

EKILEXi arenduse hetkeseis

Arvi Tavast, Aksel Tavast

Mis toimub

- EKILEX: koostamise keskkond
- Sõnaveeb: sõnastikud lõppkasutajale
- TripleDev arendab, mina olen arendajapoolne tooteomanik
- 2017 august – 2018 lõpp
- Loodetav jätkuprojekt kuni 2020 lõpp
- EKILEXi kasutajaõiguste diferentseerimine
- Terminiveeb: terminibaasid lõppkasutajale

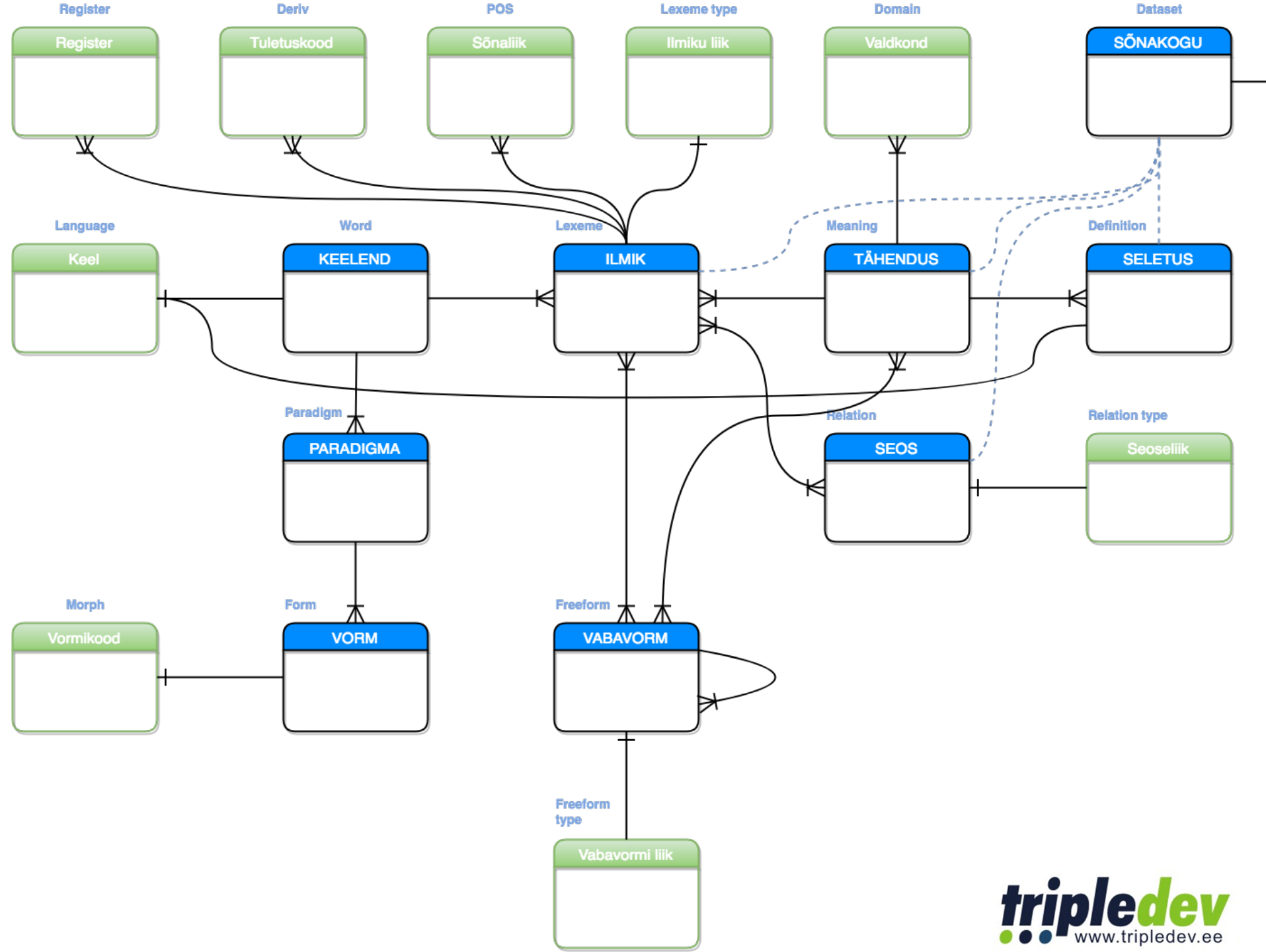
Mis meid endid rõõmustab

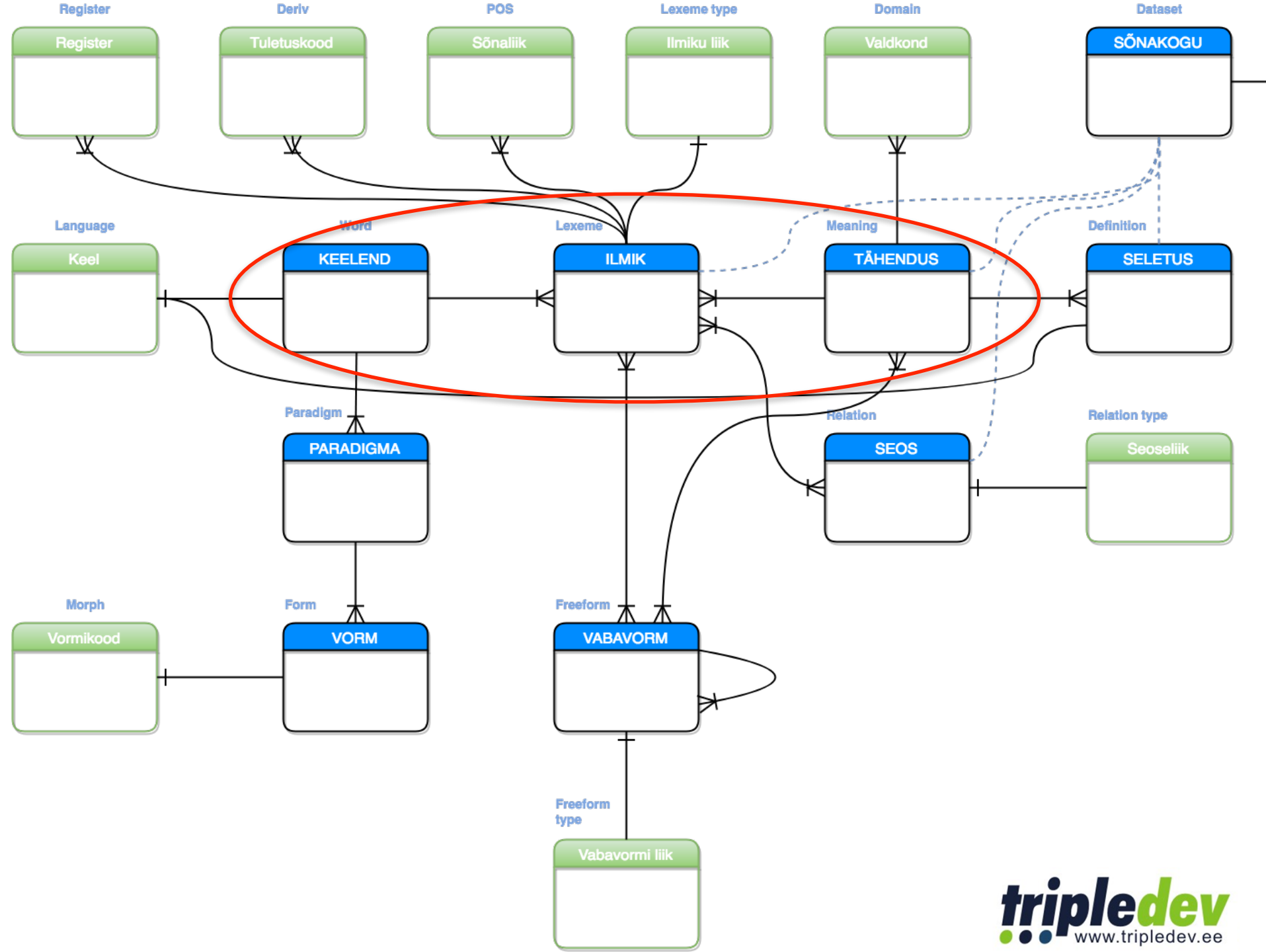
- Sõnastikud ja terminibaasid samas süsteemis
- Struktureeritud andmebaas (vs jutustus)
- Mudel nii lihtne kui võimalik
- Paberil (peaaegu) enam ei avaldata
- Korpusepõhine koostamine
- Otsepäringud korpuse
- Natuke kvantifitseerimist

