

Strojno prevođenje tekstova kao alat u diseminaciji informacija

Šimičić, Karla

Undergraduate thesis / Završni rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zadar / Sveučilište u Zadru**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:162:438514>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-12-24**



Sveučilište u Zadru
Universitas Studiorum
Jadertina | 1396 | 2002 |

Repository / Repozitorij:

[University of Zadar Institutional Repository](#)



zir.nsk.hr



DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJ

Sveučilište u Zadru
Odjel za informacijske znanosti
Sveučilišni prijediplomski studij
Informacijske znanosti

Karla Šimičić

**Strojno prevođenje tekstova kao alat u diseminaciji
informacija**

Završni rad

Zadar, 2024.

Sveučilište u Zadru
Odjel za informacijske znanosti
Sveučilišni prijediplomski studij
Informacijske znanosti

Strojno prevođenje tekstova kao alat u diseminaciji informacija

Završni rad

Student/ica:
Karla Šimičić

Mentor/ica:
doc. dr. sc. Ante Panjkota
Komentor/ica:
dr. sc. Nikolina Peša Pavlović

Zadar, 2024.



Izjava o akademskoj čestitosti

Ja, **Karla Šimičić**, ovime izjavljujem da je moj **završni** rad pod naslovom **Strojno prevođenje tekstova kao alat u diseminaciji informacija** rezultat mojega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na izvore i radove navedene u bilješkama i popisu literature. Ni jedan dio mojega rada nije napisan na nedopušten način, odnosno nije prepisan iz necitiranih radova i ne krši bilo čija autorska prava.

Izjavljujem da ni jedan dio ovoga rada nije iskorišten u kojem drugom radu pri bilo kojoj drugoj visokoškolskoj, znanstvenoj, obrazovnoj ili inoj ustanovi.

Sadržaj mojega rada u potpunosti odgovara sadržaju obranjenoga i nakon obrane uređenoga rada.

Zadar, 28. listopada 2024.

SADRŽAJ:

UVOD.....	1
1 POJMOVNO ODREĐENJE DISEMINACIJE.....	3
1.1 DISEMINACIJA PODRŽANA TEHNOLOGIJOM.....	5
2 METODE DISEMINACIJE INFORMACIJA	6
3 STROJNO PREVOĐENJE	8
3.1 POVIJEST STROJNOG PREVOĐENJA	9
3.1.1 <i>Strojno prevođenje temeljeno na pravilima.....</i>	<i>10</i>
3.1.2 <i>Strojno prevođenje temeljeno na primjerima</i>	<i>11</i>
3.1.3 <i>Statističko strojno prevođenje</i>	<i>13</i>
4 NEURONSKO STROJNO PREVOĐENJE	14
4.1 STROJNO PREVOĐENJE TEMELJENO NA DUBOKOM UČENJU	16
4.2 TRANSFORMATORI KOD NEURONSKOG STROJNOG PREVOĐENJA.....	18
5 POVEZANOST STROJNOG PREVOĐENJA I DISEMINACIJE INFORMACIJA	
21	
6 PRIMJERI UPOTREBE STROJNOG PREVOĐENJA U DISEMINACIJI	
INFORMACIJA.....	23
6.1 PRIMJENA STROJNOG PREVOĐENJA U PODRUČJU ZNANSTVENE KOMUNIKACIJE	23
6.2 PRIMJENA STROJNOG PREVOĐENJA U ZDRAVSTVU.....	24
6.2.1 <i>Strojno prevođenje u doba COVID-19 pandemije</i>	<i>26</i>
6.2.2 <i>Strojno prevođenje njemačkih zdravstvenih tekstova.....</i>	<i>27</i>
6.3 PRIMJENA STROJNOG PREVOĐENJA U OBRAZOVANJU	28
6.4 PRIMJENA STROJNOG PREVOĐENJA U POSLOVANJU	30
7 USPOREDBA ALATA ZA STROJNO PREVOĐENJE U DISEMINACIJI	
INFORMACIJA.....	32
8 ZAKLJUČAK.....	38
LITERATURA.....	39

POPIS SLIKA	42
POPIS TABLICA.....	43

SAŽETAK

Predmet rada je analiza diseminacije informacija kroz strojno prevođenje, sa posebnim naglaskom na tehnologije koje podržavaju ovaj proces. Strojno prevođenje predstavlja ključni alat u modernoj komunikaciji, omogućujući brzi i efikasan prijenos informacija između različitih jezika i kultura. Motivacija za proučavanje ovog predmeta proizlazi iz sve veće potrebe za učinkovitom komunikacijom i razmjenom informacija u globaliziranom svijetu, gdje je brzina i točnost informacija ključna, osobito u kontekstu znanstvene komunikacije, zdravstva i obrazovanja. Metodologija istraživanja uključuje analizu jednog teksta ubačenog u postojeće alate za strojno prevođenje, nakon čega je prikazana i usporedba dobivenih rezultata iz izabranih alata. Najvažniji rezultati ukazuju na to da su moderni alati za strojno prevođenje, posebno oni temeljeni na dubokom učenju, značajno poboljšali kvalitetu automatskih prijevoda, čime su se otvorila vrata za širu primjenu u raznim sektorima, uključujući znanost, zdravstvo i obrazovanje. Ova tehnologija ne samo da olakšava pristup informacijama, već i potiče globalnu suradnju, čime se smanjuju jezične barijere i poboljšava međusobno razumijevanje. Glavni zaključak rada ističe da je integracija strojnog prevođenja u procese diseminacije informacija od esencijalne važnosti za poboljšanje pristupa znanju i resursima, čime se otvaraju nove mogućnosti za suradnju i inovacije u različitim sektorima. Ove implikacije sugeriraju potrebu za daljnjim istraživanjem i razvojem tehnologija koje će dodatno unaprijediti učinkovitost komunikacijskih procesa.

Ključne riječi: diseminacija, informacije, znanje, strojno prevođenje.

UVOD

Tema diseminacije informacija kroz strojno prevođenje odražava trenutni kontekst u kojem se tehnologije prevođenja razvijaju brzim tempom, omogućujući globalnu komunikaciju i razmjenu znanja. U današnjem digitalnom dobu, brzina i točnost informacija postaju ključni, a strojno prevođenje se ističe kao važno sredstvo za prevladavanje jezičnih barijera. Ova tema je zanimljiva jer istražuje kako tehnologija može unaprijediti pristup informacijama u različitim sektorima, uključujući znanost, zdravstvo, obrazovanje i druge. Predmet ovog rada su metode i pristupi strojnog prevođenja, sa naglaskom na mogućnosti primjene istih u diseminaciji informacija. Ovaj rad istražuje različite načine primjene strojnog prevođenja te njegovu ulogu u olakšavanju pristupa informacijama na više jezika, čime se doprinosi boljoj diseminaciji informacija i znanja.

Svrha rada je pokazati ulogu strojnog prevođenja kao alata koji nadilazi jezične barijere i olakšava prijenos, a samim tim i uporabu informacija sa naglaskom na implikacije, dobrobiti, ali i izazove u primjeni istog. Ciljevi rada su pojmovno razjasniti značenje diseminacije informacija, definirati strojno prevođenje, opisati povijesni razvoj alata za strojno prevođenje, navesti ulogu strojnog prevođenja u diseminaciji informacija, predstaviti primjere korištenja strojnog prevođenja u diseminaciji informacija u različitim područjima, provesti usporedbu izabranih alata za strojno prevođenje u diseminaciji informacija na temelju jednog kraćeg teksta iz znanstvenog rada koji se bavi temom neuronskog strojnog prevođenja te prodiskutirati dobivene rezultate usporedbe.

Očekivani doprinos mog završnog rada leži u jasnijem razumijevanju diseminacije informacija i uloge strojnog prevođenja u tom procesu. Kroz analizu povijesnog razvoja i usporedbu alata za strojno prevođenje, rad će pružiti korisne uvide i praktične primjere korištenja stajnih prevoditelja, što može pomoći u odabiru najprikladnijih rješenja za različite sektore. Doseg rada definiran je fokusom na analizu dostupnih tehnologija za strojno prevođenje te primjena istih u diseminaciji informacija. Posebno se naglašava razumijevanje uloge strojnog prevođenja u olakšavanju pristupa informacijama na više jezika, s ciljem da se istraže konkretni primjeri i alati koji doprinose toj svrsi. Ovaj rad ne pokriva sve aspekte strojnog prevođenja, već se koncentrira na njegovu praktičnu primjenu u nekim sektorima, što omogućava jasniji uvid u neke načine primjene.

Struktura rada obuhvaća uvod, u 1. poglavlju je prikazan pregled literature na temu diseminacije informacija gdje je navedeno nekoliko definicija samog pojma te u 2. poglavlju

metoda širenja informacija. U 3. poglavlju je pojmovno objašnjeno strojno prevođenja koje obuhvaća povijest strojnog prevođenja te su objašnjena različitih vrsta: Strojno prevođenje temeljeno na pravilima – RBMT; Strojno prevođenje temeljeno na primjerima – EBMT; Statističko strojno prevođenje – SMT; te Neuronsko strojno prevođenje – NMT u poglavlju 4. Nakon objašnjenja pojmova strojnog prevođenja te njegovih vrsta i diseminacije, prikazana je veza između ta dva pojma u 5. poglavlju te nekoliko novijih primjera korištenja strojnih prevoditelja na području znanosti, obrazovanja, zdravstva te marketinga i poslovanja u poglavlju 6. Na kraju, u poglavlju 7, je prikazana analiza usporedbe prijevoda koristeći jedan znanstveni tekst iz područja neuronskog strojnog prevođenja i online dostupnih prevoditelja Google Translate, Microsoft Bing, Yandex Translator i Tilde MT te analiza dobivenih rezultata i zaključak.

1 POJMOVNO ODREĐENJE DISEMINACIJE

Diseminacija informacija predstavlja ključan proces u suvremenom društvu, osobito u kontekstu znanstvenog istraživanja i obrazovanja. U doba brzog tehnološkog napretka i sveprisutnog pristupa informacijama, sposobnost učinkovite diseminacije postaje sve važnija za osiguranje da rezultati istraživanja i znanja dođu do široke publike. Ovaj proces ne uključuje samo prenošenje informacija, već i njihovo oblikovanje, prilagodbu i distribuciju različitim ciljnim skupinama. U akademskom kontekstu, diseminacija igra ključnu ulogu u povezivanju istraživača s kolegama, studentima i širim društvom.

