kasutajarollid kitsendati kahele: TalTechi juhtkond ja AIRE. Intervjuude käigus kujunesid ka põhilised nõuded: semantiline hübriidotsing, graafivaade, sobivusprotsent, märksõnade mulldiagramm ja andmete eksport.

Töö sünteetiline osa hõlmas loogilise andmemudeli loomist, mis modelleerib seitse põhiobjekti (isik, projekt, publikatsioon, uurimisgrupp, märksõna, asutus, instituut) koos ajaliste seostega. Otsinguloogikaks valiti SBERT-põhise semantilise vektoriseerimise ja BM25 tekstipõhise otsingu hübriid, millele lisati sobivusprotsendi arvutamise mudel, arvestades sisulist sarnasust, teadusliku kogemuse sügavust ja informatsiooni värskust.

Kasutajaliidese prototüüp valmis Figma Design keskkonnas ning sisaldab kõiki olulisi vaateid: otsingu eelvaade, graafivaade filtreerimise ja serval klikkimisega, loendivaade sorteerimise ja ekspordiga, märksõnade mulldiagramm koos ajajoonega ning kõigi nelja objektitüübi detailvaated.

Töö peamiseks piiranguseks on see, et prototüüp põhineb näidisandmetel ning süsteemi tegelikku kasutajamugavust ja otsingutulemuste kvaliteeti ei ole pärisandmetega testitud. Edasise arenduse käigus on oluline realiseerida andmeimpordi loogika, autentimissüsteem ning viia läbi kasutajatestid tegelike sidusrühmadega. Eduka pilootprojekti korral on süsteem laiendatav teistele TalTechi teadusvaldkondadele ning potentsiaalselt ka teistele Eesti ülikoolidele.

Kasutatud kirjandus

- [1] Haridus- ja Teadusministeerium, Eesti Teadusinfosüsteem. [www] <https://www.etis.ee/>. Kasutatud: 15.05.2026.
- [2] Estonian Research Council, "Estonian Research Information System: Manual for Research Staff", Tartu: Estonian Research Council, 2015. [Online]. Loetud aadressil: https://www.etag.ee/wp-content/uploads/2012/05/ENG2_User_manual_long.pdf. Kasutatud: 01.05.2026
- [3] M. B. Aliyu, "Efficiency of Boolean Search strings for Information Retrieval," *Am. J. Eng. Res. AJER*, vol. 6, no. 11, pp. 216–222, 2017. [Online]. Loetud aadressil: https://www.researchgate.net/publication/321314341_Efficiency_of_Boolean_Search_strings_for_Information_Retrieval. Kasutatud: 13.05.2026
- [4] Tallinna Tehnikaülikool. *Struktuur*. [www] <https://taltech.ee/ulikool/struktuur>. Kasutatud: 15.05.2026.
- [5] TalTech raamatukogu digikogu. [www] <https://digikogu.taltech.ee/>. Kasutatud: 15.05.2026.
- [6] A. Möttus, "Grandiprojektide tulemuste kogumik 'Teadusrikas Eesti'," *Med24*, 2024. [Online]. Loetud aadressil: <https://www.med24.ee/uudised/kogumikust-teadusrikas-eesti-leiab-uelevaate-seitsmest-arsti-ja-terviseteaduslikust-uurimistoeest/254327>. Kasutatud: 16.05.2026.
- [7] Google Scholar. [www] <https://scholar.google.com>. Kasutatud: 16.05.2026.
- [8] R. Lillo, P. Acevedo, "Legal design and access to justice: a case study on child support payments in Chile," *TalTech J. Eur. Stud.*, vol. 15, no. 1, pp. 282–308, 2025. doi: 10.2478/bjes-2025-0014.