Kellele

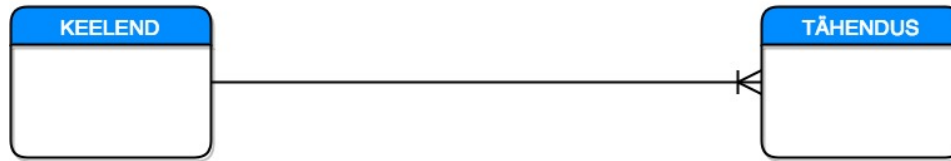
- Terminoloogid
- Leksikograafid

- Sama süsteem mõlemale? Mismoodi?



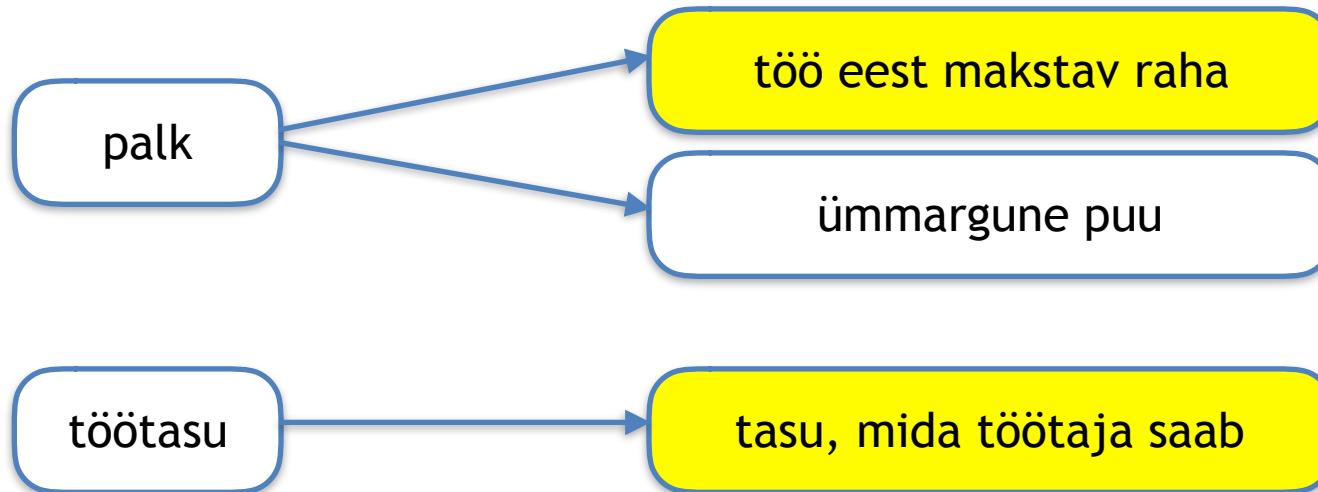


Leksikograafide soov

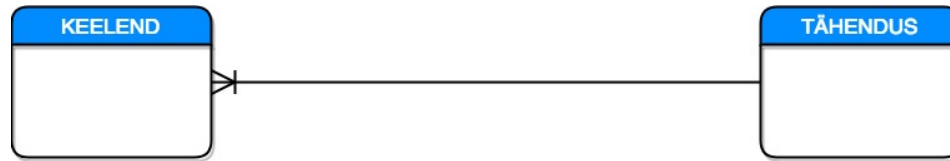