U ovom poglavlju bit će predstavljene različite definicije diseminacije, uz naglasak na njezinu važnost u akademskom i društvenom kontekstu. Definiranje ovog pojma omogućava bolje razumijevanje njezine važnosti i uloge koju igra u suvremenom informatičkom okruženju.

Pojam diseminacije odnosi se na čin širenja vijesti, informacija, ideja i drugih sadržaja među širom publikom. Kako navodi Cambridge rječnik, diseminacija je "čin širenja vijesti, informacija, ideja itd. na veliki broj ljudi". Ova definicija naglašava važnost diseminacije u prenošenju znanja i informacija, što je ključno za obrazovanje i informiranje društva.

Prema Merriam-Webster rječniku, diseminacija se opisuje kao "čin ili proces diseminiranja ili širenja nečega". Ovaj proces može uključivati različite metode i alate koji olakšavaju prenošenje informacija, od tradicionalnih medija do suvremenih digitalnih platformi. U kontekstu znanstvenog istraživanja, diseminacija rezultata postaje ključni aspekt svakog znanstvenog rada, jer omogućava širenje stečenih znanja i novih spoznaja široj javnosti.

Nadalje, The Free Dictionary objašnjava da "diseminirati" znači "raširiti široko, kao u sjetvi sjemena" te "širiti; objaviti". Ova etimološka analiza ukazuje na korijene riječi, koji potječu iz latinskog jezika, gdje "dis-" označava razdvajanje ili širenje, a "semināre" znači sijati. Ova slika sjetve može se metaforički primijeniti na način na koji se informacije šire i razvijaju u društvu.

U najširem smislu, diseminacija znači "širenje nečega" ili "prosljeđivanje kakve spoznaje većem broju primatelja". Prema Goranki Blagus Bartolec (2015), u kontekstu znanstvenih istraživanja, "transparentna i učinkovita diseminacija postaje neophodna". Potreba za učinkovitim prenošenjem informacija posebno je važna u današnjem svijetu gdje informacije igraju glavnu ulogu u donošenju odluka i zaključaka. Stoga se izrazi poput "diseminacija informacija", "diseminacija rezultata" ili "diseminacija znanja" mogu zamijeniti riječima poput

"širenje" ili "prosljeđivanje", čime se dodatno pojašnjava važnost ovog procesa u različitim kontekstima. Diseminacija nije samo tehnički proces, ona je ključna za razvoj društva i njegovo informiranje o važnim pitanjima.

Kako bi se bolje razumjelo značenje riječi "diseminacija", Goranka Blagus Bartolec (2015) navodi nekoliko primjera koji su prikazani u Tablici 1.

Tablica 1 Značenja riječi "diseminacija"

diseminacija podataka	širenje/prosljeđivanje podataka
diseminacija (projektnih/statističkih) rezultata	širenje/prosljeđivanje (projektnih/statističkih) rezultata
diseminacija znanja	širenje znanja
projektna diseminacija	širenje podataka o projektu
međunarodna/lokalna projektna diseminacija	širenje podataka o projektu na međunarodnoj/mjesnoj razini
diseminacijski plan	plan širenja podataka (o projektu)
diseminacijska radionica	radionica o širenju podataka (o projektu)

Izvor: Blagus Bartolec 2015, 33

Važno je naglasiti da diseminacija informacija nije isto što i komunikacija informacija. Ključna razlika između komunikacije i diseminacije leži u njihovim funkcijama i pristupima. Komunikacija je interaktivan proces razmjene informacija i ideja između pošiljatelja i primatelja, koji uključuje dijalog i uzajamnu interakciju. U kontekstu znanosti, komunikacija se fokusira na angažiranost sudionika, poticanje povratnih informacija i izgradnju odnosa povjerenja, čime se omogućava aktivno sudjelovanje svih strana. S druge strane, diseminacija se odnosi na jednosmjerno širenje informacija, ideja ili znanja široj publici. Ovaj proces naglašava važnost prenošenja informacija kako bi se educirala i informirala šira zajednica, često koristeći različite metode i alate za širenje informacija do ciljne publike (Vohland et al. 2021, 479). Dok komunikacija teži dvosmjernom dijalogu i suradnji, diseminacija se više fokusira na efikasno prenošenje informacija bez nužnog uključivanja primatelja u proces. Ova razlika ukazuje na to da su obje aktivnosti važne za prenošenje znanja, ali imaju različite ciljeve, ciljne skupine te metode i načine prenošenja informacija i znanja.

1.1 Diseminacija podržana tehnologijom

Diseminacija informacija podržana je tehnologijom na nekoliko ključnih načina, što znatno poboljšava brzinu i učinkovitost širenja podataka. Korištenjem interneta, e-pošte i digitalnih platformi omogućava se trenutna razmjena informacija, što je posebno važno u hitnim situacijama. Informacijsko-komunikacijska tehnologija (skr. IKT) povezuje pojedince, organizacije i vlade širom svijeta, olakšavajući suradnju i potičući globalizaciju u pozitivnom kontekstu. Također, tehnologija omogućava korisnicima pristup velikim količinama informacija putem interneta i mobilnih uređaja, čime se potiče svakodnevno pretraživanje podataka. Interaktivnost društvenih mreža i online platformi omogućava korisnicima dijeljenje informacija i sudjelovanje u raspravama, dok IKT revolucioniraju obrazovanje kroz online tečajeve i digitalne knjižnice. Praćenje i analiza podataka putem tehnologije značajno pridonose donošenju odluka temeljenih na dobivenim podacima, a istovremeno se nastoji smanjiti digitalna podjela između razvijenih zemalja i zemalja u razvoju. Sve ove karakteristike zajedno transformiraju način na koji se informacije pohranjuju, prenose i koriste, utječući na međusobnu komunikaciju i suradnju na globalnoj razini (Shonhe 2017).

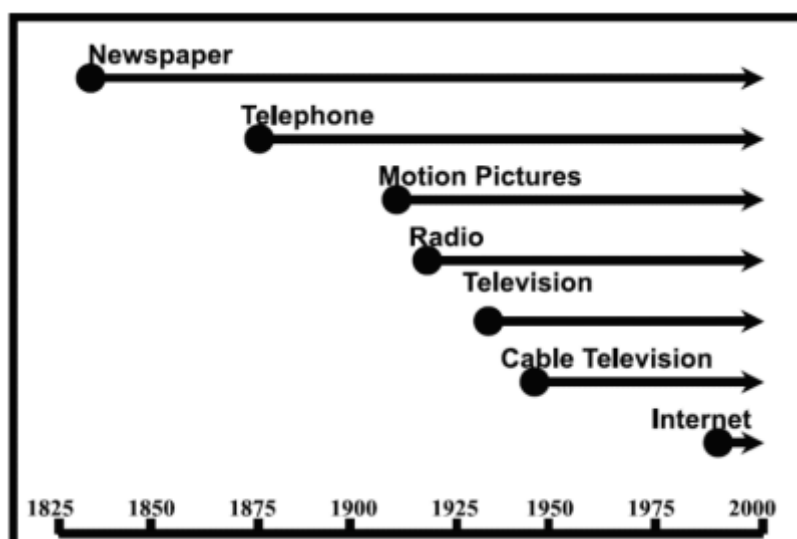
Diseminaciju informacija putem društvenih mreža također možemo objasniti i kroz sljedeću formulu: “ $IDP = (C , M , U , S , T , A)$ ” u kojoj IDP označava proces navedene diseminacije informacija, C označava sadržaj, M označava društvene mreže poput Instagrama, Tik Toka, Twittera i druge, U označava korisnike, S također označava društvene mreže ali odnosi se na manje svjetske mreže ili nasumične mreže, T označava vrijeme diseminacije informacija, i na kraju A koji označava reakcije na informacije koje se šire, poput “lajkanja”, komentiranja, spremanja i ostalo (Liu et al. 2023). Predstavljena formula opisuje sve ključne komponente diseminacije informacija putem društvenih mreža. Zanimljivo je da se pojam "S" koristi za označavanje manje poznatih ili specifičnih mreža, čime se prepoznaje raznolikost medijskih platformi i njihov utjecaj na širenje informacija. Kod diseminacije informacija ključnu ulogu igra vrijeme koje je u ovoj formuli pod oznakom “T” zato što je važno da se informacije šire i primjenjuju pravovremeno. Općenito, možemo zaključiti kako je diseminacija informacija i znanja u suvremenom društvu neodvojiva od informacijsko-komunikacijskih tehnologija.

2 METODE DISEMINACIJE INFORMACIJA

Uloga informacija doživjela je nagli rast sredinom 20. stoljeća, potaknut društvenim i tehnološkim napretkom. Razvoj znanosti i inovacija u tom razdoblju postavio je temelje za eksponencijalno širenje informacija i znanja. Danas živimo u digitalnom dobu gdje su informacije sveprisutne i lako dostupne, zahvaljujući globalnoj umreženosti putem interneta. Ljudi mogu pristupiti informacijama u realnom vremenu, jednostavnim pritiskom na tipku. Ova transformacija promijenila je način na koji se informacije prenose i koriste.

Evolucija diseminacije informacija napredovala je usporedno s razvojem medija. Od tradicionalnih oblika poput tiskanih publikacija, radija i televizije, mediji su evoluirali prema digitalnim platformama. Novine, časopisi, radio i televizija i dalje igraju važnu ulogu, ali sada se njihov sadržaj distribuira i putem interneta. Istovremeno, pojavili su se novi digitalni mediji poput društvenih mreža, blogova i internetskih portala koji omogućuju brzu i široku diseminaciju informacija.

Ova evolucija medija i tehnologija rezultirala je raznolikošću kanala za širenje informacija. Informacije se danas diseminiraju kroz mnoštvo platformi, od tradicionalnih do najmodernijih digitalnih oblika. Kontinuirani razvoj medija i tehnologija omogućio je informacijama da postanu sveprisutne i trenutno dostupne. Slika 1 ilustrira evoluciju u načinima i medijima diseminacije informacija, prikazujući kako su se mediji mijenjali od 1820-ih godina pa sve do pojave interneta i prelaska u digitalne oblike. (Rosmani et al. 2020)



Slika 1 Prikaz razvoja medija za širenje informacija (Rosmani et al. 2020)

Usljed razvoja informacijsko-komunikacijskih tehnologija (skr. IKT-a), došlo je do promjena u prikupljanju, pohrani informacija pa tako i u diseminaciji istih. U nastavku su prikazane suvremenije metode diseminacije informacija kako bi se dobila jasnija slika na koje se sve načine informacije danas šire.

