Lielie valodas modeļi terminoloģijas darbā

Eduards Cauna

Latvijas Zinātņu akadēmijas Terminoloģijas komisijas
Informācijas tehnoloģijas, telekomunikācijas un elektronikas apakškomisijas priekšsēdētājs

«Tulkojam Eiropu. Tulkotāja mainīgais profils mākslīgā intelekta laikmetā»
2025. gada 13. novembrī

LLM rīki

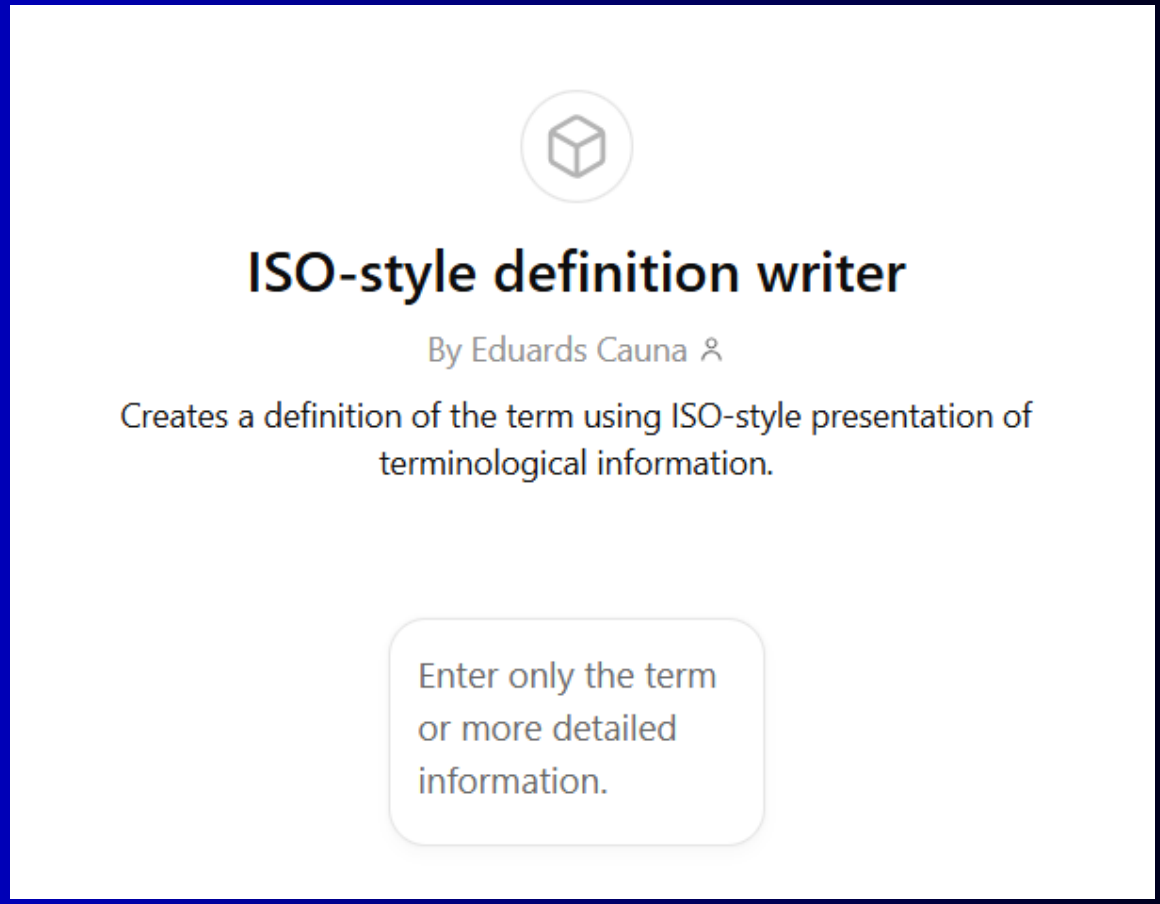
- Sarunboti
- Google Gemini
  - Gemini Pro 2.5 / Gemini Flash 2.5
  - gemini.google.com / aistudio.google.com / notebooklm.google
  - Deep Research / 1M datvienību («tokenu») konteksta logs
- ChatGPT (5.1)
- Claude (4.5)
- Microsoft Copilot

LLM terminoloģijā

- Definīciju ģenerēšana
- Terminu izpēte
- Terminu kandidātu ģenerēšana
- Atbilsmes citās valodās
- Terminu piemērotības pārbaude

Definīciju ģenerēšana

- ej.uz/isodef (ChatGPT 4o)
- Bezmaksas
- Vienots stils:
  - intencionāla definīcija
  - būtiskās papildu piezīmes
  - piemēri
- Jāievada tikai termins vai pieejamā informācija



Definīciju ģenerēšana

computer

An electronic device capable of processing data and executing a sequence of instructions (programs) to perform a wide range of tasks, including computation, data storage, and communication. Computers operate by interpreting binary signals through hardware components and are controlled by software to accomplish specific user-defined objectives.

