



Neuroninių kalbos technologijų taikymai

Applications of Neural network based Natural Language Technologies

dr. Darius Amilevičius
Vytauto Didžiojo universitetas
Skaitmeninių išteklių ir tarpdisciplininių tyrimų institutas / Informatikos fakultetas

darius.amilevicius@vdu.lt



NLT pasaulinės rinkos augimo prognozės

Report Attribute	Details
Estimated Base Year Value (2021)	US\$ 11 Billion
Expected Market Value (2022)	US\$ 14 Billion
Anticipated Forecast Value (2032)	US\$ 45 Billion
Projected Growth Rate (2022-2032)	23% CAGR

Šaltinis: <https://www.futuremarketinsights.com/reports/natural-language-processing-nlp-market>

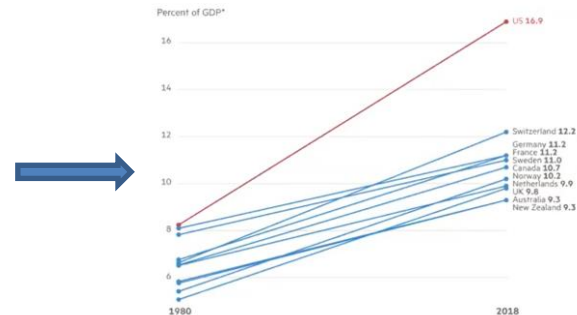


Kontekstas

AI NLT
gamyboje ir pramonėje →



AI NLT
Pvz.: Sveikatos apsaugos →
ir teisinėje sistemose





- 1948 – taisyklės
- 1990 – statistinė revoliucija
- 2001 - neuroniniai kalbos modeliai
- 2013 – įterptiniai žodžiai
- 2013 – HLT neuroniniai tinklai
- 2014 – sekvencija-į-sekvenciją modeliai
- 2015 – dėmesio sprendimai
- 2018 – iš anksto apmokyti kalbos modeliai



A. Tiuringas (1950):

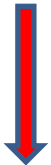
Užuot klausę, ar mašinos gali mąstyti, turėtume klausti, ar mašinos gali išlaikyti intelektualios elgsenos testą? (Tiuringo testą)

NKT Pasaulyje:

1948 m. - Taisykliniai ir heuristiniai metodai

1990 m. – statistiniai ir mašininio mokymo metodai

2013 m. – neuroninės NKT



N. Chomskis vs. G. Marcus:
natyvizmas vs. empirizmas

NKT Lietuvoje:

1994 m. pirmieji žingsniai

2012 m. – LKIV programa (Semantika-1, Liepa-1, kiti)