Prve metode diseminacije informacija bile su putem internetskih kataloga s javnim pristupom i baza podataka preko kojih su korisnici mogli prikupiti potrebne informacije u realnom vremenu kada su njima informacije potrebne. Pristup katalogima je moguć sa svih tehnologija kao što su laptop, tablet, mobilni uređaji i drugi, stoga ne moraju biti fizički unutar knjižnice kako bi mogli pristupiti katalogima i bazama podataka. Možemo dati za primjere, javni katalog i bazu podataka PubMed koja "...sadrži biomedicinsku literaturu s više od 26 milijuna citata koji mogu sadržavati cjeloviti tekst" (PubMed, 2017), te "Javna knjižnica u New Yorku, knjižnica Sveučilišta u Liverpoolu, županijska javna knjižnica Jefferson i Nacionalville javna knjižnica..." koje Nalluri i Gaddam (2016) navode kao primjere najbolje prakse korištenja mobilnih kataloga javnog pristupa i baza podataka (Shonhe, 2017).

Osim informacija dostupnih na internetskim katalogima, postoje i aplikacije za mobilne uređaje poput My Library koja korisnicima omogućava pretraživanje informacija, a svrha aplikacije je omogućavanje korisnicima provjeravanje određenog zapisa, pretraživanje stavki, praćenje međuknjižničnih pozajmica te zahtjeva za pristup dokumentima. Također je moguće postaviti e-poštu za obavještanje o novim člancima i knjigama, itd. (Saxena & Yadav, 2013).

Uz prethodno navedene metode, postoje još i suvremene metode širenja informacija koja uključuje QR kodove koji korisnike šalju na određene web lokacije i društvene mreže, poput Instagrama, Tik Toka, Facebook-a i ostale, međutim one su najvećim dijelom namijenjene za marketinške informacije (Shonhe, 2017).

3 STROJNO PREVOĐENJE

U kontekstu diseminacije informacija, strojno prevođenje (eng. Machine Translation - MT) predstavlja pristup koji omogućava brz i efikasan prijenos znanja nadilazeći jezične barijere. Ova cjelina kroz analizu strojnog prevođenja kao alata za diseminaciju informacija, razmatra njegovu primjenu, prednosti i izazove u modernom svijetu. S obzirom na sve veći protok informacija na globalnom nivou, efikasno prevođenje postalo je neophodno za osiguranje razumljivosti i dostupnosti informacija na različitim jezicima diljem svijeta.

Strojno prevođenje, predstavlja istraživačko područje koje kombinira računalnu lingvistiku (eng. Computational Linguistics - CL) i obradu prirodnog jezika (eng. Natural Language Processing - NLP). Slično je procesu razgovora između čovjeka i stroja na prirodnom jeziku, jer generiranje odgovora stroja na prirodnom jeziku također uključuje elemente prevođenja (Šuman, 2021).

Od 1940-ih MT je postao glavna tema mnogih istraživanja (Kirchhoff et al. 2011), a glavni cilj strojnog prevođenja je prevesti tekst s izvornog jezika na ciljni jezik, uz očuvanje originalnog značenja što je bolje moguće. Prema tome, izazov kvalitetnog prijevoda leži u tome da se zadrži smisao originalnog teksta u ciljnom jeziku, uz osiguranje da novi tekst bude prikladno formuliran u skladu s jezikom ciljnog područja, a ne samo doslovno prevedena riječ po riječ (Šuman, 2021).

Europska udruga za strojno prevođenje navodi definiciju po kojoj je strojno prevođenje računalna aplikacija čiji je zadatak prevođenje tekstova s jednog prirodnog jezika na drugi prirodni jezik (AI-Salman, 2004).

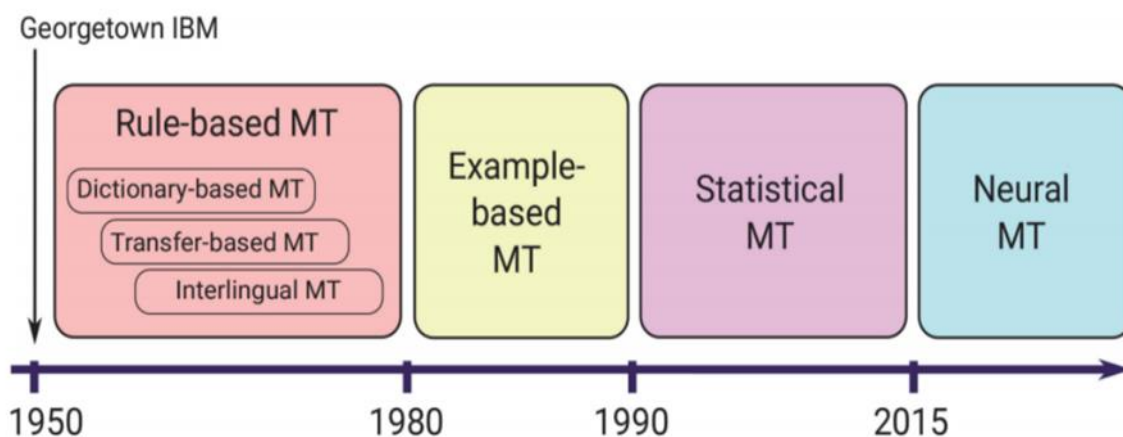
Dahal i Aoun (2023) u svom radu definiraju strojno prevođenje kao računalni sustav dizajniran za izradu prijevoda, sa ili bez uključivanja ljudske pomoći, te naglašava kako je važno razlikovati strojno prevođenje od računalnih alata za prevođenje koji pomažu ljudskim prevoditeljima tako što im omogućuju pristup internetskim rječnicima, bazama podataka i olakšavaju prijevod teksta. Također tvrde da strojno prevođenje koristi računalne algoritme i jezične modele za analizu izvornih tekstova i stvaranje odgovarajućih prijevoda. Za razliku od tradicionalnih metoda prevođenja, koje uvelike ovise o ljudskoj stručnosti i uključenosti, sustavi strojnog prevođenja imaju za cilj osigurati brze i učinkovite automatizirane prijevode.

3.1 Povijest strojnog prevođenja

Područje strojnog prevođenja obuhvaća proces obrade prirodnog jezika koje se bavi pretvaranjem teksta izvornog jezika u tekst ciljanog jezika pomoću automatskih tehnika, a prvi MT sustavi su bili namijenjeni za vojne potrebe (Wang et al. 2022). Slocum (1985) u svom radu navodi kako je početak strojnog prevođenja počeo 1952. godine kada je stvoren “Georgetown Automated Translation”, a onda nakon njega su se pojavili i “Centre d'Études pour la Traduction Automatique”, “Mechanical Translation and Analysis of Languages”, “Traduction Automatique de l'Université de Montréal” i “Automated Language Processing”.

Međutim, Haifeng Wang (et al. 2022) tvrdi kako početak strojnog prevođenja datira još od 1947. godine kada je zapravo stvoreno prvo stolno računalo, te se od tada na strojno prevođenje gleda kao na najizazovniji zadatak u području NLP. U razdoblju od prvih ideja o strojnom prevođenju do 1990-ih godina, metoda strojnog prevođenja temeljena na pravilima (RBMT) bila je dominantna u odnosu na druge metode.

Kroz desetljeća, metode strojnog prevođenja su se značajno razvile, kako je prikazano na slici 2, počevši od jednostavnih pravila zasnovanih na rječnicima, do sofisticiranih neuronskih mreža koje omogućuju visoko precizne prijevode. Ovaj pregled, odnosno slika 2, prikazuje glavne faze razvoja strojnog prevođenja, ističući ključne metodologije koje su oblikovale ovaj dinamični istraživački prostor.

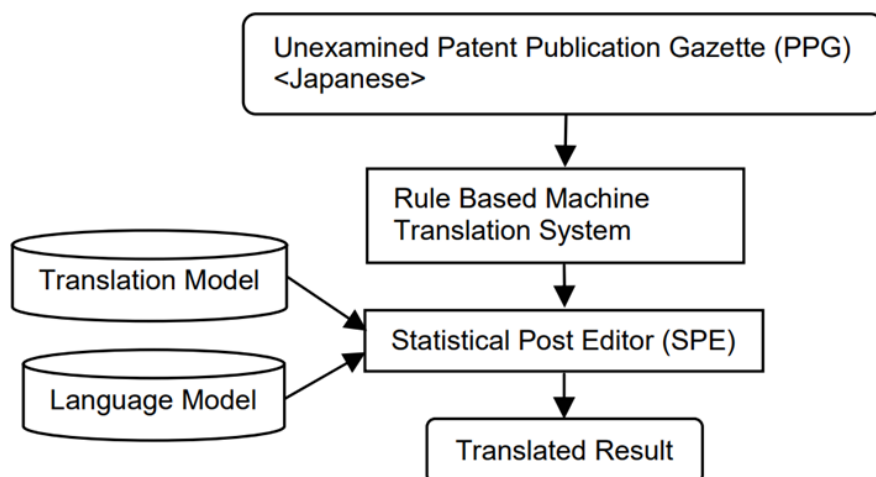


Slika 2 Kronološki prikaz razvoja strojnog prevođenja (Maučec i Donaj 2019, 145)

3.1.1 Strojno prevođenje temeljeno na pravilima

Strojno prevođenje temeljeno na pravilima (engl. Rule Based Machine Translation - RBMT) prvi put se pojavilo 1950-ih godina što je vidljivo na slici 2, a poznato je kao strojno prevođenje temeljeno na znanju ili klasični pristup MT-u, koji koristi lingvističke informacije o izvornom i ciljnom jeziku koje su osigurane putem “dvojezičnih rječnika i ručno pisanih pravila za prevođenje tekstova izvornog jezika u tekstove ciljnog jezika (Wang et al. 2022).”

RBMT sustav pokriva osnovne semantičke, morfološke i sintaktičke strukture svakog jezika, te prima ulazne rečenice na izvornom jeziku i generira izlazne rečenice na ciljnom jeziku putem morfološke, sintaktičke i semantičke analize oba jezika uključena u proces prevođenja. RBMT pristup koristi skup jezičnih pravila u fazama analize, prijenosa i generiranja, zahtijevajući sintaktičku i semantičku analizu te sintaktičko i semantičko generiranje. Nedostaci strojnog prevođenja temeljeno na pravila su sljedeći: nedostatak kvalitetnih rječnika dok se novi stvaraju, potreba za ručnim postavljanjem određenih jezičnih informacija, kompleksnost interakcija pravila u velikim sustavima, izazovi s dvosmislenostima i idiomatskim izrazima, te teškoće u prilagodbi na nove domene (Šuman, 2021).