- [9] ELNET Konsortium. E-kataloog ESTER. [www] <https://www.elnet.ee/teenused/e-kataloog-ester/>. Kasutatud: 16.05.2026.
- [10] ELNET Konsortium. E-varamu. [www] <https://www.e-varamu.ee/et/>. Kasutatud: 15.05.2026.
- [11] F. Buttle, S. Maklan, *Customer Relationship Management: Concepts and Technologies*, 4th ed. New York: Routledge, 2019. doi: 10.4324/9781351016551.
- [12] E. W. T. Ngai, "Customer relationship management research (1992-2002): An academic literature review and classification", *Mark. Intell. Plan.*, vol. 23, no. 6, pp. 582–605, 2005. doi: 10.1108/02634500510624147.
- [13] M. H. Quzwen, R. Hurriyati, M. A. Sultan, "Customer Relationship Management in Business-to-Business Marketing: The Impact on Corporate Performance," in *5th Global Conference on Business, Management and Entrepreneurship (GCBME 2020)*, Bandung, Indonesia, 2021. doi: 10.2991/aebmr.k.210831.101.
- [14] A. Payne, *Handbook of CRM: Achieving excellence in customer management*. London: Routledge, 2006.
- [15] I. J. Chen, K. Popovich, "Understanding customer relationship management (CRM): People, process and technology," *Bus. Process Manag. J.*, vol. 9, no. 5, pp. 672–688, 2003. doi: 10.1108/14637150310496758.
- [16] W. Boulding, R. Staelin, M. Ehret, W. J. Johnston, "A Customer Relationship Management Roadmap: What is Known, Potential Pitfalls, and Where to Go," *J. Mark.*, vol. 69, no. 4, pp. 155–166, 2005. doi: 10.1509/jmkg.2005.69.4.155.
- [17] P. P. Chen, "The entity-relationship model — toward a unified view of data," *ACM Trans. Database Syst.*, vol. 1, no. 1, pp. 9–36, 1976. doi: 10.1145/320434.320440.
- [18] J. Xu, C. Lin, H. Yu, T. Feng, J. You, "ResearchArcade: Graph Interface for Academic Tasks," 2025. doi: 10.48550/arXiv.2511.22036.

- [19] M. Fowler, "Bitemporal history," martinowler.com. [Online]. Loetud aadressil: <https://martinowler.com/articles/bitemporal-history.html>. Kasutatud: 16.05.2026.
- [20] R. Kimball, M. Ross, *The data warehouse toolkit: the definitive guide to dimensional modeling*, 3rd ed. Indianapolis, IN: John Wiley & Sons, Inc, 2013.
- [21] Joint Task Force Interagency Working Group, "Security and Privacy Controls for Information Systems and Organizations," National Institute of Standards and Technology, 2020. doi: 10.6028/NIST.SP.800-53r5.
- [22] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, C. E. Youman, "Role-based access control models," *Computer*, vol. 29, no. 2, pp. 38–47, 1996. doi: 10.1109/2.485845.
- [23] M. Biehl, *API Architecture*. API-University Press, 2015.
- [24] M. A. Brown, A. Gruen, G. Maldoff, S. Messing, Z. Sanderson, M. Zimmer, "Web scraping for research: Legal, ethical, institutional, and scientific considerations," *Big Data Soc.*, vol. 12, no. 4, 2025. doi: 10.1177/20539517251381686.
- [25] M. E. J. Newman, *Networks*, 2nd ed. Oxford: Oxford University Press, 2018.
- [26] S. Wasserman, K. Faust, *Social Network Analysis: Methods and Applications*, 1st ed. Cambridge University Press, 1994. doi: 10.1017/CBO9780511815478.
- [27] E. Di Bella, L. Gandullia, S. Preti, "Analysis of scientific collaboration network of Italian Institute of Technology", *Scientometrics*, vol. 126, no. 10, pp. 8517–8539, 2021. doi: 10.1007/s11192-021-04120-9.
- [28] A. L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, T. Vicsek, "Evolution of the social network of scientific collaborations," *Phys. Stat. Mech. Its Appl.*, vol. 311, no. 3–4, pp. 590–614, 2002. doi: 10.1016/S0378-4371(02)00736-7.
- [29] S. P. Borgatti, M. G. Everett, "A Graph-theoretic perspective on centrality", *Soc. Netw.*, vol. 28, no. 4, pp. 466–484, 2006. doi: 10.1016/j.socnet.2005.11.005.
- [30] R. S. Burt, "Structural Holes and Good Ideas", *Am. J. Sociol.*, vol. 110, no. 2, pp. 349–399, 2004. doi: 10.1086/421787.