2013 m. – VLKK Gairės

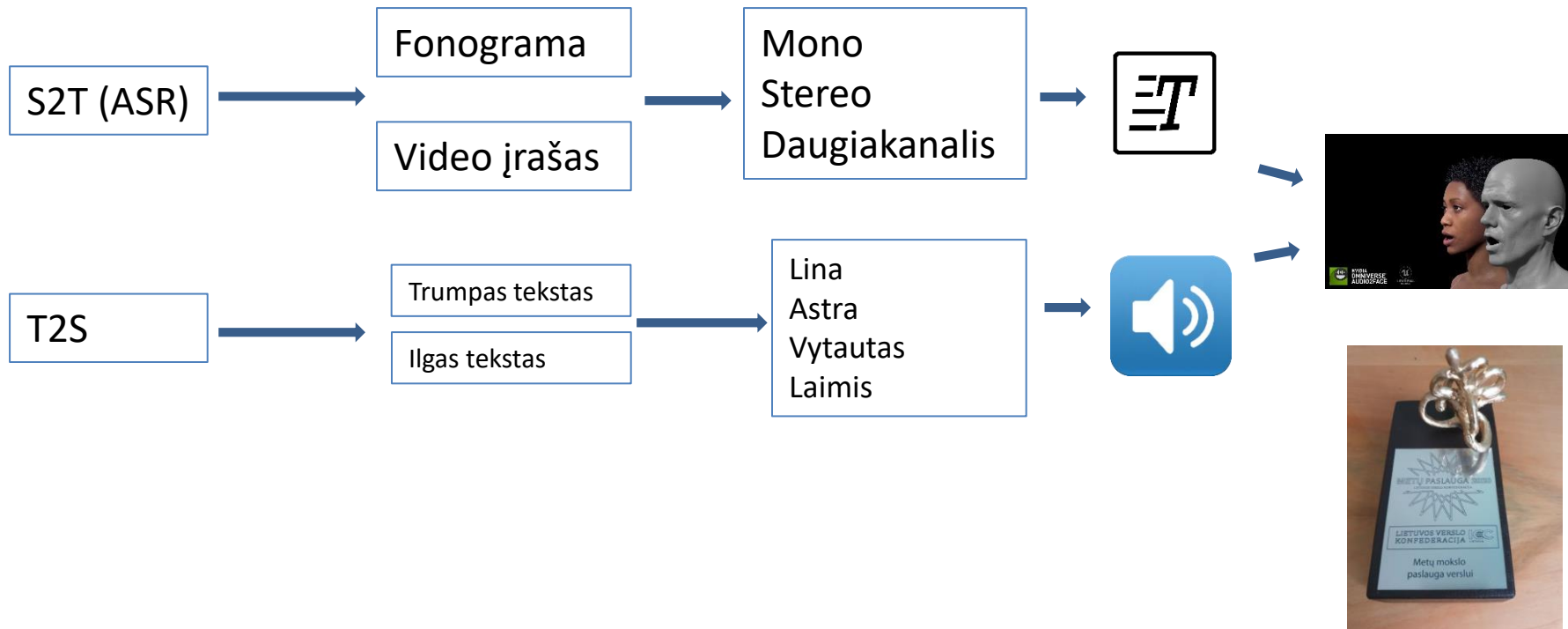
2018 m. – LKIT programa (Semantika-2, Liepa-2, kiti)

2020 m. – VLKK-Seimas Gairės 2021–2027

???? m. - „Naujos kartos Lietuva“

(Lietuvių kalbos ištekliai dirbtinio intelekto ir inovatyvių technologijų vystymui)

VDU taikomieji tyrimai: Šneka





VDU taikomieji tyrimai: tekstynai, teksto analizė

- Morfologijos ir sintaksės analizė
- Automatinis kirčiavimas, vienareikšminimas
- Įvardytų esybių atpažinimas, sentimentų analizė, kiti
- Tekstynai ir neuroninis kalbos modeliavimas
- Automatinis santraukų sudarymas, aspektais grįsta sentimentų analizė
- Neapykantos kalbos atpažinimas, klaidinančios informacijos tekstų atpažinimas, kiti
- kiti

Automatizavimas: sveikatos apsauga

Kūno odoje, daugiau juosmens srityje, numuliarinės plokščios papulės, uždegiminės, su žvynu. Alkūnių, kelių srityje jos susilieja į stambesnius židinius. Tarpgliuteralinėje raukšlėje paraudęs, infiltruotas židiny su įtrūkimu ir maceracija. Kojų nagai distrofiški, padų ir delnų srityje sausos pustulės, padengtos žvynu. Aušpico triada teigiama. Galvoje pavienės plokštelės. Bendras kraujo tyrimas 2003-08-13: leukocitai - $5,5 \times 10^9/l$, eritrocitai - $4,38 \times 10^{12}/l$, hemoglobinas- 134 g/l, trombocitai - $232 \times 10^9/l$, leukocitų formulė: lazdeliniai- 1%, segmentuoti- 55%, bazofilai- 0, eozinofilai- 1%, limfocitai- 33%, monocitai- 10%, ENG pagal Westergreną- 8 mm/val. Šlapimo tyrimas normos ribise. Tyrimas dėl grybelio iš pakitusių rankų nagų: grybelio nerasta. Biocheminis kraujo tyrimas: bilirubinas bendras- 6,2 mkmol/l, ASAT- 23 u/l, ALAT- 21 u/l, gliukozės kiekis kraujyje- 5,7 mmol/l. PUVA terapija, vartojant Oksaraleno po 10 mg 4 tbl 1,5 val prieš PUVA terapijos procedūrą, viso gavo 8 procedūras. Tepėsi deguto 2 % tepalu pusiau su Sinaflanu vieną kartą dnoje



