

"Tehisintellekti rakendamine riigikaitseterminoloogia arendamisel"

Eleri Aedmaa, Reet Hendrikson

Sandra Eiche, Esta Prangel, Mari-Liis Tikerperi

”Kava

- I. Taust: arenduse ajend ja projekti üldandmed
- II. Prototüübi meetod
- III. Prototüübi funktsionaalsus
- IV. Kahe uuringu lühitutvustus
 - Eesmärk ja meetod
 - Tulemused
- V. Kokkuvõtteks

I. Taust: arenduse ajend ja projekti üldandmed

Taust ja arenduse ajend

- 2004 – riigikaitse, sõjanduse, julgeolekupoliitika terminivara alalise arendamise algus (**Kaitseministeeriumi terminoloogiakomisjon**) + terminibaasi **Militerm** loomine
- 2017. a lõpp – vastutus valitsemisala terminitöö eest **Kaitseväge Akadeemiale**
 - Militerm alates 2020. a Sõnaveebis, eraldi vārav <https://sonaveeb.ee/ds/mil>
 - 2018. a algus – **sõjandusterminoloogia töörühma** loomine
 - al oktoobrist 2024 – **riigikaitseterminoloogia töörühm**
- Erialakeele arendamisel ja terminibaasi loomisel on üks ajamahukaid tööloike **alginfo leidmine mõistete ja terminite kohta** dokumentidest, digiväljaannetest, (elektroonilistest) keelekogudest jt **eriilmelistest andmehulkadest**.
- Alginfo = definitsioonid, mõistetunnused, mõistet täpsustavad kontekstid, terminivariandid, võimalikud sünonüümid + allikaviited → **terminiarutelude toormaterjal**