- [31] *html.to.design (by <div>RIOTS)*. [www] <https://www.figma.com/community/plugin/1159123024924461424/>. Kasutatud: 17.05.2026.
- [32] Figma, *Figma Design*. 2026. [www] <https://www.figma.com>. Kasutatud: 10.05.2026
- [33] Sparx Systems, *Enterprise Architect*. 2026. [www] <https://sparxsystems.com>. Kasutatud: 13.05.2026
- [34] Google, *Gemini 3 Flash*. 2026. [Online]. Loetud aadressil: <https://gemini.google.com>. Kasutatud: 15.05.2026
- [35] Anthropic, *Claude Sonnet 4.6*. 2026. [Online] <https://www.anthropic.com>. Kasutatud: 15.05.2026.
- [36] Adapter. *Meist*. [www] <https://adapter.ee/meist/>. Kasutatud: 15.05.2026.
- [37] Eesti Maaülikool. *Adapter*. [www] <https://www.emu.ee/adapter>. Kasutatud: 15.05.2026.
- [38] K. Praakli, "Intervjuu keeleandmete kogumise meetodina," *Eesti Raken. Ühin. Aastaraam. Est. Pap. Appl. Linguist.*, vol. 5, pp. 209–223, 2009. doi: 10.5128/ERYa5.14.
- [39] A. Osterwalder, *Value proposition design: how to create products and services customers want*. Hoboken: John Wiley & Sons, 2014.
- [40] N. Reimers, I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," 2019. doi: 10.48550/ARXIV.1908.10084.
- [41] S. Bagui, R. Earp, *Database design using entity-relationship diagrams*, 3rd ed. Milton: Auerbach Publishers, Inc, 2023.
- [42] "Projekt", *Sõnaveeb*. [www] <https://sonaveeb.ee/search/unif/dlall/dsall/projekt/1/est>. Kasutatud: 15.05.2026.

- [43] "Publikatsioon", *Sõnaveeb*. [www]
<https://sonaveeb.ee/search/unif/est/eki/publikatsioon/1/est>. Kasutatud: 15.05.2026.
- [44] "Uurimisgrupp", *Sõnaveeb*. [www]
<https://sonaveeb.ee/search/unif/est/dsall/uurimisgrupp/1/est>. Kasutatud: 15.05.2026.
- [45] E. Tsoukanara, G. Koloniari, E. Pitoura, "Skyline-based Exploration of Temporal Property Graphs," *Inf. Syst. Front.*, 2024. doi: 10.1007/s10796-024-10505-x.
- [46] Object Management Group, *Unified Modeling Language (UML)*, Version 2.5.1. 2017. [Online]. Loetud aadressil: <https://www.omg.org/spec/UML/2.5.1/>. Kasutatud: 09.05.2026
- [47] *ISO 8601: Date and Time Format*. International Organization for Standardization, 2019. [Online]. Loetud aadressil: <https://www.iso.org/iso-8601-date-and-time-format.html>. Kasutatud: 05.05.2026
- [48] J. Rumbaugh, I. Jacobson, G. Booch, *The unified modeling language reference manual: the definitive reference to the UML from the original designers*, 5th print. Reading, Mass.: Addison-Wesley, 2003.
- [49] A. Bindle, P. Singla, S. Sharma, A. Khakimov, R. I. Alkanhel, A. Muthanna, "A Hybrid Large Language Model for Context-Aware Document Ranking in Telecommunication Data," *IEEE Access*, vol. 13, pp. 120345–120359, 2025. doi: 10.1109/ACCESS.2025.3585637.
- [50] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proc. Natl. Acad. Sci.*, vol. 102, no. 46, pp. 16569–16572, 2005. doi: 10.1073/pnas.0507655102.
- [51] N. Craswell, D. Hawking, S. Robertson, "Effective site finding using link anchor information," in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, New Orleans, Louisiana, USA: ACM, 2001, pp. 250–257. doi: 10.1145/383952.383999.