keelendil on mitu tähendust

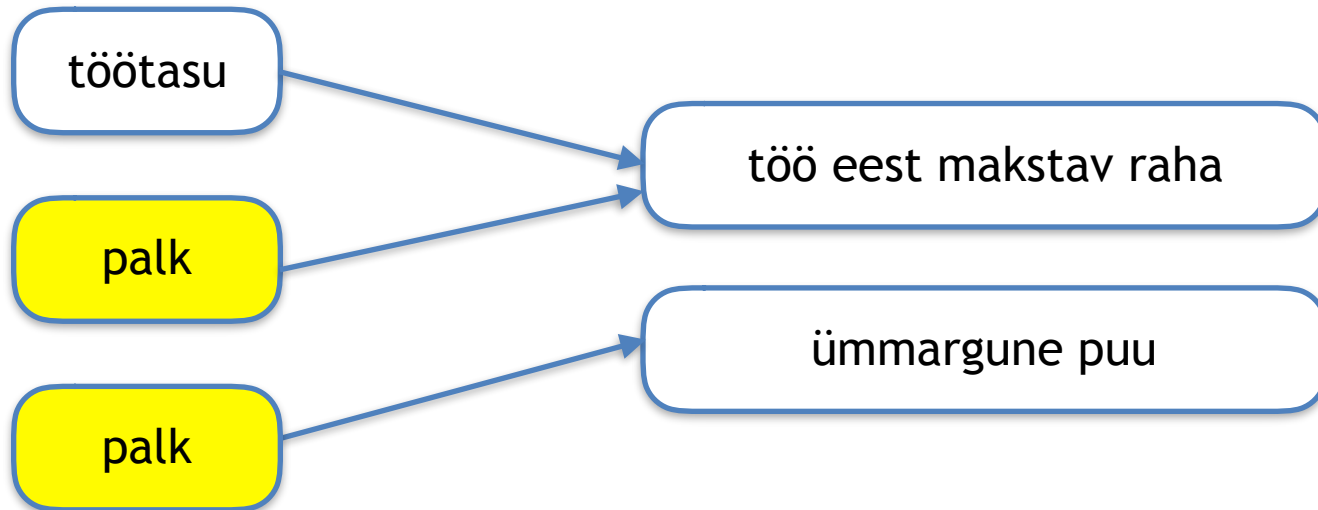
kokkulangevaid tähendusi on korratud



Terminoloogide soov



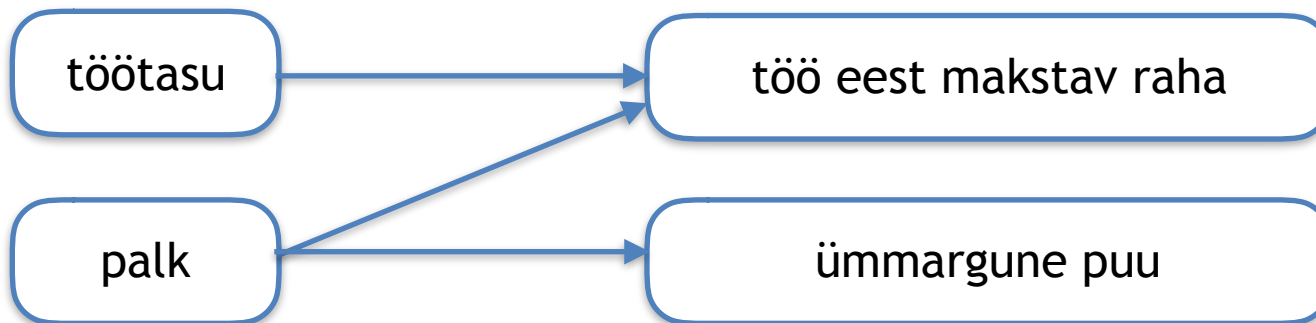
tähendusel on mitu keelendit
samakujulisi keelendeid on korratud