Projekti üldandmed

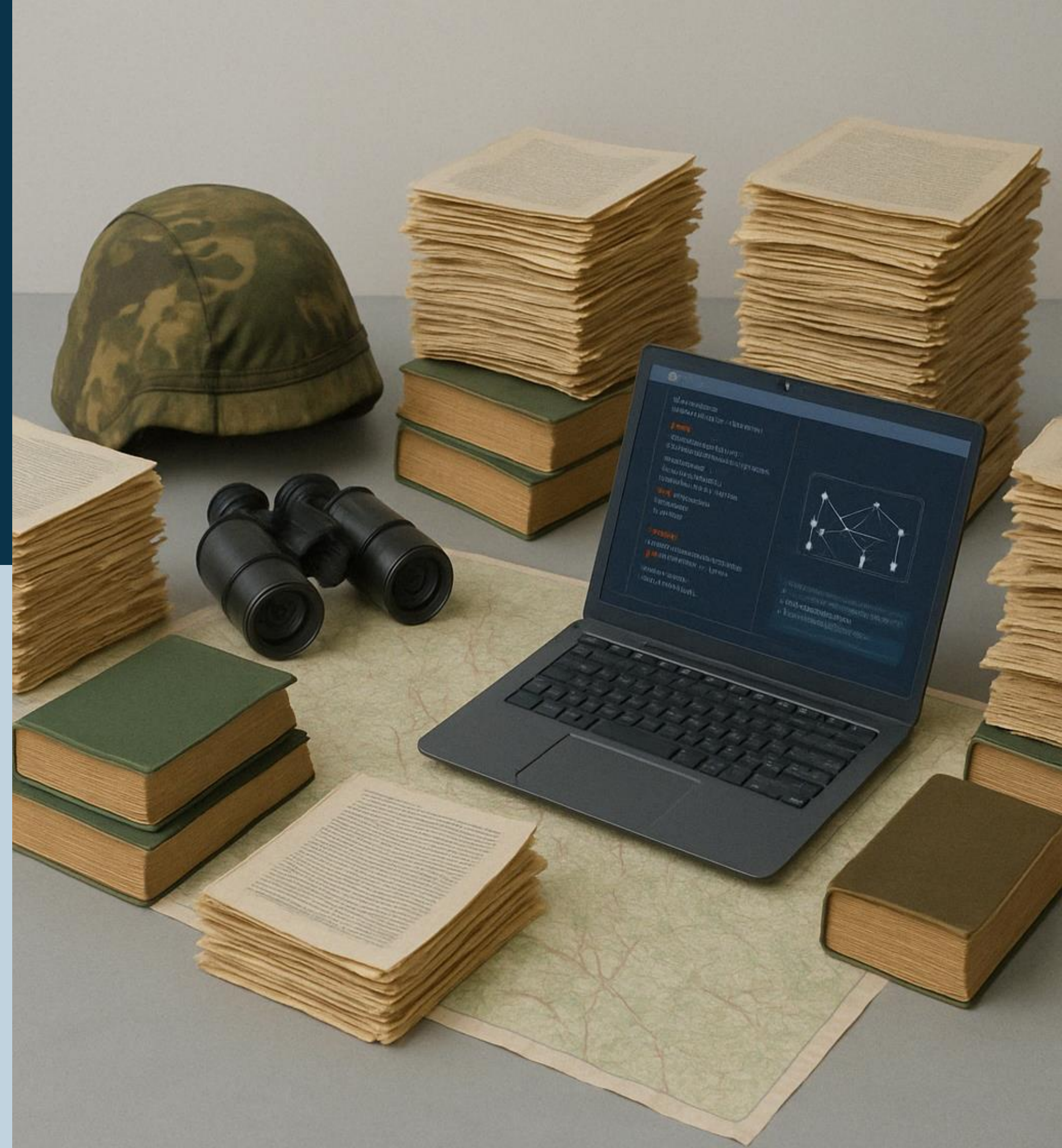
- Kestus 01.01.2024–31.12.2024/31.12.2025
- Pilootprojekti ellukutsuja ja koordinaator **KVA**, vastutav täitja ja elluviija **EKI**
- Eesmärk: "...välja töötada terminitööd toetava, tehisintellektil baseeruva tööriista prototüüp, et vabastada inimtööjõudu sisuliseks, loovamaks tööks ja suurendada nii töö tõhusust (korrastatud mõistekirjete hulka)."
- Seis 2024. a lõpuks ...
 - valmis tehisintellektil põhineva tööriista prototüüp, mida saab kasutada **kõigi erialakeelte arendamisel**; tulemus antud üle EKI-le
 - tehnoloogilise valmiduse tase (TVT) 4
- Järgmine siht TVT 6