- [52] B. Long, Y. Chang, *Relevance ranking for vertical search engines*. Amsterdam: Morgan Kaufmann, 2014.
- [53] K. E. Wiegers, *Software requirements*, 3rd ed. Redmond, WA: Microsoft Press, 2013.
- [54] C. Lee, *The art of crafting user stories: unleash creativity and collaboration to deliver high-value products with a delightful user experience*, 1st ed. Packt Publishing, 2023.
- [55] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychol. Rev.*, vol. 63, no. 2, pp. 81–97, 1956. doi: 10.1037/h0043158.

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

Meie, Kristin Salu ja Karoliine Tarjus

- 1 Anname Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „AI-kompetentside suhete põhise kaardistussüsteemi analüüs ja kasutajaliidese disain Tallinna Tehnikaülikooli näitel“, mille juhendaja Ago Luberg
 - 1.1 reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2 üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
- 2 Oleme teadlikud, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
- 3 Kinnitame, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

¹ Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal vastavalt üliõpilase taotlusele lõputööle juurdepääsupiirangu kehtestamiseks, mis on allkirjastatud teaduskonna dekaani poolt, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil. Kui lõputöö on loonud kaks või enam isikut oma ühise loomingulise tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtajaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. ja 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.

Lisa 2 – Poolstruktureeritud intervjuu küsimused

Käesolevas lisas on esitatud küsimustik, mida kasutati töö raames 17. märtsil 2026 läbi viidud poolstruktureeritud intervjuudes. Intervjuu eesmärgiks oli kaardistada esialgsed ootused ja nõuded loodavale AI-kompetentside kaardistamise süsteemile. Intervjuude tulemused on peatükis 3.3.3.

1. Kes on loodava lahenduse peamised kasutajad?
2. Milliste allikate vahel on info killustunud?
3. Kas Teil on visioon, kuidas andmeid visualiseerida?
4. Missugused võiks olla lõpplahenduse vaated?
5. Millised rollid ja õigused loodavas süsteemis oleks?
6. Milliseid otsinguid ja analüüse soovite teha, kuid mis hetkel killustatuse tõttu pole võimalik?
7. Kui tihti tuleks andmeid süsteemis uuendada?
8. Kas kompetentside arengut ajas on vaja kuvada?

Lisa 3 – Kristin Salu eneseanalüüs

Käesolevas peatükis annan ülevaate enda panusest lõputöösse ning hindan meeskonnatöö kulgu kogu lõputöö protsessi vältel.

Tööd alustasime kaasautoriga ühiselt. Algus nõudis kasutajavajaduste kaardistamist, seega viisime läbi esimese intervjuu ning seejärel saime süveneda töö kirjutamisse ning prototüübi loomisse.

Suurem osa peatükke on kirjutatud tihedas koostöös ja sisu läbi arutledes, kuid omamoodi tööjaotus meil siiski tekkis. Minu ülesandeks sai olemasolevate andmeallikate kaardistamine, kirjeldamine peatükis 2.1 ning andmeallikate hindamine peatükis 4.1. Loomulikuks edasiseks sammuks sai ka loogilise andmemudeli loomine. Objektid mõtlesime kaasautoriga koos, kuid objektide vaheliste seoste ja atribuutide loomine jäi minu ülesandeks. Samuti tegelesin mina Enterprise Architectis andmemudeli visuaalse kuvamisega ning andmemudeli kirjeldamisega peatükis 5.2.4.

