

Europos kalbų išteklių konsorciumo seminaras

Kalba ir kalbos technologijos

Lietuvoje

Daiva VAIŠNIENĖ
Laimutis TELKSNYS
Renata ŠPUKIENĖ



- Automatinio vertimo sistemų tobulinimas



- Šnekamosios kalbos technologijų tobulinimas



- Sintaksinės ir semantinės analizės tobulinimas



- Bazinių kalbos išteklių ir įrankių kūrimas



- Infrastruktūros plėtra

Faktai

- Imigrantų skaičius 2014 metais - 24 294.
- 80,8 procento miesto ir 73,8 procento kaimo gyventojų moka bent vieną užsienio kalbą. Anglų kalbą moka 30,4 proc., o rusų – 63 proc. gyventojų. Dažniausiai užsienio kalbų visai nemokėjo gyventojai iki 15 metų amžiaus bei 80 metų ir vyresni.

Sprendiniai

Automatinis vertimas;
Semantinė paieška, sujungta su vertimo įrankiais;
Kalbinės pagalbos paslaugos (gramatikos tikrintuvai, tartuvai, kirčiuoklės);
Mokomosios paslaugos (tekstynai, žodynai, nuotolinio mokymo(si) programos).
Informacijos išgavimo ir apdorojimo paslaugos (nuomonių analizė, tekstų santraukų kūrimo įrankiai terminų identifikavimas)



Faktai

- Lietuvoje yra apie 8 tūkst. kurčiųjų ir beveik 40 tūkst. neprigirdinčiųjų.
- Lietuvos aklujų ir silpnaregių sąjunga vienija 6,5 tūkst. narių, tačiau šalyje gali būti iki 60 tūkst. žmonių, turinčių vienokių ar kitokių regėjimo problemų.

Sprendiniai

Kompiuterio valdymas balsu (balso atpažintuvas);
Šnekos sintezavimas (skaitytuvas)
Semantinė paieška, sujungta su vertimo įrankiais.



- **57 proc.** interneto naudotojų Europoje perka prekes ir paslaugas ne savo gimtąja kalba (anglų kalba yra labiausiai paplitusi užsienio kalba, po jos – prancūzų, vokiečių ir ispanų kalbos);
- **55 proc.** naudotojų skaito internete pateikiamą informaciją užsienio kalbomis;
- vos **35 proc.** užsienio kalbomis rašo elektroninius laiškus arba komentarus interneto tinklalapiuose užsienio kalba.



**Interneto vartai,
leidžiantys atvirą
prieigą prie visų per
šią programą
sukurtų kalbos
išteklių ir
technologijų**

Naršytuvas

„Naršytuvas“ – paslauga, leidžianti naršyti internete duodant kompiuteriui nurodymus komandomis balsu.

Paslauga „Naršytuvas“ diegiama vartotojo kompiuteryje. Ji veikia Linux operacinėje sistemoje, bei Windows 7/8/8.1 operacinėse sistemose.



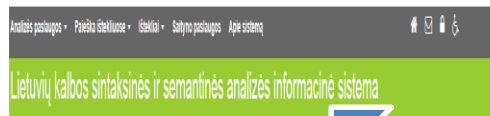
**Šnekamosios
kalbos atpažinimo ir
sintezės įrankių
tobulinimas ir
diegimas į
paslaugas**



**Skaitmeniniai
kalbos ištekliai
(skaitmeniniai
žodynai ir pan.)**



**Nauji automatinio
vertimo įrankiai:
lietuvių–anglų–
lietuvių, lietuvių–
prancūzų–lietuvių**



**Semantinės ir
sintaksinės analizės
ir paieškos įrankiai**



Turtinant lietuvių šnekamosios kalbos išteklius, sulydžius lietuvių filologijos ir informatikos žinias, sukurti lietuvių šnekamosios kalbos plėtros *technologinė bazė ir infrastruktūros pamatiniai elementai*:

- **Atpažintuvai**
- **Sintezatoriai**
- **Šnekos duomenų bazė - garsynas**
- **Balso įrašų laboratorija**