Unstructured data



Text mining



Interoperable & Aggregated

Structured information



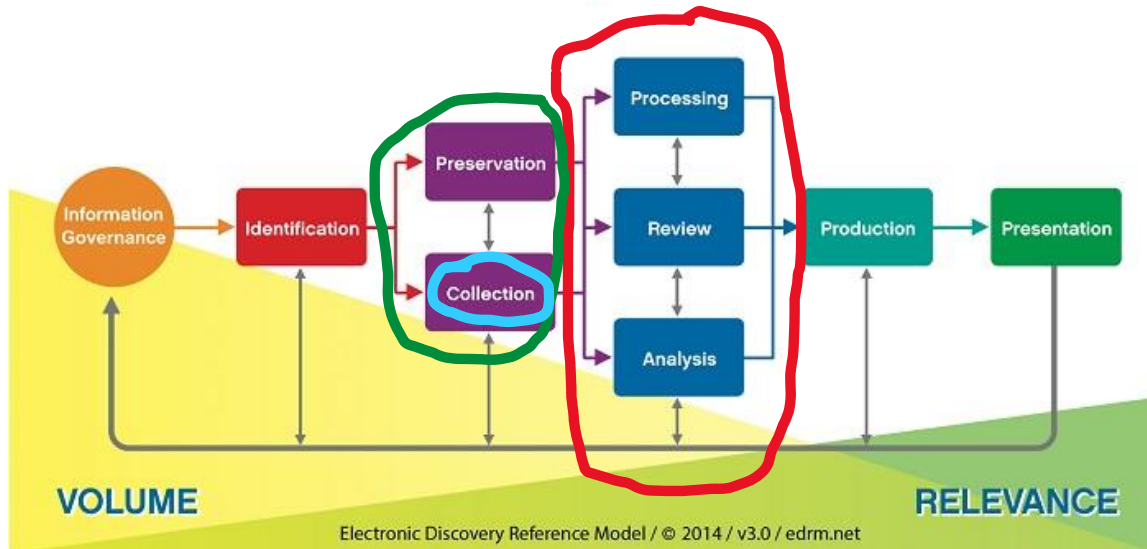
Insights Generation



Actionable knowledge

Automatizavimas: teisinė sistema

Electronic Discovery Reference Model





Problemos (1)

Remiantis R. Susskindo nuomone, visų pirma būtina atsakyti į penkis esminius klausimus:

- ar technologijos jau pakankamai brandžios, kad galėtų pakeisti gydytoją/teisininką
- ar moraliai jau priimtina, kad mašinos perimtų bent kai kurias gydytojo/tesininko funkcijas
- ar tokios sistemos jau komerciškai perspektyvios, ar jų ekonominė nauda bus didesnė už jų kūrimo išlaidas
- ar tai bus tvaru (nepaskatins atmetimo reakcijos), atsižvelgiant į kultūrinį kontekstą
- ar yra teisinis pagrindas kurti tokias sistemas, t. y. ar tam tikroje nacionalinėje sistemoje sprendimų priėmimo procedūra nesusijusi su kokia nors specifika, dėl ko sprendimo priėmimas iš dalies ar visiškai negali būti patikėtas mašinai



Problemos (2)

- Natūralios kalbos supratimas
- NNKT mažus išteklius turinčioms kalboms
- Didelės apimties arba kelių dokumentų analizė
- Duomenų ištekliai ir jų validavimas



Děkoju už děmesj!