Algse andmemudeli põhjal sai kaasautor alustada figma prototüübi loomist, kus mina alguses veidi kaasa aitasin. Kuna kaasautoril oli eelnev Figma kogemus olemas, siis üsna pea tekkis tööjaotus, kus kaasautor tegeles Figmaga ja mina parandasin seoseid andmemudelis ning lisasin puuduolevaid välju.

Võrdlusanalüüs, kus keskendusime CRM-süsteemide põhimõtetele, kujunes koostöös. Kaasautor andis ülevaate CRM-idest ning kuidas need suhteid modelleerivad. Minu ülesandeks jäi kirjeldada CRM-ide ajaloo ja õiguste haldamist. Võimalike CRM-põhimõtete ülekandmise võimaluste analüüsi koostasid mina.

Metoodika peatüki kirjutasime ühes koos ilma tööjaotuseta. Kuna olime mõlemad intervjuudel kohal, oli kõige kergem koos nendest ülevaade anda. Samuti kirjeldasime intervjuude põhjal kujunenud süsteeminõudeid ning seejärel kasutajavajaduste analüüsi.

Sünteesi peatükk jagunes eelnevalt kirjeldatud viisil, et mina tegelesin andmemudeliga ning kaasautor kasutajaliidesega. Rollid, õigused ja objektid kirjeldasime koos, et ei tekiks ebakõlasid. Samamoodi talitasime nõuete ja kasutuslugude puhul. Otsingupeatükis

keskendusin semantilisele vektoriseerimisele ning SBERT ja BM25 algoritmi kasutamisele töös ja andsin ülevaate, kuidas seda matemaatilist mudelit teistele objektidele laiendada.

Käesoleva töö kõige keeruliamaks osaks osutus andmemudeli loomine. Andmemudeli valmimine ei olnud sujuv protsess ning parandusi ja täiustusi tuli selles teha peaaegu kuni töö esitamiseni. Üsna aeganõudev oli analüüsimine, milliseid seoseid tuleb näidata ajakomponendiga. Samuti oli segane, mis seosed on allikates olemas. Näiteks algselt jäi mulje, et ETIS-es pole publikatsioonide ja projektide vahel seost, sest ETIS-t analüüsides sattusid ette ainult ilma seoseta projektid. Selliseid olukordi esines töö jooksul korduvalt, kus avastasin, et seos on kas puudu või hoopiski, et seost pole võimalik kuvada, sest infot pole. Nendega toime tulemine nõudis vahel kaasautori abi ning ideid, aga iga probleem sai lahendatud ning edaspidi on mul rohkem julgust abi küsida kui varem.

Üheks segadust tekitavaks aspektiks käesoleva töö juures oli struktuur. Peatükkide ümber tõstmist ja ümber struktureerimist oli mitmeid kordi. Oli palju arusaamatust, kus teatud alapeatükk asuma peaks. Lõpuks oli abiks juhendi lugemine ning iga peatüki puhul mõtestamine, mis infot see sisaldab ja kas see on ülevaade teemast või meie järeldused ja analüüs.

Minu meelest üheks töö huvitavamaks peatükiks on sobivusprotsendi loogika. Kuna see peatükk valmis tehisintellekti abiga, sain ka rakendada kriitilist mõtlemist, et mõista kas pakutud komponendid ka loodavasse mudelisse sobiksid. Selle jaoks tuli süveneda erinevatesse matemaatilistesse teadustekstidesse, mida ma varem sellisel määral teinud polnud. Seega oli see minu jaoks üks arendavamaid osi.

Lisa 4 – Karoliine Tarjus eneseanalüüs

Käesolevas peatükis annan ülevaate enda panusest lõputöösse ning hindan meeskonnatöö kulgu kogu protsessi vältel.