Sukurti lietuvių šneka valdomų *paslaugų pavyzdžiai* (1):

- Elektroninius tekstus galima **klausyti skaitomus lietuviškai** dviem moteriškais ir dviem vyriškais balsais
- **Aklieji**, dirbantys kompiuteriais, girdi taisyklinga lietuvių šneka nusakomus vaizdus, kurie rodomi jų kompiuterių ekranuose
- **Moksleiviai**, šnekėdami su kompiuteriais lietuviškai, pradeda mokytis pažinti ląstelės sandarą, susipažinti su UNESCO pasaulyje globojamu kultūros paveldu

Sukurti lietuvių šneka valdomų *paslaugų pavyzdžiai* (2):

- Kompiuterio lietuviškas balsas vadovavo **Lietuvos nacionalinio dramos teatro** tarptautinio festivalio Sirenos 2014 ir 2015 metų spektakliams Remote Vilnius – Nutolęs Vilnius
- **Žmonės, stokojantys gebėjimų bendrauti** vieni su kitais gali mokytis jų įgyti, pasitelkę lietuviškai šnekančius ir lietuvių šneką atpažįstančius kompiuterizuotus įrankius. Gali mokytis teisingai lietuviškai vadinti rodokus įvairius daiktus
- Kiekvienas žmogus, būdamas bet kurioje pasaulio vietoje, pasitekęs internetą, gali išgirsti, kaip turi būti lietuviškai **taisyklingai tariamas naujažodis**



Padaryti darbai yra reikšmingi, tačiau jie tik **praveria** kelią lietuvių šnekai į pasaulio daugiakalbę elektroninę erdvę.

Tęsiant darbus ir vykdant projektą **LIEPA 2 – Lietuvių šneka valdomų paslaugų plėtra**, būtina:

- Išplėsti žmonių bendravimo su kompiuteriais **žodyną**
- Patobulinti elektroninio teksto **skaitytuvo galimybes**
- Perkelti lietuvių šneką į **mobiliają įrangą**
- Pasirūpinti lietuvių šnekos **tarpkalbiniais ryšiais**
- Parengti šneka valdomas priemones **skirtingoms operacinėms sistemoms**



Igyvendinti darbai, atlikti vykdant projektą
LIEPA - *Lietuvių šneka valdomos paslaugos*,
buvo padaryti sutelktomis pastangomis:

- **Vilniaus universiteto**
- **Lietuvos edukologijos universiteto**
- **Lietuvių kalbos instituto**
- **Šiaulių universiteto**

Tęsiant darbus numatoma bendradarbiauti su projektais:

Semantika 2 - *Lietuvių kalbos teksto sintaksinės-semantinės analizės informacinės sistemos viešųjų paslaugų vystymas*

Mašininio vertimo sistemų ir lokalizavimo paslaugų tobulinimas ir plėtra

Raštija 2 - *Integruotų lietuvių kalbos ir raštijos išteklių informacinės sistemos plėtra*

Mašininio vertimo ištekliai

Renata Špukienė, Tilde IT



Dvidešimt ketverius metus dirbame *Tilde* įmonių grupėje Baltijos šalyse

Šešiolika metų dirbame **kalbinių technologijų srityje** Lietuvoje:

Pažangių kalbinių technologijų vystymas ir sprendimai

Kalbos tvarkymo priemonė *Tildės Biuras*

Vertimas ir lokalizavimas

Esame įvairių asociacijų nariai ir Lietuvoje, ir užsienyje

Atliekame pažangiausius mokslinius tyrimus kartu su partneriais visame pasaulyje





- ✓ Kalbos tikrinimo priemonės
 - Lietuvių kalbos rašybos tikrintuvas
 - Lietuvių kalbos gramatikos tikrintuvas
- ✓ Mašininio vertimo sistemos
- ✓ Semantinės sistemos
- ✓ Kiti kalbiniai įrankiai
 - Morfologijos analizatorius
 - Sintaksės analizatorius
 - Lemuoklis
 - Kalbos dalių žymėjimo įrankis
 - Morfologijos ir sintaksės anotavimo įrankis
 - Įvardintų esybių atpažintuvas
 - Sakinių atskyriklis





„Tilde“ vertyklė – tai žodyno ir teksto automatinio vertimo derinys.