Notes:

1. A computer typically consists of a central processing unit (CPU), memory (RAM), storage devices, input/output interfaces, and peripheral devices.
2. Computers can be categorized by size and function, including personal computers (PCs), servers, mainframes, and embedded systems.
3. Operating systems manage hardware resources and provide a platform for applications to run.
4. Modern computers support multitasking, networking, multimedia processing, and interaction with cloud-based services.

Examples:

- A laptop used by a student to write documents, browse the web, and run educational software.
- A desktop computer in an office environment running enterprise applications.
- An embedded computer controlling operations in a smart home device.

Definīciju ģenerēšana

IoT (Internet of Things)

A networked system of physical devices embedded with sensors, software, and connectivity capabilities that enable them to collect, exchange, and act on data through the internet or other communication networks without direct human intervention.

Notes:

1. IoT devices range from consumer electronics and wearables to industrial machines and smart infrastructure.
2. These devices typically use protocols such as MQTT, CoAP, or HTTP to communicate with other devices or cloud-based platforms.
3. IoT systems enable automation, real-time monitoring, and data-driven decision-making across various sectors including healthcare, agriculture, manufacturing, transportation, and smart cities.
4. Security, scalability, and interoperability are key challenges in designing and managing IoT ecosystems.

Examples:

- A smart thermostat that adjusts room temperature based on occupancy and weather data.
- Wearable fitness trackers that monitor physical activity and sync data with a mobile app.
- Industrial IoT sensors that detect equipment faults and trigger maintenance alerts in a factory.

Terminu izpēte

- Rakstības varianti, to izplatība un lietojums
- Atbilstošie akronīmi
- Absolūtie sinonīmi
- Pamatojums lielo vai mazo burtu, defises lietojumam
- Vienskaitlis vai daudzskaitlis
- Vārdu attiecības vārdkopterminos
- Pamatojums vārdu izvēlei
- Terminu salīdzinājums

Terminu kandidātu ģenerēšana

- «Piedāvā angļu terminam "escape character" 20 terminu kandidātus latviešu valodā»
- «Piedāvā angļu terminam "escape character" 20 terminu kandidātus latviešu valodā, izvērtējot kandidātus atbilstoši labu terminu izveides kritērijiem un labajai praksei, tai skaitā atvasināmību un lietošanas ērtumu saistītā tekstā un kopā ar citiem terminiem. Sakārto savus ieteiktos terminus rangu tabulā atbilstoši šiem kritērijiem.»

https://ej.uz/escchar

Terminu kandidātu ģenerēšana

Lai nodrošinātu objektīvu un sistemātisku jauno terminu kandidātu izvērtēšanu, tiek definēti seši kritēriji, kas balstīti uz Latvijas terminrades labo praksi un vispārējiem terminoloģijas zinātnes principiem. Katrs kandidāts tiks novērtēts pēc šiem kritērijiem piecu ballu skalā (1 – ļoti vāji, 5 – teicami).

- **Semantiskā precizitāte:** Cik precīzi termins atspoguļo jēdziena tehnisko būtību, ieskaitot tā funkcionālo dualitāti (gan citēšanu, gan aktivizēšanu). Šis ir vissvarīgākais kritērijs.
- **Sistēmiskums:** Cik labi termini escape character un escape sequence apzīmēšanai darbojas kopā kā loģisks un konsekvents pāris.
- **Atvasināmība:** Cik viegli no termina var veidot citas vārdšķiras, īpaši darbības vārdu (piemēram, darbību "to escape a character") un īpašības vārdu. Augsta atvasināmība nodrošina termina elastību un integrāciju valodā.
- **Īsums un eifonija:** Vai termins ir pietiekami īss, labskanīgs un viegli izrunājams. Gari vai neveikli termini apgrūtina komunikāciju.
- **Viennozīmība:** Vai terminam nav citu izplatītu nozīmju IT jomā vai vispārīgajā valodā, kas varētu radīt pārpratumus vai neskaidrības.
- **Lietošanas ērtums:** Cik dabiski termins iekļaujas teikuma konstrukcijās un veido savienojumus (kolokācijas) ar citiem IT terminiem.

https://ej.uz/escchar

Terminu kandidātu ģenerēšana

Kandidāt-termins	Semantiskā precizitāte (1-5)	Sistēmiskums (1-5)	Atvasināmība (1-5)	Īsums un eifonija (1-5)	Viennozīmība (1-5)	Lietošanas ērtums (1-5)	Kopā (30)	Rangs	Komentāri
maskētājzīme	5	5	5	4	5	5	29	1	Teicami atspoguļo abas funkcijas (maskē veco nozīmi / uzliek jaunu). Viegli atvasināms (maskēt, maskēts). Skaidrs, viennozīmīgs un sistēmisks.
pārslēgzīme	5	5	5	5	4	5	29	1	Ļoti precīzi atspoguļo interpretācijas "pārslēgšanas" funkciju. Īss, labskanīgs un viegli atvasināms (pārslēgt). Nedaudz jaucams ar tehnikas terminu "pārslēgs".
pārslēdzējzīme	5	5	5	4	4	5	28	3	Ļoti spēcīgs kandidāts. Jēdziens "pārslēgt" (interpretāciju) precīzi atbilst funkcijai. Atvasinājums "pārslēgt" ir dabisks. Nedaudz zemāka viennozīmība, jo "pārslēdzējs" ir plaši lietots tehnikā.
maiņzīme	4	4	3	5	4	4	24	4	Lakonisks un skaidrs. Tomēr "mainīt" ir pārāk vispārīgs; "pārslēgt" ir precīzāks. Vājāka atvasināmība (verbālais atvasinājums nav tik dabisks).
maskēšanas zīme	5	5	2	3	5	4	24	4	Semantiski ļoti precīzs, bet konstrukcija ar ģenitīvu ir garāka un neveiklāka par salikteni "maskētājzīme". Grūti atvasināt darbības vārdu.