Ühine lahendus



mitu-mitmele seos
midagi ei ole korratud



Keelendid



kõik keelendid on keelendid, ka

- vasted
- sünonüümid
- kollokaadid jne

kõik need on meie baasis ühesugused,
märksõnataolised olemid

Olemasoleva sõnastiku import

.töö+tasu (palk) заработная плата, заработок;
вознаграждение за труд; оплата труда

.palk зарплата, оклад



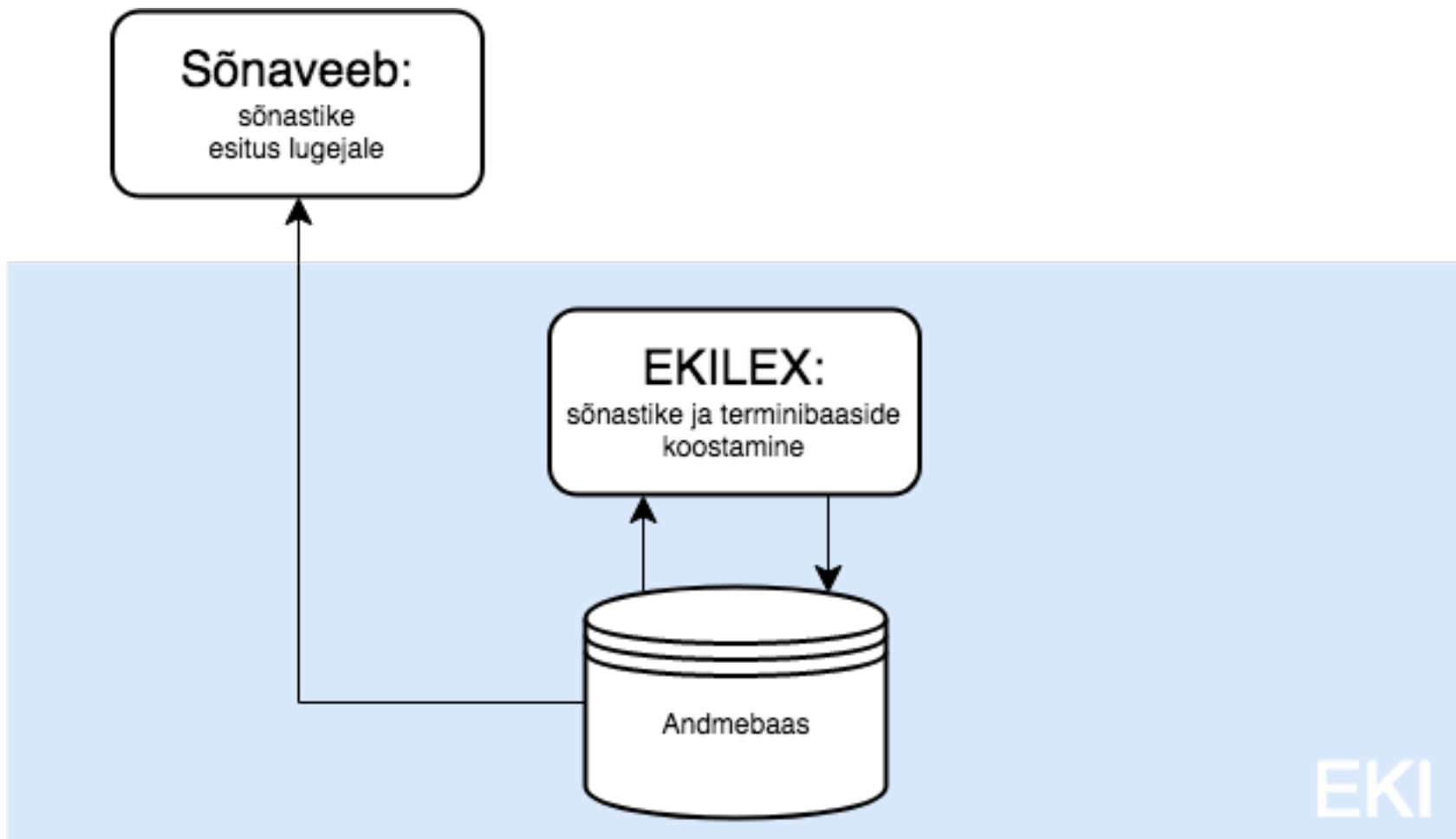
palk 1 (töötasu) заработная плата, заработок;
вознаграждение за труд; оплата труда