II. Prototüübi meetod

RAG

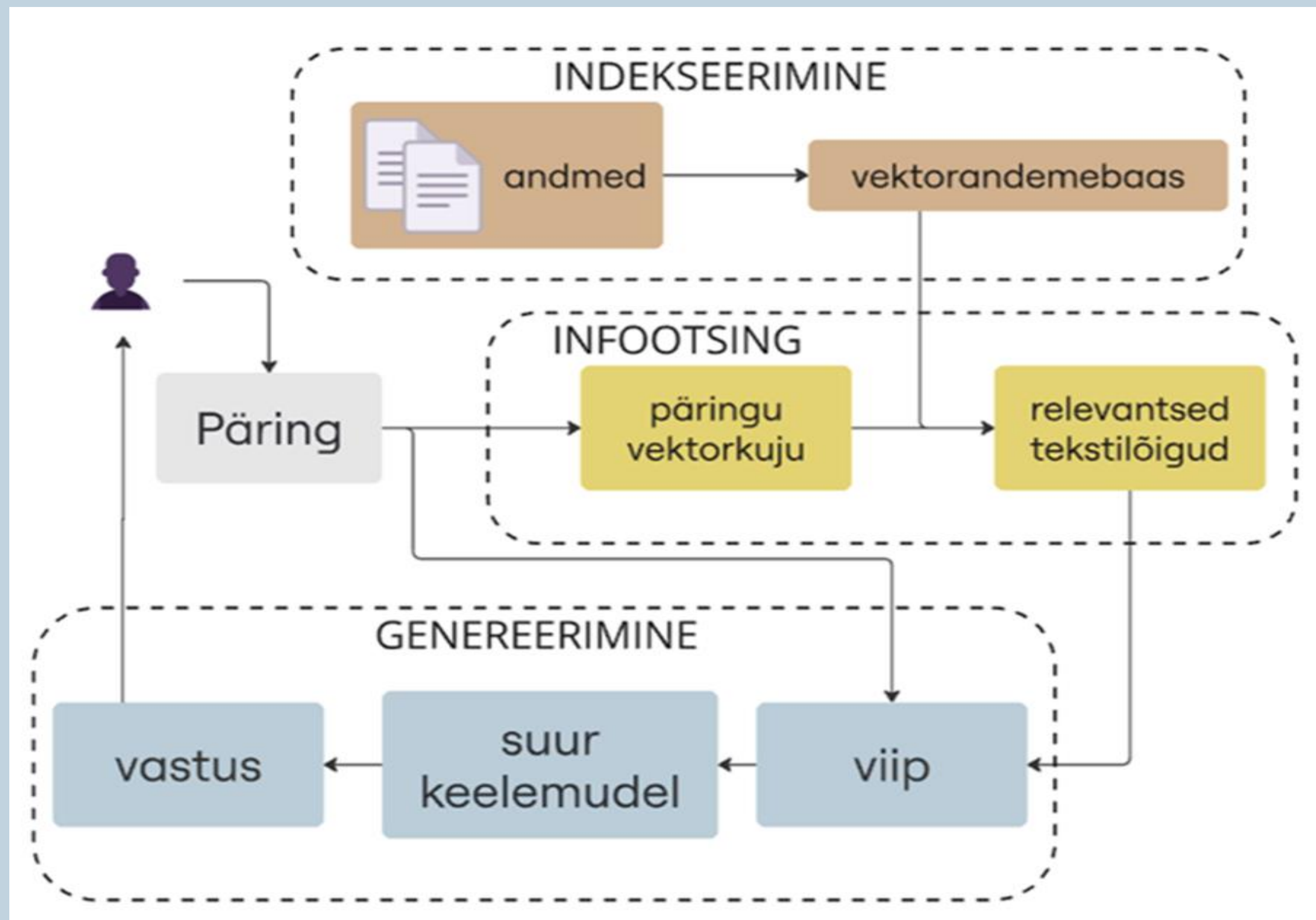
RAG (*Retrieval-Augmented Generation*),
eesti keeles *päringgenereerimine*,
välisloeline/allikapõhine genereerimine

Üldmudelit (eeltreenitud suurt keele-
mudelit) täiendatakse lisaandmetega,
mis võimaldavad sel genereerida
kvaliteetsemaid vastuseid.





RAGi etapid



III. Prototüübi funktsionaalsus

Valminud prototüübi iseloomustus

- Andmed: riigikaitsevaldkonna terminitöös kasutatavad ingliskeelsed dokumendid (STANAGid, ATPd, AJPd, FMid jpt)
- Suur keelemudel: GPT-4o
- Filtrid:
 - märksõnad (dokumentide metaandmetest),
 - dokumentide nimed,
 - mudelisse minevate relevantsete tekstilõikude arv.
- Sisaldab ka APIde (IATE, sõnastikud) ühispäringu ja dokumentide, metaandmete lisamise funktsioonid.

Dokumendiotsing

Päring: termin(ikandidaat)

Vastus:

- Toob andmetest välja sõnasõnalisi definitsioon(ikandidaat)e ja lisab allikaviited.
- Pakub välja seotud termineid, täpsustab tähendusseost ja lisab allikaviited.
- Toob välja kasutusnäiteid ja lisab allikaviited.
- Pakub välja, missuguseid teisi märksõnu/termineid veel uurida (mudeli n-ö omalooming).