Slika 3 Proces prevođenja japanskog teksta temeljen na RBMT sustavu (Terumasa 2007, 2)

Proces strojnog prevođenja prikazan na slici 3. prikazuje RBMT sustav koji kombinira pravila i statističke metode kako bi se postigla veća točnost i prirodnost prijevoda, s posebnim fokusom na prijevod japanskih patenata iz neispitanih patentnih biltena. Proces započinje korištenjem sustava strojnog prevođenja temeljenog na pravilima (RBMT), koji prevodi

izvorni japanski tekst na ciljani jezik koristeći unaprijed definirana lingvistička pravila i leksikone. Iako ovaj pristup može pružiti strukturalno ispravan prijevod, često mu nedostaje fleksibilnost u prilagodbi složenim ili idiomatskim izrazima, što može dovesti do neprirodnog ili netočnog prijevoda.

Da bi se prevladala ova ograničenja, prevedeni tekst iz RBMT sustava podvrgava se daljnjoj obradi putem statističkog post-urednika (eng. Statistical Post-Editor - SPE). SPE koristi dva ključna modela za poboljšanje prijevoda: model prijevoda i jezični model. Model prijevoda analizira statističke podatke o tome kako su riječi i fraze prevedene u prošlim primjerima, što omogućuje preciznije prilagodbe postojećeg prijevoda. Jezični model, s druge strane, ocjenjuje gramatičku ispravnost i točnost rečenica, osiguravajući da prijevod zvuči prirodno i da je kontekstualno ispravan.

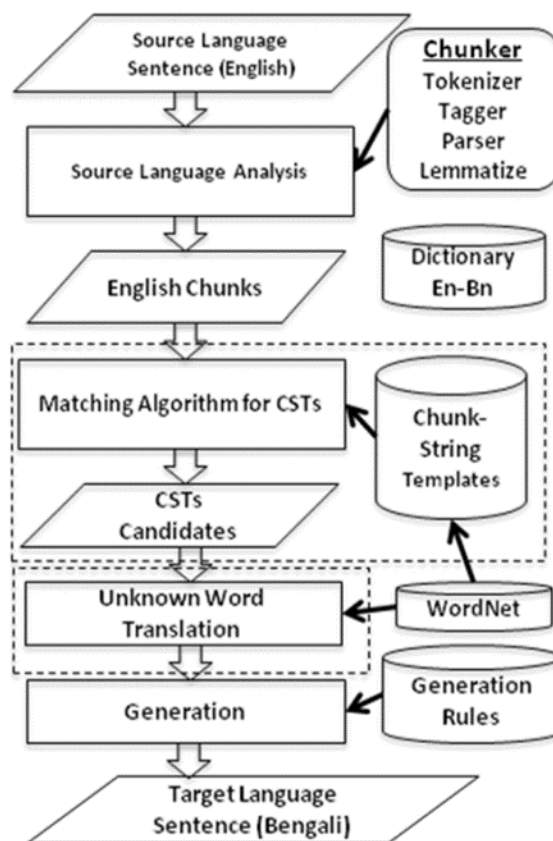
Kombiniranjem ovih pristupa, hibridni sustav omogućuje generiranje konačnog prijevoda koji je ne samo točniji u pogledu značenja izvornog teksta, već je i prirodniji i točniji, što je posebno važno u zahtjevnim područjima poput pravne i tehničke dokumentacije. Na taj način, hibridni pristup nudi najbolje od oba svijeta – preciznost strojnog prevođenja temeljenog na pravilima i fleksibilnost i prilagodljivost statističkih metoda.

3.1.2 Strojno prevođenje temeljeno na primjerima

EBMT (eng. Example-Based Machine Translation) je metoda strojnog prevođenja koja se pojavila 80-ih godina 20. stoljeća te koristi brojne pohranjene podatke/primjere za prevođenje iz dvojezičnog ili višejezičnog korpusa. U korpusu se nalaze uzorci rečenica iz izvornog i ciljnog jezika, koji služe kao primjeri za buduće prijevode. Kada se pojavi potreba za prijevodom novog sadržaja, sustav pretražuje korpus kako bi pronašao odgovarajući uzorak ili bilo koji povezani tekst koji može olakšati prevođenje. Ako sustav otkrije tekst koji još nije preveden ili ne može pronaći bolji primjer prijevoda, korpus se ažurira od strane administratora kako bi bio koristan za buduće potrebe prevođenja. (Anugu i Ramesh, 2020).

Arhitektura EBMT (Example-Based Machine Translation) metode prevođenja, koja je prikazana na slici 4, prikazuje primjer procesa prevođenja engleskih rečenica na bengalski jezik koristeći niz specifičnih koraka. Prvo, ulazna rečenica na izvornom jeziku (engleskom) prolazi kroz analizu izvornog jezika, gdje se koristi alat pod nazivom Chunker. Chunker obuhvaća različite zadatke, poput podjele teksta na manje rečenične jedinice, dodjeljivanje gramatičkih oznaka (poput glagola, predikata, objekta...), analiza sintaktičke strukture rečenice i proces

smanjenja riječi na njihov osnovni oblik, čime se rečenica dijeli na manje dijelove. Navedeni dijelovi se potom uspoređuju s postojećim segmentima iz korpusa pomoću algoritma za podudaranje koji traži odgovarajuće strukture djelova. Nakon identifikacije potencijalnih prijevoda dijelova rečenice, proces prevođenja se nastavlja s prevođenjem nepoznatih riječi koje nisu ranije identificirane u korpusu. Ovdje se koristi baza podataka poput WordNet za pružanje dodatne informacije o značenju nepoznatih riječi. Na kraju, pomoću pravila generacije i odgovarajućih predložaka za djelove rečenica, sustav generira rečenicu na ciljnom jeziku koja predstavlja konačni rezultat ovog procesa. Ova arhitektura omogućuje sustavu da koristi postojeće primjere iz korpusa za prevođenje, dok istovremeno rješava nove izazove u prijevodu putem dodatnih jezičnih resursa i pravila.



Slika 4 Arhitektura EBMT sustava (Salam et al. 2017, 178)

3.1.3 Statističko strojno prevođenje

Statističko strojno prevođenje (eng. Statistical Machine Translation - SMT) se pojavilo 1990-ih godina te je specifičan po tome što se temelji na korištenju velikih korpusa dvojezičnih podataka kako bi se identificirao najvjerojatniji prijevod za dani ulazni tekst. Ovi sustavi funkcioniraju tako da statistički analiziraju odnose između izvornih tekstova i njihovih postojećih ljudskih prijevoda, učeći iz tih podataka kako generirati nove prijevode. (Harison 2023) Osim toga, koristi podatke i kvantitativne metode za analizu odnosa u rečenicama. Za ovakve pristupe potrebni su tekstualni skupovi kako bi se osigurala pouzdanost i razvili modeli koji mogu automatizirati različite jezične kontekste, koristeći parametre za opisivanje jezičnih pojava. Općenito, procesiranje prirodnih jezika (engl. Natural Language Processing - NLP) temeljen na statistici, ne samo da koristi osnovne statističke koncepte iz teorije vjerojatnosti, već uključuje i različite kvantitativne pristupe za automatsku obradu jezika, kao što su modeliranje vjerojatnosti, teorija informacija i linearna algebra (Šuman 2021).

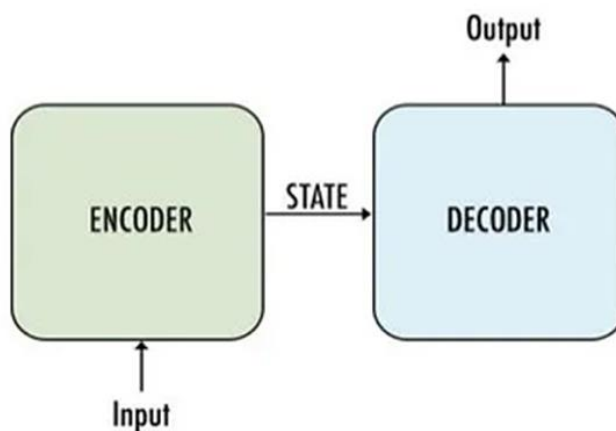
Statističko strojno prevođenje obuhvaća 4 pristupa automatskog prevođenja tekstova. To su prevođenje temeljeno na riječima (Word-based translation), prevođenje temeljeno na frazama (Phrase-based translation), prevođenje temeljeno na sintaksi (Syntax-based translation) i prevođenje temeljeno na jezičnim modelima (Translation based on language models) (AnyChart 2022).

Zbog važnosti koju ima na kvalitetu i brzinu strojnih prijevoda, te općenito na područje strojnog prevođenja, neuronsko strojno prevođenje (NMT) je obrađeno u zasebnoj cjelini.

4 NEURONSKO STROJNO PREVOĐENJE

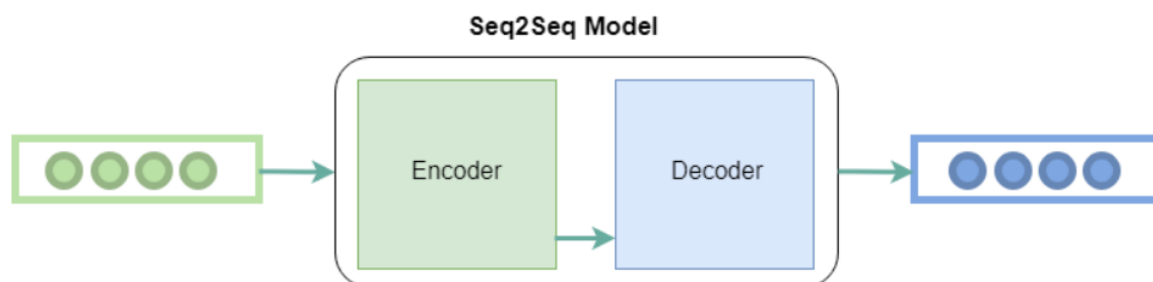
Neuronsko strojno prevođenje (eng. Neural Machine Translation - NMT) predstavlja proces prevođenja rečenice s izvornog jezika na ciljni jezik korištenjem specifičnih modela neuronskih mreža za prijevod. Obično su za prijevode potrebne velike količine paralelnih podataka, no visokokvalitetni podaci tog tipa često su teško dostupni. S obzirom na to da generiranje sintetičkih paralelnih podataka iziskuje mnogo vremena i financijskih resursa, ovaj problem postaje posebno izazovan kod jezika ili područja s ograničenim resursima, gdje su mogućnosti za prikupljanje kvalitetnih podataka ograničene (Oh et al. 2023).