palk 2 зарплата, оклад

Kui sõnastikke on palju

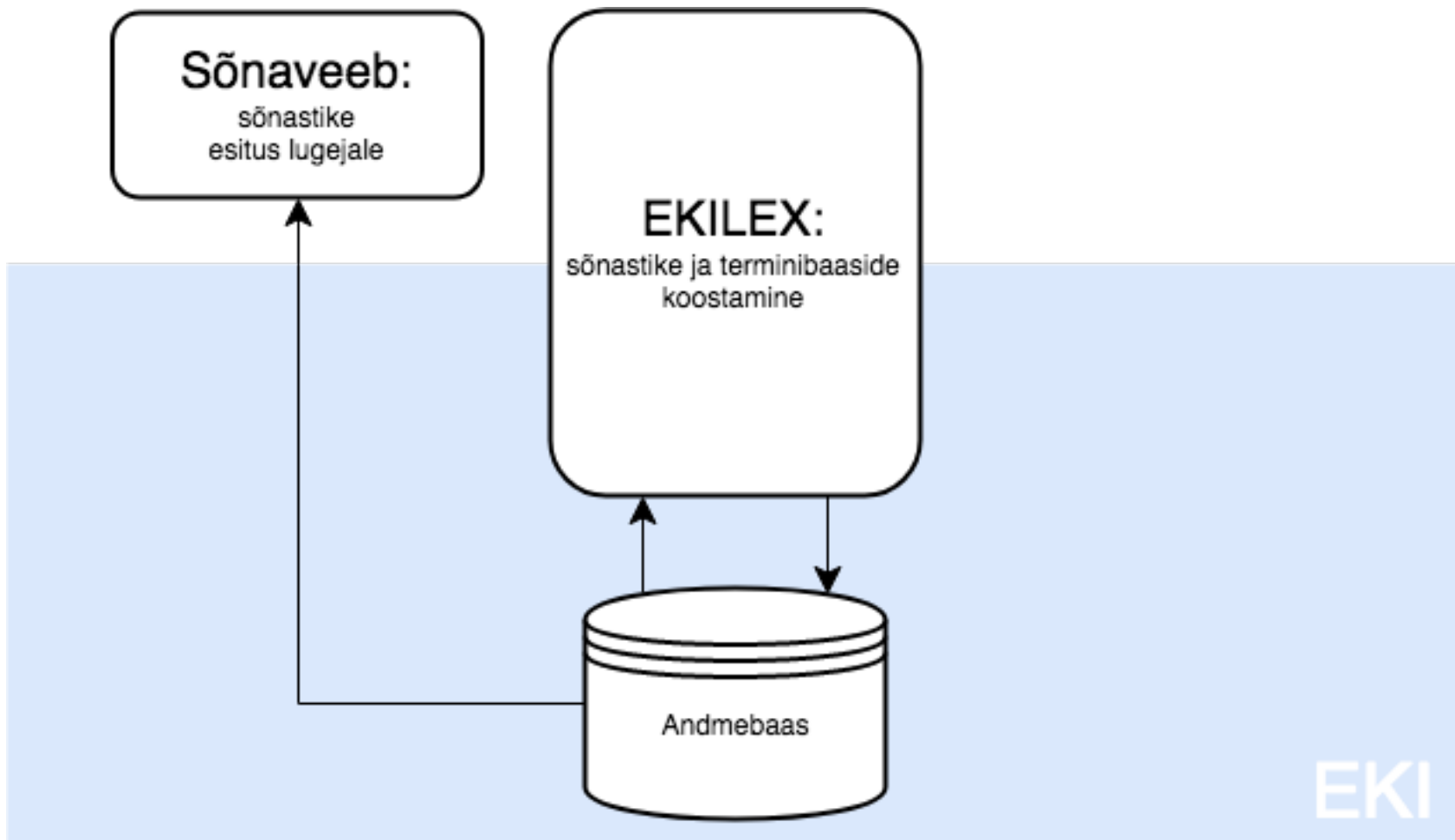
- palk 1 (raha-tähendus Estermist)
 - palk 2 (duplikaatne raha-tähendus Estermist)
 - palk 3 (puu-tähendus Estermist)
 - palk 4 (raha-tähendus raamatupidamise terminibaasist)
 - palk 5 (puu-tähendus metsanduse terminibaasist)
 - palk 6 (raha-tähendus EKSSist)
 - palk 7 (puu-tähendus EKSSist)
 - palk 8 (raha-tähendus ÕSist)
 - palk 9 (puu-tähendus ÕSist)
- Lõpuks tahaks kasutaja infot keele, mitte sõnastike kohta. Sellel sõnal on 2 tähendust, mitte 9