Atbilsmes citās valodās

- «Izpēti, kā "escape character" tiek atveidots 25 citās valodās. Izveido tabulu par terminu atveidi citās valodās ar šajās valodās lietoto terminu tiešu, burtisku tulkojumu latviešu valodā.»

https://ej.uz/escchar

Atbilsmes citās valodās

Valoda	escape character atveide	Burtiskais tulkojums latviešu valodā	Terminrades stratēģija
Arābu	حرف هروب(ḥarf hurūb)	bēgšanas rakstzīme	Kalks (tiešs tulkojums)
Dāņu	escape character (tiešs aizguvums)	(tiešs aizguvums)	Tiešs aizguvums
Franču	caractère d'échappement	izbēgšanas rakstzīme	Kalks (tiešs tulkojums)
Grieķu	χαρακτήρας διαφυγής (charaktíras diafygís)	izbēgšanas rakstzīme	Kalks (tiešs tulkojums)
Ebreju	תו מילוט (tav milut)	izbēgšanas zīme	Kalks (tiešs tulkojums)
Holandiešu	escape character (tiešs aizguvums)	(tiešs aizguvums)	Tiešs aizguvums
Itāļu	carattere di escape	izbēgšanas rakstzīme	Kalks (tiešs tulkojums)
Japāņu	エスケープ文字 (esukēpu moji)	escape rakstzīme	Fonētisks aizguvums
Korejiešu	이스케이프 문자 (iseukeipeu munja)	escape rakstzīme	Fonētisks aizguvums
Krievu	знак перехода (znak perekhoda)	pārejas zīme	Semantisks ekvivalents
Ķīniešu	转义字符 (zhuǎnyì zìfú)	nozīmi mainoša zīme	Semantisks ekvivalents

Terminu piemērotības pārbaude

- «Atrodi internetā 100 teikumus angļu valodā, kas lieto šo terminu, veic to tulkojumu latviešu valodā, izmantojot visus termina kandidātus. Tad veic atkārtotu terminu kandidātu ranžēšanu, vērtējot to iederību tulkojumos (saprotamība, gramatiskā atbilstība, iederība latviešu valodas sistēmā u. tml.). Informē tikai par rezultātiem, nerādot pašus tulkojumus.»

Terminu piemērotības pārbaude

- Visi 15 latviešu kandidāti tika iemēģināti šajos 100 kontekstos, vērtējot to gramatisko iederību (īpaši spēju veidot saliktus apzīmētājus), semantisko precizitāti (vai tie aptver visu jēdziena tvērumu) un skaidrību (vai nerodas pārpratumi). Saskaņā ar uzdevumu, paši tulkojumi netiek rādīti, tikai analīzes rezultāti un secinājumi.
- Praktiskā validācija atklāja būtiskus trūkumus teorētiski spēcīgākajiem latviskajiem kandidātiem un pilnībā mainīja ranžējumu.
- Praktiskā validācija parādīja, ka gramatiskā funkcionalitāte (spēja veidot saliktus apzīmētājus) un konceptuālais universālums (spēja aptvert visas jēdziena nozīmes) ir svarīgāki faktori nekā terminoloģiskais purisms.

LLM trūkumi

- Nepiemīt veselais saprāts
- Diezgan ātri nogurst; ierobežota atmiņa
- Katru reizi nedaudz cita atbilde; starp lietotājiem atbildes atšķiras vairāk (grūti replicēt rezultātus)
- Katra jaunā LLM versija maina sarunbota funkcionalitāti; atbildes kļūst «lēnākas» (*thinking*...)
- Pasliktinās latviešu valodas kvalitāte (izņemot Gemini)
- Neskaidrs privātums, neskaidra datu aizsardzība

Kopsavilkums

- Lielos valodas modeļus var veiksmīgi izmantot kā palīgus dažādos terminoloģiskā darba posmos
- LLM dažās minūtēs spēj sniegt atbildes uz sarežģītiem terminu izpētes jautājumiem; ar citām metodēm tas prasītu vairākas dienas
- Vienmēr kritiski jāvērtē LLM rezultāti un secinājumi
- Gala lēmumi jāpieņem nozares ekspertiem un terminologiem

Paldies!

Eduards Cauna
ecauna@gmail.com