Terminoloog

main effort

09:20

Assistent

Main Effort

Definitions:

1. "A concentration of forces or means in a particular area and at a particular time to enable a commander to bring about a decision." (AAP-06 2014, Page 137; AAP-06 Edition 2013, Page 131; AAP-06 2017, Page 92; AAP-06 2018, Page 101; AAP-06 2019, Page 105; ATP-3.2.1 EDC V1 E, Page 134; ATP-3.2.1 EdC V1 LAND TACTICAL OPERATIONS, Page 134)
2. "A designated subordinate unit whose mission at a given point in time is most critical to overall mission success. It is usually weighted with the preponderance of combat power." (ADP 3-0, Page 56; FM 5-0, C1, Page 20)
3. "The activity that the commander considers crucial to the success of his mission at that time." (ATP-3.2.2 (B) Command and Control of Allied Land Forces, Page 93)

Related Terms:

1. Priority of effort (Synonym, AAP-06 2014, Page 137; AAP-06 Edition 2013, Page 131; AAP-06 2017, Page 92; AAP-06 2018, Page 101; AAP-06 2019, Page 105)
2. Supporting effort (Related term, ADP 3-0, Page 56; FM 101-5-1 - OPERATIONAL TERMS AND GRAPHICS, Page 153)
3. Main attack (Related term, FM 101-5-1 - OPERATIONAL TERMS AND GRAPHICS, Page 100)
4. Decisive operation (Related term, ADP 3-90 OFFENSE AND DEFENSE, Page 35)

Usage evidence:

1. "Designating a main effort temporarily prioritizes resource allocation. When commanders designate a unit as the main effort, it receives priority of support and resources to maximize combat power. Commanders establish clear priorities of support, and they shift resources and priorities to the main effort as circumstances and the commander's intent require." (ADP 3-0, Page 56)
2. "Commanders must designate and appropriately allocate resources to the force assigned to the main effort, including all kinds of support to accomplish the mission. This is a sign of clear, determined leadership. No operation must be conducted without a dedicated main effort." (ATP-3.2.1 EdC V1 LAND TACTICAL OPERATIONS, Page 43)
3. "The main effort is a designated subordinate unit whose mission at a given point in time is most critical to overall mission success. The main effort is weighted with the preponderance of combat power—the total means of destructive, constructive, and information capabilities that a military unit or formation can apply at a given time." (FM 5-0, C1, Page 20)

See Also: Combat power; Commander's intent; Resource allocation; Tactical operations

Prompt

Role: You are a terminologist searching for terminological information about a keyword.

Objective: You've collected key sections from various documents about the keyword. Your task is to analyze these sections in order to create a comprehensive term entry. ALWAYS use EXACT QUOTES unless otherwise stated. You must remain focused on linguistic accuracy.

Instructions: Extract and organize terminological information as follows:

1. "Definitions"

Extract ALL definitions describing the term. Keep ORIGINAL WORDING and include all relevant information.

Keep source identifiers, if a definition is exactly the same in multiple documents, include all of the sources.

2. "Related terms and lexical relations"

From the passages try to identify:

- Synonyms - spelling or form variations of the keyword and precise synonyms
- Broader term - keyword is a type/part/subcategory of a broader term
- Narrower term - narrower term is a type/part/subcategory of the keyword
- Abbreviations
- Other related terms - frequently appearing in same context

3. "Usage evidence"

From the retrieved passages find paragraphs that contain the keyword and help to identify domain-specific applications or characteristic patterns of use of the term. The output paragraph must be coherent and must include the keyword. Keep original wording and source identifiers.

4. "See also"

Identify terms, abbreviations, or synonyms that may be useful for further exploration of the keyword. These don't necessarily have to be mentioned in key sections.