U usporedbi s prethodnim metodama poput pravila temeljenih na strojnom prevođenju (RBMT) i statističkog strojnog prevođenja (SMT), neuronsko strojno prevođenje (NMT) ima značajnu prednost. Za razliku od RBMT-a koji se oslanja na ručno napisana pravila i SMT-a koji zahtijeva ručno odabrane značajke, NMT modeli mogu samostalno naučiti semantičko predstavljanje i znanje prevođenja izravno iz velikih skupova podataka za trening. Ovo je omogućeno zahvaljujući end-to-end arhitekturi NMT-a koja može automatski otkriti i iskoristiti kompleksne obrasce u podacima. Ovaj pristup učenja iz sirovih podataka, bez potrebe za ručnim oblikovanjem značajki, čini NMT znatno efikasnijim i skalabilnijim u odnosu na tradicionalne metode strojnog prevođenja. Model neuronskog strojnog prevođenja (NMT) sastoji se od dvije ključne komponente. Prva komponenta, poznata kao enkoder, koji pretvara izvorni tekst u vektor stvarnih vrijednosti koji predstavlja semantičko značenje rečenice. Nakon toga, druga komponenta, dekoder, koji koristi taj vektor kako bi generirao prevedeni tekst na ciljanom jeziku. Ovaj proces omogućava modelu da efikasno uči i prenosi značenje između jezika, stvarajući tako točan i prirodan prijevod (Wang et al. 2022).



Slika 5 Prikaz Encoder-Decoder arhitekture (Chauhan 2020)

Slika 5 prikazuje encoder-decoder arhitekturu koja se koristi u neuronskom strojnom prevođenju (NMT). Ova arhitektura sastoji se od dva osnovna dijela: enkodera (engl. encoder) i dekodera (engl. decoder). Enkoder prima ulazni tekst na izvornom jeziku i postupno ga obrađuje, pretvarajući svaku riječ u numerički prikaz. Nakon što obradi cijelu rečenicu, enkoder stvara sažeti prikaz te rečenice, poznat kao skrivena faza (engl. state), koje sadrži sve ključne informacije potrebne za razumijevanje značenja rečenice. Taj sažetak se zatim prosljeđuje dekoderu kako bi on mogao generirati prijevod na ciljanom jeziku. Ta skrivena faza služi kao most između enkodera i dekodera, prenoseći informaciju koja omogućava dekoderu da generira odgovarajući izlaz na ciljanom jeziku. Dekoder koristi primljeni prikaz rečenice kako bi proizveo rečenicu na ciljanom jeziku, riječ po riječ, uzimajući u obzir kontekst cijele rečenice. Ova struktura omogućava modelu da zadrži značenje rečenice tijekom procesa prevođenja, osiguravajući veću preciznost i koherentnost u usporedbi s ranijim pristupima strojnog prevođenja. Encoder-decoder arhitektura čini temelj modernih neuronskih modela za prevođenje, kao što su sequence-to-sequence (seq2seq) modeli, te je ključna tehnologija u razvoju naprednih sustava za prevođenje poput Google Translate-a.



Slika 6 Prikaz modela seq2seq (AnyChart 2024)

Seq2seq model, ili model "sekvenca-u-sekvencu", prikazan na slici 6 predstavlja naprednu tehnologiju koja se koristi u području obrade prirodnog jezika, posebno za zadatke poput prevođenja. Ova arhitektura funkcionira tako da prvo analizira ulaznu rečenicu koju želimo prevesti i pretvara je u sažetak, poznat kao "kontekstualni vektor". Ovaj vektor sadrži ključne informacije iz ulazne rečenice, omogućujući modelu da zadrži izvorno značenje rečenice.

Nakon što je rečenica sažeta, seq2seq model koristi taj kontekstualni vektor za generiranje izlazne rečenice na ciljanom jeziku. U ovom procesu, mehanizam pažnje igra ključnu ulogu, omogućujući modelu da se dinamički fokusira na različite dijelove ulazne

rečenice tijekom prevođenja. Ova sposobnost omogućava modelu da bolje razumije odnose između riječi i strukturu rečenice, čime se postiže veća točnost i smislenost u prijevodu.

Seq2seq modeli su značajno unaprijedili kvalitetu strojnog prevođenja, prelazeći s jednostavnog zamjenjivanja riječi na kompleksnije razumijevanje jezika i konteksta. Umjesto da se oslanjaju na doslovno prevođenje, ovi modeli uzimaju u obzir cijelu rečenicu i međusobne odnose riječi, što rezultira prirodnijim i točnijim prijevodima. Ukratko, seq2seq model može se smatrati sofisticiranim alatom koji ne samo da prepoznaje riječi, već i razumije njihovo značenje unutar šireg konteksta (Sai et al. n.d.).

4.1 Strojno prevođenje temeljeno na dubokom učenju

Duboko učenje se odnosi na višeslojne neuronske mreže, koje su naprednije od plitkih metoda strojnog učenja, poput stabala odlučivanja ili standardnih neuronskih mreža s nekoliko slojeva. Napredak u algoritmima i računalnim resursima omogućilo je primjenu dubokog učenja na složene zadatke, posebno u obradi prirodnog jezika i klasifikaciji slika, gdje su duboke neuronske mreže pokazale superiorne rezultate (LeCun et al. 2015). Duboko učenje je način na koji računala mogu učiti iz podataka, slično kao što ljudi uče iz iskustva. Umjesto da im se daju stroga pravila, računala koriste velike količine podataka kako bi sama otkrila obrasce i informacije. Ova tehnika se koristi u raznim područjima, kao što su prepoznavanje slika (npr. prepoznavanje lica), analiza velikih skupova podataka (npr. pronalaženje važnih informacija), prepoznavanje govora (npr. glasovni asistenti) i strojno prevođenje (npr. automatsko prevođenje jezika) (Singh et al. 2017). Osnovne vrste neuronskih mreža uključuju višeslojni perceptron (MLP), konvolucijske neuronske mreže (CNN), povratne neuronske mreže (RNN) i Long Short-Term Memory (LSTM). MLP predstavlja klasičnu arhitekturu u kojoj neuroni u slojevima međusobno komuniciraju, koristeći ponderirane prosjeke i aktivacijske funkcije za uvođenje nelinearnosti, a trenira se putem povratnog širenja pogreške (eng. backward propagation of errors or backpropagation). Svaki neuron izračunava ponderirani prosjek signala koje prima, što znači da se svaki ulazni signal množi s odgovarajućom težinom (ponderom) prije nego što se zbroji. Ovaj izračun omogućuje mreži da odredi koliko je svaki ulaz važan za konačni ishod. CNN su specijalizirane za analizu vizualnih podataka, posebno u području računalnog vida i prepoznavanja slika, koristeći konvolucijske slojeve za ekstrakciju značajki. RNN su dizajnirane za obradu sekvencijalnih podataka, koristeći iste težine za svaki vremenski korak kako bi mogle "pamtiti" informacije iz prošlosti. Međutim, suočavaju se s

problemom slabljenja gradijenta, što otežava učenju dugoročnih ovisnosti jer se gradijenti postaju vrlo mali i neefikasni za ažuriranje težina. Kao rezultat, RNN-ovi često imaju poteškoća s prepoznavanjem informacija koje su udaljene u vremenskim sekvencama. Kako bi se prevladali ovi problemi, razvijene su LSTM arhitekture koje koriste mehanizme za pamćenje informacija kroz vrijeme, omogućujući bolju kontrolu nad protokom podataka i zadržavanje dugoročnih ovisnosti. Ove različite arhitekture omogućuju učinkovito rješavanje složenih problema u raznim domenama, uključujući obradu jezika i analizu vremenskih serija. (Neu et al. 2022).

Osnovna zamisao je bila razviti sustav koji djeluje kao prevoditelj, oslanjajući se na prethodna iskustva i podatke za trening, umjesto da se koristi velikim skupovima pravila. Strojno prevođenje obuhvaća nekoliko međusobno povezanih procesa, uključujući usklađivanje riječi, reorganizaciju rečenica i modeliranje jezika. Svaki od ovih procesa može se rješavati pomoću specifičnih rješenja koja koriste duboke neuronske mreže (DNN).

Singh et al. (2017) u svom radu navode kako se proces prevođenja temeljem DNN-a, sastoji se od 4 koraka:

1. Usklađivanje riječi

Ovaj korak uključuje identifikaciju parova riječi iz izvornog i ciljnog jezika koji su međusobno povezani. U ovom procesu, RNN (povratne neuronske mreže) se često koriste zbog svoje sposobnosti da pamte prethodne informacije i predviđaju točna usklađivanja na temelju povijesti.

2. Odabir pravila i preuređenje

Nakon usklađivanja, odabiru se pravila za prevođenje na temelju usklađenih riječi, a zatim se trenira model za preuređenje rečenica kako bi se osiguralo da struktura rečenice u ciljanom jeziku bude ispravna. Ovdje se koriste metode kao što su povratne neuronske mreže (RNN) koje pomažu u održavanju povijesti i generiranju struktura rečenica.

3. Modeliranje jezika

Ovaj korak uključuje predviđanje vjerojatnosti sljedeće riječi u rečenici na temelju prethodnih riječi. RNN-ovi su posebno učinkoviti za ovu zadaću jer mogu obraditi sekvencijalne podatke i održati kontekst tijekom prevođenja.

4. Zajedničko modeliranje jezika i prevođenja

Ova metoda predviđa ciljne riječi koristeći informacije iz prošlih riječi u izvoru i cilju, čime se dodatno poboljšava točnost prijevoda.

Navedenim koracima, duboke neuronske mreže omogućuju bolje i točnije prevođenje između jezika, smanjujući potrebu za opsežnim skupovima pravila i oslanjajući se na proces učenja iz podataka.

S napretkom dubokog učenja, razvijeni su mnogi otvoreni alati za neuronsko strojno prevođenje (NMT), koji su značajno unaprijedili istraživanja u ovom području. Među najpoznatijim alatima za strojno prevođenje su Tensor2Tensor, FairSeq, NMT, OpenNMT, Sockeye, Nematus, Marian, THUMT, NMT-Keras i Neural Monkey.