Koos kaasautoriga leidsime lõputöö teema Protsessorist, mille järel võtsime aja, et omaette teemaga tutvuda. Seejärel viisime koos läbi esimese intervjuu tellijatega, mis andis suuna, mille järel said meie rollid töö tegemisel üsna loomulikult kujunema hakata. Töötasime suures osas koos ühistes sessioonides, mis tegi asja lihtsamaks, sest küsimuste tekkimisel sai kohe mõtteid jagada. See oli eriti kasulik kohtades, kus meie töösad kattusid, näiteks nõuete ja kasutuslugude kirjutamisel, kus oli oluline, et andmemudeli ja kasutajaliidese otsused oleksid omavahel kooskõlas. Rollid ja õigused said samuti koos läbi arutatud, et ebakõlasid ei tekiks.

Minu põhivastutusalaks kujunes kasutajaliidese disain ning sellega seotud analüütiline töö. Kuna olin Figma keskkonnaga varasemalt kokku puutunud, olin valmis selle osa enda peale võtma, seega jäi prototüübi loomine minu teha, samal ajal kui kaasautor süvenes andmemudeli loomisesse. See tööjaotus ei olnud algusest peale kokku lepitud, vaid tekkis orgaaniliselt vastavalt meie tugevatele külgedele.

Teoreetilise osa raames kujunes minu vastutusalaks CRM-süsteemide sissejuhatav ülevaade ning suhete modelleerimise kirjeldamine peatükkides 2.2.1 ja 2.2.2. Graafianalüüsi peatükk 2.3 valmis samuti minu poolt, kuna graafipõhine lähenemine oli tihedalt seotud sellega, kuidas kavandsin tulemuste visuaalset kuvamist prototüübis, seega oli loogiline, et sama inimene, kes disainis graafivaate, kirjutas ka selle teoreetilise aluse. Andmete kogumise viiside ja juriidiliste piirangute analüüs peatükis 3.2.2 oli samuti minu teha.

Sobivusprotsendi puhul oli minu roll komponentide kaardistamine. Defineerisin, milliseid tegureid peaks hindamisel arvestama, ning selle põhjal palusin tehisintellektilt abi matemaatilise mudeli koostamisel. See nõudis kriitilist mõtlemist, kuna pakutud komponente tuli hinnata selle järgi, kas need loodavasse süsteemi ka tegelikult sobivad. Oli põnev rakendada sellist lähenemist valdkonnas, kus mul varasem kogemus puudus.

Siiski oli prototüübi loomine töö aeganõudvamaid, aga ka huvitavamaid osi. Erinevalt tavapärasest disainiprotsessist, kus alustatakse tühjalt lehelt, tuli siin luua kasutajaliides, mis sobib orgaaniliselt olemasolevale TalTech AI veebilehele. Selleks konverteerisime olemasoleva veebilehe Figma plugina abil disainilauale, kuid kuna veebilehe struktuur on iseenesest kompleksne, kulus alguses päris palju aega selleks, et mõista, kuidas uusi elemente lisada nii, et kogu ülejäänud vaade paigast ei läheks. Sessioone, kus kaasautor kõrval istus ja sai kohe küsida, oli siin eriti vaja, sest andmemudeli otsused mõjutasid otseselt seda, mida kasutajaliideses kuvada tuli ning vastupidi. Lõpptulemuse üle oleme kaasautoriga pigem uhked ning see töö kinnitas minus huvi disaini ja selle täpsuse vastu.

Töö keerulisemaks küljeks osutus ootamatult struktuur. Teoreetilise tausta ja analüüsi vahel piiri tõmbamine ei olnud alati üheselt selge ning peatükkide ümberstruktureerimist esines töö vältel rohkem kui algselt ootas. Tagantjärele vaadates oli see siiski vajalik protsess, sest iga kord, kui midagi ümber tõstsime, sai töö loogika veidi selgemaks. Ja eks see ongi üks asi, mida sellisest tööst kaasa võtta. Struktuur ei teki iseenesest, selle nimel tuleb vaeva näha.