Output Format:

[INSERT TERM OF INTEREST]

Definitions:

1. Definition text (Document title, Page no)
2. Additional definitions as found. (Document title, Page no)

Related Terms

1. Related term (Relation type, Document title, Page no)

Usage evidence

1. Contextual paragraph text (Document title, Page no)

See Also: [Insert terms, abbreviations, or synonyms, separate terms using semicolons]

IV. Uuringute lühitutvustus

Uuring: eesmärk, meetod ja tulemused

Eesmärk ja meetod

- Eesmärk: välja selgitada, kui täpselt **eristab** süsteem **definitsioone ja allikaviiteid**
- **(Pool)automaatne hindamine**, kus võrdlesime infootsingu vastuseid genereeritud vastustega
- Keskmis küsimus, **kas ja kui palju suur keelemudel hallutsineerib** (ehk muudab teksti sisu või esitab sisu, mis ei vasta andmetele)

Tulemused

- ~ 90% juhtudest annab definitsiooni(kandidaadi) originaalsõnastuse
- ~ 99% juhtudest annab täpse allikaviite

II uuring: eesmärk ja meetod

- Eesmärk: hinnata prototüübi **kasutatavust ja funktsionaalsust terminoloogi tööriistana**
- Meetod: **küsimustik**, kus palusime ekspertidel hinnata prototüübi tulemi sisu (etteantud kuvad)
- Terminivalim: 31 keelendit, sh ...
 - (a) **lühendi-väljakirjutuse** paare (nt KIA ja *killed in action*; MEDEX ja *medical exercise*);
 - (b) **sõna-termini piiril** olevaid keelendeid (nt *movement*, *enclosure*);
 - (c) spetsiifilise **alamvaldkonna** termineid, mille kohta on dokumentides vähe infot (nt *combatant*);
 - (d) termineid, millele seni **pole leitud (asjakohast)** ingliskeelset **definitsiooni** (nt *will to defend*).
- Hindajate valim: **3+2** hindajat
- Saadud ja analüüsitud vastusteid **155**
- Metoodilisi õpituvastusi: (a) ilmselt **liiga suur terminivalim** (väsisas hindajaid); (b) paratamatu **arvamispõhisus**; (c) ette antud **kindlad kuvad**, mistap ei saanud vastaja mängida ridade arvuga; (d) vastajad **ei hinnanud väljundeid samas järjekorras**.

II uuring: tulemused (1/2)

Definitsioonid

- 130 (84%) juhul **pakub välja definitsiooni**. Neist 101 (78%) on sellised, millega **saab** küsitletute hinnangul **edasi minna**.
- 14% juhtudest **ei esita** süsteem **definitsiooni**, mis vastajate hinnangul peaks kindlasti andmetes leiduma.
- 20% juhtudest **pole andmetest mõtet definitsiooni otsida**, sest vastajate hinnangul seda pole andmetes.

Seotud terminid

- 59% juhul olid **kõik** pakutud terminid **asjakohased** ja 38% olid pakutud terminite seas **vähemalt üks asjakohane seotud termin**.
- Viiel korral polnud pakutud seotud terminite seas **mitte ühtegi asjakohast** terminit.
- **Asjakohatuteks** peeti termineid, kui need olid (a) otsitud terminiga **väga kaudselt** või **mitte üldse seotud**, (b) **prantsuse keeles**, (c) liiga **detailsed** või **üldised** või siis (d) polnud **terminiväärtusega**. Toodi välja ka **kõhutunnet**.
- 41% juhtudest vastati, et dokumentides on kindlasti rohkem seotud termineid ning need toodi ka vastustes välja. 13% kordadest arvati, et neid seal kindlasti ei ole. 46% kordadest ei osatud öelda.

II uuring: tulemused (2/2)

Seotud terminid

- 67% asjakohastest terminitest **korrektsete semantiliste suhetega**; 31% kohta leiti, et esitatud seoste seas on **nii õigeid kui ka valesid seoseid**.

Kasutusnäited

- 62% pakutud kontekstidest olid **täielikult**, 28% osaliselt **asjakohased**.
- Asjakohatuteks peeti kontekste, mis (a) **kitsendasid** mõistet, (b) esitasid definitsiooni väljal **juba esitatud infot**, (c) olid mõistega **liiga nõrgalt** või **mitte üldse seotud**.