Tensor2Tensor (T2T) je biblioteka dubokog učenja koja podržava različite arhitekture za strojno prevođenje, uključujući Transformatore, i omogućava rad na CPU-u, GPU-u i TPU-u.

FairSeq, razvijen od Facebook AI Research-a, omogućuje treniranje prilagođenih modela koristeći RNN i arhitekturu Transformatora. NMT je alat od Google Research-a koji nudi GNMT arhitekturu i vodiče za razvoj modela.

OpenNMT je rezultat suradnje Sveučilišta Harvard i SYSTRAN-a, s implementacijama OPENNMT-PY i OPENNMT-TF te brzim inferencijskim motorom CTRANSLATE2. Sockeye, razvijen od Amazona, koristi se u Amazon Translate-u i nudi moderne modele prevođenja. Nematus podržava RNN i Transformer arhitekture za 13 pravaca prevođenja.

Marian je okvir razvijen od Microsoft Translator tima, dok THUMT uključuje implementacije u Theano, TensorFlow i Pytorch te omogućuje vizualizaciju analize. NMT-Keras se fokusira na napredne primjene sustava strojnog prevođenja, a Neural Monkey je otvoreni sustav temeljen na TensorFlow-u za učenje sekvenci.

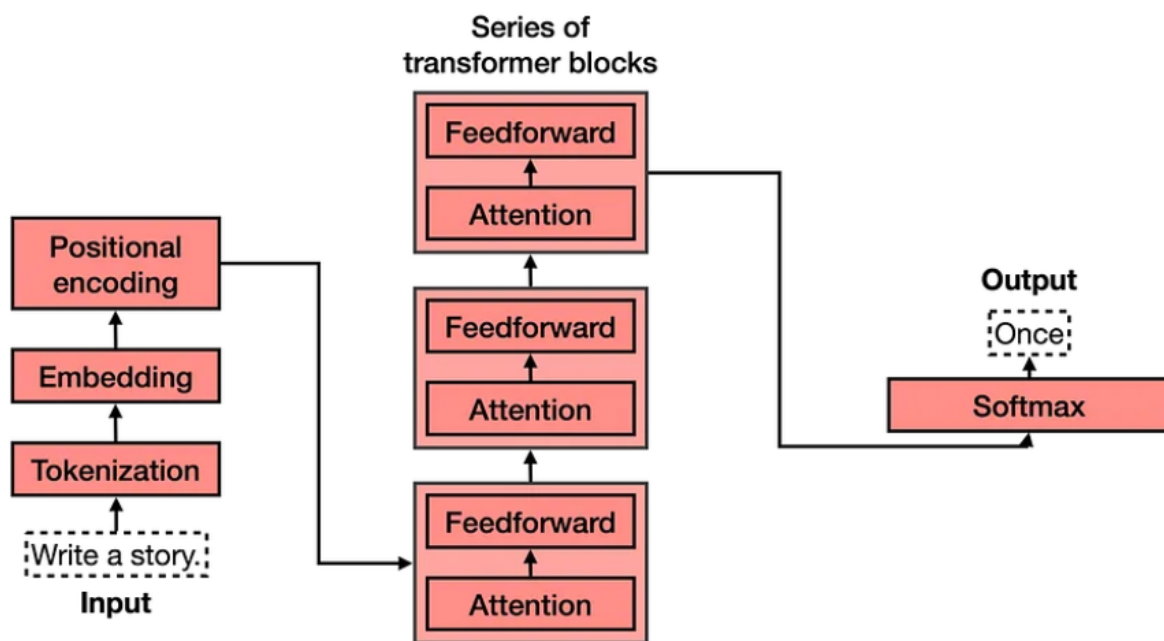
Osim alata za prevođenje, važni su i alati za evaluaciju rezultata. Ručna evaluacija može biti skupa i nepraktična za velike jezične parove, pa se koriste automatske metrike poput BLEU. Alati kao što su SACREBLEU, COMPARE-MT i MT-COMPAREVAL omogućuju analizu i usporedbu rezultata više sustava, čime se poboljšava kvaliteta strojnog prevođenja (Tan et al. 2020).

4.2 Transformatori kod neuronskog strojnog prevođenja

Transformator je nova struktura za neuronsko strojno prevođenje (NMT) koju su predložili Vaswani i suradnici (2017). Ova arhitektura napustila je standardne RNN i CNN strukture i uvela inovativne višeslojne blokove samopažnje, koji su kombinirani s metodom pozicijskog kodiranja. Ovaj pristup koristi prednosti oba modela, omogućujući bolju inicijalizaciju ulazne reprezentacije za različite zadatke obrade prirodnog jezika (NLP). Ključne komponente

transformatora uključuju jedinstvenu reprezentaciju ulaza koja koristi vektore Ključ (Key), Vrijednost (Value) i Upit (Query), kao i višekratnu samopažnju koja omogućava modelu da obraća pažnju na informacije iz različitih reprezentacijskih podprostora (Yang, Wang i Chu 2020).

Struktura transformatora sastoji se od enkodera i dekodera. Enkoder se sastoji od šest identičnih komponenti, svaka s jednim slojem višekratne pažnje, dok je dekodera složeniji, s dodatnim slojevima koji osiguravaju da model ne gleda unaprijed u ciljani niz prilikom predviđanja trenutnih riječi. Zbog značajnog unapređenja učinkovitosti koje pruža transformator, istraživači su se fokusirali na prilagodbe ove arhitekture kako bi prevladali njene slabosti, kao što su nedostatak ponavljajuće strukture i visoka složenost. Različite modifikacije uključuju dublje transformator modele s poboljšanim mehanizmima pažnje, kao i korištenje unaprijed treniranih modela kao što su BERT i GPT, koji su pokazali poboljšanja u raznim zadacima obrade prirodnog jezika (Yang, Wang i Chu 2020).



Slika 7 Arhitektura Transformatora (AnyChart 2023)

Slika 7 prikazuje arhitekturu Transformatora koja je dio neronskog strojnog prevođenja (NMT). Tokenizacija je prvi i osnovni korak u obradi prirodnog jezika (NLP). To je proces koji dijeli tekst na manje dijelove, poznate kao "tokeni". Tokeni mogu biti riječi, interpunkcijski znakovi, prefiksi ili sufiksi. Tokenizacija je ključna jer omogućava računalima da razumiju ljudski jezik. Bez ovog koraka, modeli ne bi mogli analizirati ili interpretirati tekst na način koji je smislen za daljnju obradu, poput prevođenja, sažimanja ili analize sentimenta.

Nakon podijele riječi na tzv. tokene, slijedi korak ugradnje (embedding). Ugradnja pretvara riječi u numeričke vektore, što omogućava računalima da bolje obrade i razumiju tekst. Svaki token se mapira na vektor u visoko-dimenzionalnom prostoru, a slični tokeni imaju slične vektore. Ova tehnika pomaže modelima da hvataju kontekst i odnose između riječi, što je ključno za zadatke poput prevođenja ili analize sentimenta. Pozicijsko kodiranje je sljedeći korak, a to je tehnika koja omogućuje transformatoru da prepozna redosljed tokena u rečenici. Kada se ulazni tokeni pretvore u vektore, pozicijsko kodiranje dodaje unaprijed definirane vektore koji predstavljaju pozicije svakog tokena. Ova metoda osigurava da rečenice s istim riječima, ali različitim redosljedom, dobiju jedinstvene vektore, čime se sprječava da slične rečenice rezultiraju istim vektorom. Na taj način, pozicijsko kodiranje omogućuje modelu da bolje razumije kontekst i odnos između riječi unutar rečenice. (AnyChart 2023)

Funkciju pažnje možemo definirati kao proces koji mapira upit i skup parova ključ-vrijednost na izlaz, pri čemu su upit, ključevi, vrijednosti i izlaz predstavljeni kao vektori. Izlaz se dobiva kao težinski zbroj vrijednosti, gdje se težina koja se dodjeljuje svakoj vrijednosti izračunava koristeći funkciju kompatibilnosti između upita i pripadajućeg ključa (Vaswani et al. 2017).

Softmax sloj je posljednji korak u transformatoru koji pretvara ocjene za sve riječi u vjerojatnosti koje se zbrajaju na 1. Ovaj sloj osigurava da se najviše ocjene pretvaraju u najviše vjerojatnosti, omogućujući modelu da nasumično odabere sljedeću riječ na temelju tih vjerojatnosti. Na primjer, ako transformator dodijeli najvišu vjerojatnost riječi "Jednom" (0.5), a riječi "Negdje" i "Tamo" imaju vjerojatnosti 0.3 i 0.2, tada će se pri uzorkovanju odabrati riječ "Jednom", jer ima najveću vrijednost i time postaje izlaz transformatora. Ovaj proces omogućuje modelu da donese informirane odluke o tome koja bi riječ trebala uslijediti u rečenici, čime se poboljšava kvaliteta generiranog teksta (AnyChart 2023).

5 POVEZANOST STROJNOG PREVOĐENJA I DISEMINACIJE INFORMACIJA

Strojno prevođenje i prevoditeljski alati imaju sve veću ulogu u diseminaciji znanstvenih informacija. Prevođenje sažetaka, ključnih riječi i cijelih članaka na različite jezike može postati uobičajena praksa, povećavajući pristup istraživanjima za širu publiku. Ovo bi ravnomjernije rasporedilo posao prevođenja između govornika engleskog i drugih jezika.

Autori mogu koristiti alate za strojno prevođenje poput DeepL ili Google Translate kao prvi nacrt, a zatim ga ručno urediti. Časopisi mogu pomoći normaliziranjem prakse prevođenja i priznanjem doprinosa prevoditelja. Prijevodi prethodno objavljenih članaka često su ograničeni autorskim pravima, ali mogu se uključiti kao dodatni materijal uz objavu ili pohraniti na dostupnim platformama. Pisanje jednostavnijih sažetaka, manje strukturno složenih ali bogatih značenjem, olakšava prijevod alatima za strojno prevođenje ili ljudskim prevoditeljima.