EKILEXi ökosüsteem



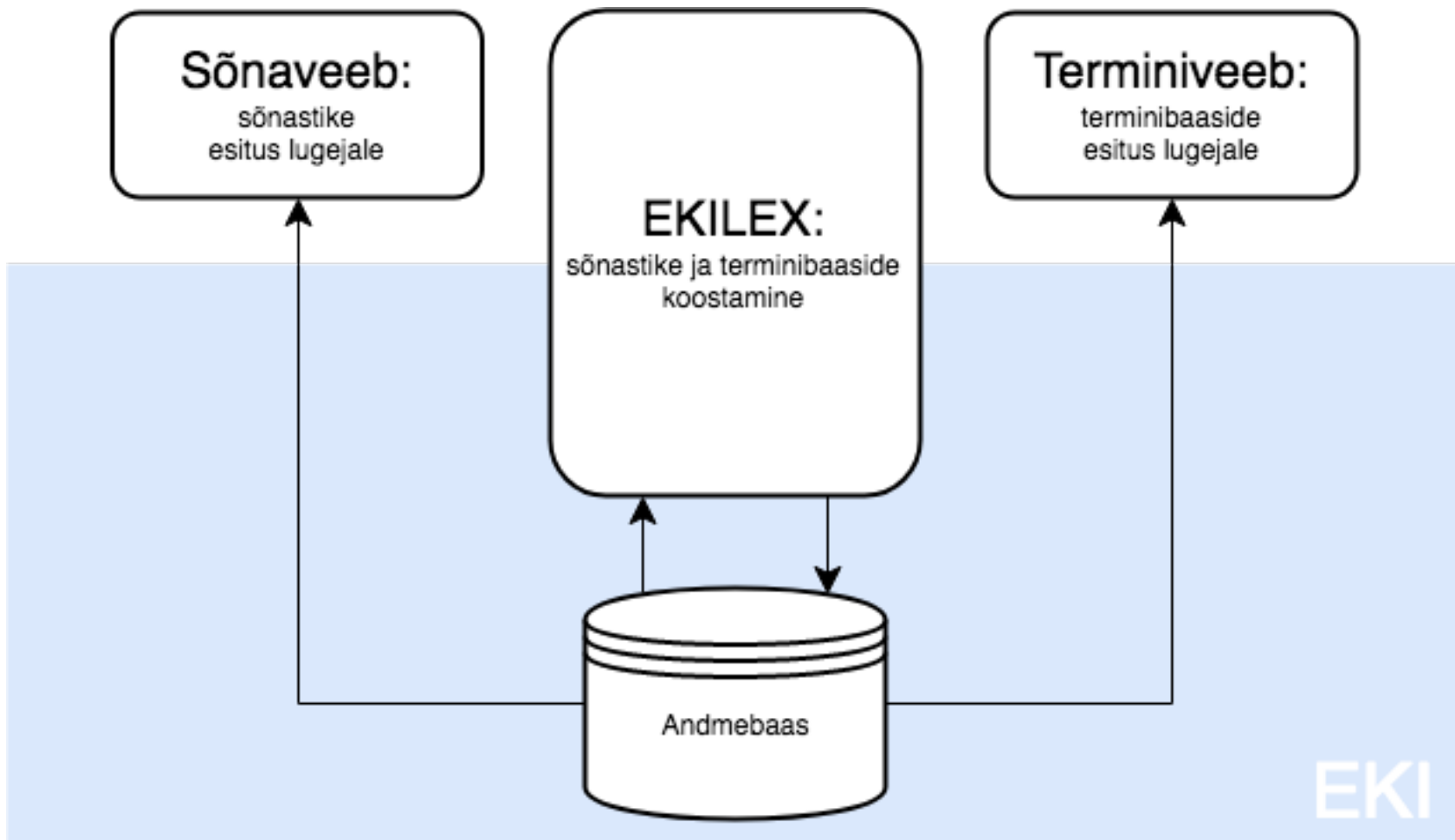
2017-2018

EKILEXi ökosüsteem



2019-2020

EKILEXi ökosüsteem



2019-2020

Hetkeseis

- EKILEX: üldiselt olemas, v.a
 - sõnakogude mestimine
 - uue sõnakogu loomine
 - unifitseeritud import
 - kasutajate haldus
 - mõned muutmisfunktsioonid
 - palju mugavusfunktsioone
- Sõnaveeb: kohe valmimas
- Terminiveeb: finantseerimistaotlus esitatud

Unistusi

- Üks keskne sõnastik ehk tähenduste ühendamine
 - kõigepealt homonüümitasemel: peaaegu ~~tehtav~~ tehtud
 - siis tähenduse tasemel: väga palju käsitööd; vaieldavused
- Palju rohkem korpust
 - sh erialakorpused
- Toorsõnastiku automaatne koostamine
- Sõnastiku kvantifitseerimine
 - mitte tähendab või ei tähenda
 - vaid 65% seninähtud juhtudest on tähendanud

Probleeme

- Tähendused ühendamata
 - sõnastike vahel
 - sõnastike sees
- Korpused semantiliselt märgendamata
 - (siinse) keelendi tähendus pole teada
- Lähtesõnastikud põhinevad sisekaemusel
- Koostaja ütleb muutmispalvele sirge "ei"
- Koostaja pahandab meie peale, kui talle tema vigu näitame
 - see on kõlbmatu info, seda ei tohi avaldada
- Kasutaja soovib teada, kuidas peaks olema
 - mitte kuidas on

Millega praegu pead vaevame

- Kuidas esitada infot, mille asukoht mudelis pole teada või võib muutuda?
 - väljendid
 - näited
 - etümoloogia
 - isegi sõnaliik
- Kuidas esitada ebakindlat infot?
 - oletatav etümoloogia
 - nõrgalt seotud kollokaadid
- Kuidas esitada ajutist infot?
 - käsitsi koostatud sõnakogu import

Aitäh