Lisasoovitused („vaata ka“)

- 45% olid pakutud märksõnad **täielikult** ja 43% osaliselt **asjakohased ja informatiivsed**.
- 68% juhtudest arvati, et **pakutut võiks edasi uurida**.
- Plussid: (a) pakub viiteid **mõistepesa lähimõistetele**, mida seotud terminite osa ei pakkunud; (b) annab **uusi (tähendus)vihjeid**; (c) mõnel korral lisab täpseid allikaviiteid.
- Kriitikat: (a) liiga palju lühendeid, millel **pole lahtikirjutusi**, ja/või ei teatudki, mis on lühendi täiskuju; (b) vahel annab **kasutuid seoseid** (nt NATO, NATO standard).

II uuring: tähelepanekuid ja põnevamaid leide

- Süsteem **muutis** ise **otsisõna**: nt *combatant* -> *combatant commander* ja *combatant command*.
- Definiitsiooni(kandidaati)de puhul annab kohati **ammendava variatiivsesega tulemi**, nt *main effort* (leiab nii üksuse, tegevuse kui ka tegevuse tulemuse tähenduse).
- Kui promptile vastavat definiitsiooni(kandidaati) pole, **ei taha süsteem alati seda tunnistada** ja hakkab midagigi otsima. Nii tekivad ka **hallutsinatsioonid** (2 juhtu: MEDEX, CIMIC). Aga vahel tunnistab (*public safety*).
- Kohati on **allikate variatiivsus** väga hea (nt *direct support*, *main effort*), kohati jääb ühe-kahe allika keskseks (nt *combatant*).
- Esines kohti, kus relevantssed **seotud terminid** ilmusid alles kasutusnäidete või vaata-ka-plokis (nt *urban military operations* ja *urban combat* termini *urban warfare* juures).
- Nimetas pealtnäha ***mittesamaviitelist*** lühendit **samaviiteliseks** otsituga –> **osutas kahe mõiste hägusale piirile** (RSOM ja RSOMI: terminit ja lühendit on dokumentides muudetud, aga mõiste sisuliselt sama).
- Inimfaktor ehk **erialateadmised, kogemus, erialainimeste arutelud/vaidlused ja erialainimeste kõhutunne (!)** – neid aspekte tehismõistus ei asenda.

Subjektiivsus ja erialainimese kõhutunne

- "Saan aru, et sel üksusel (JLSG) [*joint logistics support group*] on oluline roll ja vastutus RSOM [*reception, staging and onward movement*] läbiviimisel, aga ma ütleks, et liiga kaugel seos leitud (**umbes sama kaugel nagu abivajaja evakuatsioon põlevast majast on seotud siseministeeriumiga**)."
- "POD [*port of debarkation*] on küll üks punkte/etappe, mida RSOM teekonnal läbitakse, aga neid on veel palju teisi – miks see siin seotud termin on... (jälle päästeoperatsiooni võrdluses – **abivajaja päästmine põlevast majast on tõesti seotud tuletõrje veevõtukohaga teekonnal päästesündmusele**)."

Kokkuvõtteks ...

<https://shorturl.at/ULZY4>

eki.ee

"Tehisaru näitab meile, et inimese tegelik elu on samavõrra unikaalne, variatiivne, emotsionaalne, teistsugune, otsiv, mittestatistilist rõhutav. Sest mis asi on keelemudel? See lähtub sellest, mis on statistiliselt tõenäosuslik, korduv, millest on võimalik teha uusi variante. /../ Kui me võtame tehisaru kui abimeest, nagu ehitustellingut, mis võimaldab ehitada järjest ja järjest kõrgemale, on meil ees väga põnevad ajad."

(Marju Lauristin, Sirp, 11. aprill 2025)

Uuri lisaks

Eiche, Sandra; Hendrikson, Reet; Aedmaa, Eleri; Prangel, Esta; Tikerperi Mari-Liis 2025. Tehisintellekti rakendamine erialakeele arendamisel riigikaitseterminoloogia näitel. – Eesti Rakenduslingvistika Ühingu aastaraamat = Estonian papers in applied linguistics. DOI: 10.5128/erya.1736-2563.



Github: <https://shorturl.at/9RVM9>