Časopisi mogu pomoći ublažavanjem ograničenja broja riječi, što često uvodi složenost. Suradnja između autora, časopisa i prevoditelja ključna je za učinkovitu diseminaciju istraživanja na više jezika koristeći strojno prevođenje i prevoditeljske alate. (Steigerwald et al. 2022). Možda je najveći ograničavajući faktor u prijevodu informacija monopol engleskog jezika u prijevodima znanstvenih radova, što stvara nejednakost između izvornih govornika tog jezika i onih kojima engleski nije primarni (materinski) jezik. Istraživanja pokazuju da se ljudi kojima engleski jezik nije izvorni jezik govora, suočavaju s većim izazovima u akademskom pisanju. Alati za strojno prevođenje mogu pomoći u prevladavanju ovih prepreka, omogućujući autorima da komuniciraju na engleskom jeziku s manje poteškoća. Istraživanja su pokazala da kvaliteta MT alata varira ovisno o jeziku i tipu teksta, ali prilagođeni alati za znanstvenu komunikaciju često nadmašuju opće alate poput Google Translate-a.

Nadalje, primjena MT-a u akademskom pisanju pokazuje potencijal za poboljšanje kvalitete i učinkovitosti. Istraživanja su pokazala da pravilno uređeni tekstovi prije prevođenja mogu značajno poboljšati kvalitetu konačnog rezultata. Također, razvoj domenski specifičnih korpusa za MT može poboljšati performanse u prevođenju specijaliziranih znanstvenih tekstova (Macken, De Wilde i Tezcan 2024). U sljedećem poglavlju prikazani su primjeri kako strojno prevođenje doprinosi diseminaciji informacija, posebno u znanstvenom kontekstu, analizirajući njegovu učinkovitost, izazove i buduće smjerove razvoja.

Važno je istaknuti kako nije samo diseminacija znanstvenih informacija i znanja profitirala od strojnog učenja već i strojno prevođenje. Glavni primjer veze između strojnog učenja i strojnog prevođenja je učenje stranih jezika. Online strojno prevođenje (eng. Online Machine Translation - OMT) postalo je popularan alat (u ovom slučaju) među studentima koji ga koriste kao resurs za neovisno učenje jezika. Iako je zabrinjavajuća činjenica da zbog toga neki studenti OMT koriste za pisanje i prevođenje zadataka, važno je napomenuti da se u doba interneta potiče višejezična komunikacija. OMT omogućava korisnicima da komuniciraju na stranim jezicima, često putem društvenih mreža. S obzirom na sveprisutnost OMT-a u svakodnevnom životu, poznate su njegove prednosti i ograničenja kao alata za učenje jezika. Cilj ovog istraživanja je istražiti kako studenti ocjenjuju korisnost rezultata strojnog prevođenja i kako koriste OMT kao resurs za neovisno učenje jezika, uključujući njihove preferencije i izbjegavanja u vezi s razumijevanjem čitanja i slušanja na engleskom jeziku te prevođenjem na ciljni jezik (Niño 2020).

Strojno prevođenje i strojno učenje imaju ključnu ulogu u diseminaciji informacija, posebno u kontekstu sveprisutne digitalizacije. OMT omogućava brzu i učinkovitu komunikaciju na različitim jezicima, čime se olakšava pristup informacijama široj publici. Korištenjem strojnog učenja, algoritmi se neprestano usavršavaju, što rezultira točnijim i kontekstualno relevantnijim prijevodima. Ova tehnologija ne samo da poboljšava kvalitetu komunikacije, već i potiče razmjenu znanja među korisnicima iz različitih jezičnih sredina. Na taj način, strojno prevođenje postaje alat koji ne samo da prevodi riječi, već i širi razumijevanje i interakciju među ljudima. U konačnici, OMT i strojno učenje doprinose globalnoj diseminaciji informacija, omogućujući korisnicima da prevladaju jezične barijere i sudjeluju u globalnom dijalogu.

6 PRIMJERI UPOTREBE STROJNOG PREVOĐENJA U DISEMINACIJI INFORMACIJA

U ovom poglavlju fokusirat ćemo se na primjenu strojnog prevođenja kao ključnog alata u diseminaciji informacija. U globaliziranom svijetu, gdje se informacije šire brže nego ikad, potreba za učinkovitim prevođenjem postaje sve važnija. Strojnim prevođenjem omogućava se pristup znanju i resursima na različitim jezicima, čime se smanjuju jezične barijere i potiče međunarodna suradnja. Kroz analizu konkretnih primjera, istražiti ćemo kako ova tehnologija doprinosi širenju informacija u različitim područjima, uključujući znanost, obrazovanje, zdravstvo, marketing i poslovanje.

6.1 Primjena strojnog prevođenja u području znanstvene komunikacije

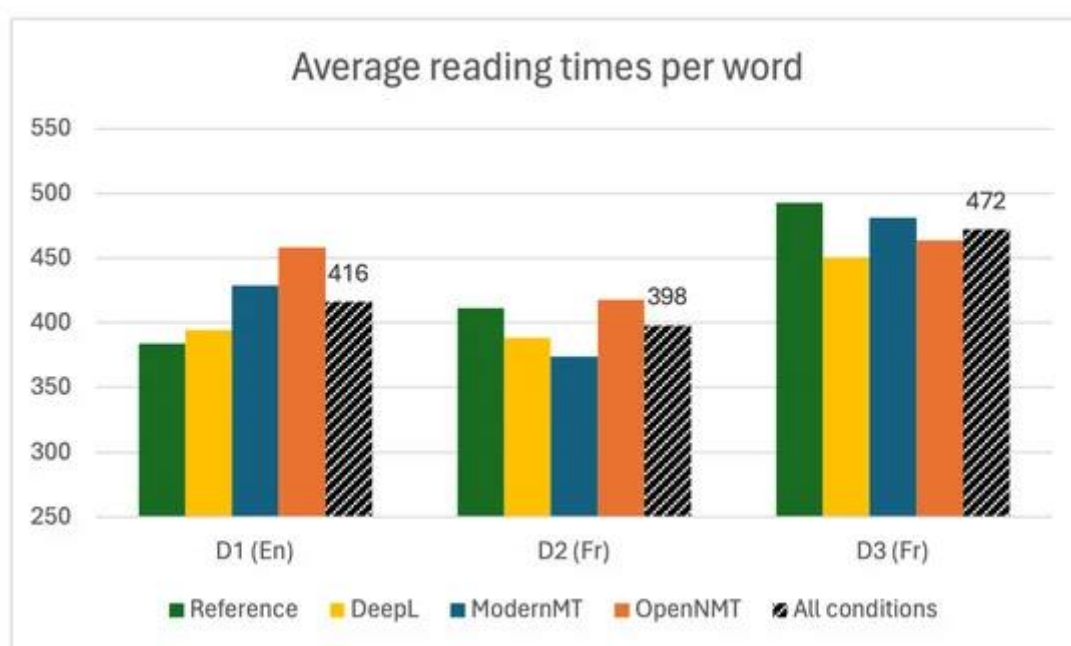
Lieve Macken, Vanessa De Wilde i Arda Tezcan (2024) su proveli istraživanje o primjeni strojnog prevođenja za širenje rezultata istraživanja na različitim jezicima, s posebnim fokusom na jezični par engleski – francuski u tri znanstvene discipline: 1. ljudska mobilnost, 2. neuroznanosti i 3. klimatologija. U istraživanju su korištena tri različita sustava za strojno prevođenje: DeepL, ModernMT i OpenNMT, pri čemu je ModernMT dodatno prilagođen specifičnim domenama. Istraživanje procjenjuje upotrebljivost izlaza strojnog prevođenja kroz tri slučaja upotrebe: za istraživače koji koriste strojne prevoditelje za prevođenje publikacija, profesionalne prevoditelje koji strojne prevoditelje koriste za naknadno uređivanje, te nestručne čitatelje koji žele steći uvid u sadržaj znanstvenih radova. Ovo istraživanje fokusira se na procjenu kvalitete strojnog prijevoda za nestručne čitatelje, koristeći eksperimente čitanja u vlastitom tempu kako bi se prikupili podaci o vremenu čitanja i percepciji kvalitete prijevoda.

Cilj istraživanja je istražiti odnos između kvalitete prijevoda i napora čitatelja, mjerenog vremenom čitanja, kako bi se utvrdilo može li kvaliteta prijevoda predvidjeti vrijeme čitanja.

Rezultati su pokazali sljedeće:

1. Prosječno vrijeme čitanja: Ukupno prosječno vrijeme čitanja bilo je znatno veće od vremena čitanja navedenog u literaturi, s prosječnim vremenom čitanja od 472 milisekunde po riječi za disciplinu 3 (Klimatologija i klimatske promjene), dok su discipline 1 i 2 imale prosječna vremena od 416 i 398 milisekundi po riječi.

2. Točnost odgovora: Sudionici su dali 52% netočnih odgovora na pitanja s višestrukim izborom o razumijevanju u disciplini 3, dok su discipline 1 i 2 imale znatno manje netočnih odgovora (11% i 25%).
3. Percipirane ocjene upotrebljivosti: Ocjene upotrebljivosti prijevoda bile su dobre za discipline 1 (81%) i 2 (81%), dok je disciplina 3 imala nižu ocjenu (67%).
4. Učinkovitost različitih vrsta prijevoda: U disciplini 1, ljudski referentni prijevodi čitani su brže od strojnog prevođenja. DeepL je postigao najbolju ocjenu (89%) za disciplinu 1, dok je OpenNMT imao znatno lošiji rezultat (65%) za disciplinu 2, a za treću disciplinu, razlike između različitih uvjeta prevođenja bile su minimalne.



Slika 8 Prikaz prosječnog vrijeme čitanja (Macken, De Wilde i Tezcan 2024)

6.2 Primjena strojnog prevođenja u zdravstvu

Dahal i Aoun (2023) u svom radu objašnjavaju na koji način strojno prevođenje pomaže u širenju važnih informacija u zdravstvenim ustanovama. Tvrdi kako strojno prevođenje, u zdravstvenim ustanovama, olakšava točnu i pravovremenu razmjenu ključnih informacija, kao što su zdravstveni zapisi, dijagnostička izvješća i planovi liječenja, između zdravstvenih djelatnika i pacijenata. Omogućuje liječnicima i medicinskim sestrama da komuniciraju s pacijentima koji govore različite jezike, poboljšavaju pružanje zdravstvenih usluga i

poboljšavaju ishode pacijenata. Uklanjanjem jezičnih prepreka između zdravstvenih djelatnika i pacijenata, strojnim prevođenjem pacijenti dobivaju odgovarajuću pomoć i skrb i razumiju upute i informacije dobivene od zdravstvenih djelatnika. Ističu kako ograničena jezična pismenost određenih segmenata populacije otežava učinkovitu komunikaciju kritičnih informacija o bolestima, preventivnim mjerama, kampanjama cijepljenja i zdravstvenopromicateljskim ponašanjima. Međutim, pojavom strojnog prevođenja došlo je do značajne transformacije koja je promijenila način širenja materijala za zdravstveno obrazovanje i podizanje svijesti u lingvistički raznolikim zajednicama.