Vertimo kryptis:

Įveskite norimą išverst

anglų-lietuvių
anglų-latvių
lietuvių-rusų
rusų-lietuvių
lietuvių-anglų
latvių-rusų
estų-anglų
latvių-anglų
rusų-latvių

Versti

- ✓ Viešai prieinama internete www.tilde.lt
- ✓ 10 vertimo krypčių iš arba į anglų, lietuvių, rusų, latvių, estų kalbas
- ✓ Bendrosios srities vertimai (nespecializuoti)
- ✓ Neprofesionalams





„Tilde“ vertyklės mobilioji versija „Tilde Translator“



- ✓ Žodžių, frazių ir teksto vertimas
- ✓ Kalbos: lietuvių, anglų, rusų, latvių, estų
- ✓ Vertimo istorijos peržiūra
- ✓ Transliteracija, skirta įrenginiams, nepalaikantiems diakritinių ženklų (š, ž, kt.)
- ✓ Sąsajos kalbos: lietuvių, anglų, rusų, latvių, estų,





„Tildės biuras“ – kasdien dirbantiems kompiuteriu



- ✓ Automatinis tekstų ir dokumentų vertimas
- ✓ Rašybos ir gramatikos tikrintuvas
- ✓ Įvairių kalbų ir sričių žodynai
- ✓ Interaktyvus semantinis žodžių tinklas
- ✓ Dokumentų šablonai
- ✓ Vaizdų atpažinimas
- ✓ Šriftai
- ✓ Enciklopedijos

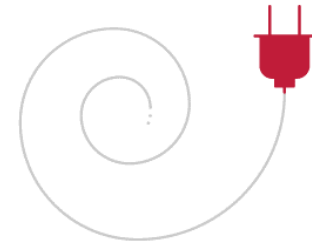




Tilde MV platforma – mašininio vertimo sistemos pagal poreikį

Galimybė naudotis esamomis MV sistemomis:

- ✓ Įkelti ir versti tekstus, dokumentus, puslapius
- ✓ Rinktis vieną iš daugelio sukurtų mašininio vertimo sistemų:
 - ✓ 15 skirtingų sričių
 - ✓ 100 kalbų krypčių
- ✓ Naudoti terminų rinkinius tikslumui didinti
- ✓ Įvertinti mašininio vertimo kokybę





Tilde MV platforma – mašininio vertimo sistemos pagal poreikį

Galimybė užsisakyti MV sistemą:

- ✓ Apdorojami turimi užsakovo resursai
- ✓ Surenkama papildomų resursų
- ✓ Sukuriama ir tobulinama kliento MV sistema
- ✓ Integruojama kliento terminologija
- ✓ Įvertinama mašininio vertimo kokybė
- ✓ Įdiegiama kliento infrastruktūroje





Tilde MV platforma – mašininio vertimo sistemos pagal poreikį

Galimybė patiems susikurti MV sistemą:

- ✓ Visos reikalingos techninės priemonės
- ✓ Patogi ir lengvai valdoma vartotojo sąsaja
- ✓ Didžiuliai kiekiai įvairių kalbų ir sričių duomenų
- ✓ Vartotojo išteklių pridėjimas
- ✓ Terminologijos integravimas
- ✓ Kalbinių įrankių panaudojimas
- ✓ Įdiegiama kliento infrastruktūroje





Tilde MV platforma – mašininio vertimo sistemos pagal poreikį

Skirta naudoti:

- ✓ Vertimo profesionalams
- ✓ Įmonėms
- ✓ Viešajam sektoriui
- ✓ Akademinio bendradarbiavimo tikslais

Privalumai:

- ✓ Pateikia aukštesnės kokybės vertimus
- ✓ Garantuoja konfidencialumą
- ✓ Užtikrina nuoseklią terminiją
- ✓ Išlaiko originalų dokumentų formatą
- ✓ Palaikomi įvairūs failų formatai
- ✓ Integruojama į užsakovų infrastruktūrą

Dėkojame už dėmesį.