Dew (2018) u svom radu nastoji opisati načine na koje se strojno prevođenje razvija kako bi se uklonile jezične prepreke u zdravstvenim ustanovama na temelju sustavnog pretraživanja i kvalitativne analize 27 studija. Ciljni korisnici za aplikacije strojnog prevođenja bili su isključivo pacijenti koji ne govore dominantan jezik svojeg medicinskog okruženja. Većina tih studija fokusirala se na jednosmjerne prijevode s jednog izvornog jezika na jedan ciljni jezik, pri čemu je englesko-španjolski jezik najčešći par jezika. Od analiziranih 27 studija, 11 ih je opisalo da je cilj sustava za strojno prevođenje poboljšati komunikaciju između pružatelja zdravstvenih usluga/osoblja i pacijenata u kliničkim situacijama u kojima postoji jezična barijera.

Oba rada naglašavaju kako strojno prevođenje uklanja jezične prepreke koje otežavaju razmjenu važnih informacija između zdravstvenih radnika i pacijenata, čime se poboljšavaju ishodi liječenja i pružanje zdravstvenih usluga. Dahal i Aoun (2023) ističu da strojno prevođenje omogućava točnu i pravovremenu razmjenu zdravstvenih zapisa i dijagnostičkih izvješća, dok Dew (2018) naglašava da se većina istraživanja fokusira na jednosmjerne prijevode, posebno englesko-španjolski jezik, kako bi se poboljšala komunikacija u kliničkim situacijama.

Oboje su došli do zaključka da strojno prevođenje predstavlja značajan alat za smanjenje zdravstvenih nejednakosti i poboljšanje pristupa zdravstvenim informacijama za lingvistički raznolike populacije. Iako postoje izazovi u vezi s točnošću prijevoda, posebno u vezi s medicinskom terminologijom, strojno prevođenje se pokazuje kao učinkovito rješenje za osnaživanje pacijenata i omogućavanje bolje zdravstvene skrbi. Oba istraživanja sugeriraju da bi se strojno prevođenje trebalo koristiti kao dopunski alat uz ljudske tumače kako bi se osigurala točnost i kvaliteta komunikacije u zdravstvenim ustanovama.

6.2.1 Strojno prevođenje u doba COVID-19 pandemije

Dana Khalid Mahadin i Sameer Naser Olimat (2022) su proveli anketu na profesionalnim jordanskim prevoditeljima. Ukupno je bilo 106 ispitanika. Svrha ove ankete je bila istražiti stajališta i mišljenja jordanskih profesionalnih prevoditelja o korištenju strojnih prevoditelja tijekom COVID-19 pandemije kada je širenje informacija o pandemiji bilo na vrhuncu te je bilo neophodno prevođenje uputa o zaštiti od same bolesti. Prevoditelji su informacije i druge materijale o COVID-19 virusu prevodili sa drugih jezika na arapski jezik. Anketa se sastojala od 3 dijela.

Prvi dio odnosi se na korištenje strojnih prevoditelja za COVID-19 informacija, gdje su rezultati pokazali da MT nije potpuno pouzdano rješenje u ovom slučaju jer nema semantička, sintetička ni kulturna obilježja. „Na primjer, arapski prijevod MT „imuniteta stada” nepoželjan je, a prednost se daje „imunitetu zajednice (Mahadin i Olimat 2022, 37)” Riječ “stada” u ovom primjeru je nepoželjan iz razloga što je korištenje životinjskih izraza, u arapskoj kulturi, neprihvatljivo te se smatra i uvredom arapskom narodu. Drugi dio, u kontekstu ovog rada, nije važno spomenuti jer se odnosi na korištenje ili izradu rječnika s terminologijom COVID-19 bolesti. Nas zanima treći dio koji se odnosi na korištenje strojnih prevoditelja za prijevod COVID-19 terminologije na arapski jezik. Rezultati trećeg dijela ankete su pokazali da je 60% ispitanika imalo problema u korištenju MT-a za prevođenje informacija. Mahadin i Olimat (2022) pretpostavljaju da je razlog tome brzina uvođenja novih termina/izraza te nedostatak sadržaja na internetu na arapskom jeziku. Također navode kako se otvara pitanje o jednakosti otvorenih informacija i njihovim pristupima te kako dominiraju informacije na europskim jezicima.

Međutim, imamo još jedan primjer gdje se strojno prevođenje nije pokazalo kao najbolja opcija za prevođenje informacija o COVID-19 bolesti. Naime Zakaryia Almahasees (et al. 2021) su proveli istraživanje gdje su testirali Google Translate, s engleskog jezika na arapski. Rezultati njihovog istraživanja su pokazali kako Google Translate ima najviše semantičkih grešaka što onemogućava razumijevanje teksta, a nakon toga slijede i gramatičke i leksičke greške. Također se pokazalo i da neke riječi krivo prevodi.

Iz oba prikazana primjera možemo doći do zaključka kako strojni prevoditelji nisu korisni u kriznim zdravstvenim situacijama, u ovom slučaju je to COVID-19 pandemija, iz razloga što korišteni prevoditelji ulazne tekstove prevode doslovno od riječi do riječi bez

ikakve smislenosti, te neke pojmove čak i prevodi krivo i daje im skroz neko drugo značenje unutar rečenice.

6.2.2 Strojno prevođenje njemačkih zdravstvenih tekstova

Silvana Deilen (et al. 2024) su proveli istraživanje na 30 tekstova s web stranice njemačkog zdravstvenog magazina Apotheken Umschau, koji pokrivaju razne teme poput uboda insekata, cijepljenja, cistitisa i trovanja hranom. Za sve tekstove već su postojali ljudski prijevodi na jednostavnom jeziku, koje su pregledali stručnjaci iz medicine i farmacije, osiguravajući točnost sadržaja. Ovi tekstovi su zatim prevedeni pomoću sustava strojno prevođenje SUMM AI. U analizi su uspoređeni strojno prevedeni tekstovi s ljudskim prijevodima i izvorima, koristeći tri kriterija: točnost sadržaja, čitljivost i sintaktička složenost. Prvi kriterij primijenjen je samo na strojne prijevode, dok su ostala dva korištena za sve tri skupine (izvorni tekstovi, ljudski prijevodi i strojni prijevodi).

Rezultati istraživanja točnosti strojnih prijevoda pokazuje da je samo 1 od 30 analiziranih tekstova bio ispravno preveden. Preostalih 29 tekstova imalo je različite probleme, što ukazuje na značajnu varijabilnost i neusklađenost rezultata. Tekstovi na kojima se provelo istraživanje ne prate dosljednu strukturu i nisu usmjereni na praktičnu primjenu, što zahtijeva potpunu opsežnu doradu. Uočene su gramatičke i pravopisne pogreške, propušteni prefiksi ili negacije, netočna objašnjenja tehničkih izraza, nepotpuni popisi te kontradiktorne tvrdnje.

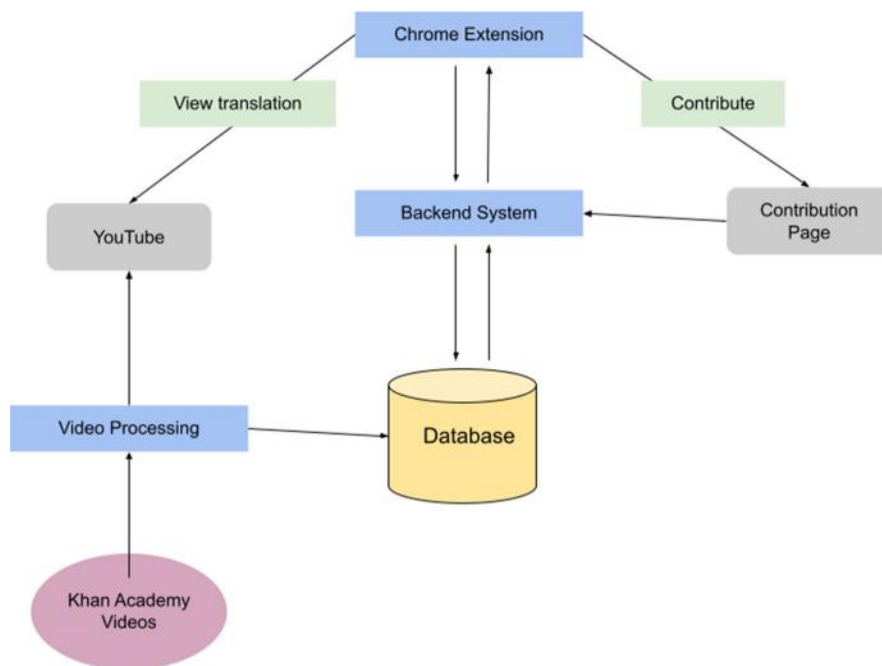
Važno je spomenuti da nije provedena kvantitativna analiza, jer su pogreške same po sebi predstavljale rizik za primarne korisnike. Što se tiče razumljivosti tekstova, usporedba ljudskih i strojnih prijevoda, kao i izvornih tekstova, pokazala je da su strojni prijevodi bili najrazumljiviji, s prosječnom HIX vrijednošću (mjera za razumljivosti teksta) od 19,15. Ljudski prijevodi imali su prosjek od 17,74, dok su izvorni tekstovi bili najmanje razumljivi s prosjekom od 10,46. S obzirom na malu varijancu u strojnom prijevodu, svih 30 tekstova može se klasificirati kao tekstovi na jednostavnom jeziku. Međutim, samo 83% ljudskih prijevoda dostiglo je standard Plain German. Ljudski prijevodi pokazuju veću varijaciju u HIX vrijednostima nego strojni. Važno je napomenuti da HIX vrijednosti uzimaju u obzir samo očitu složenost, pa ih treba dopuniti dodatnom kvalitativnom analizom za potpuniju procjenu razumljivosti. Na kraju istraživanja, analizirana je raspodjela sintaktičkih odnosa u ljudskim i strojno prevedenim tekstovima, kao i u izvorima. Rezultati pokazuju da izvorni tekstovi i strojni prijevodi imaju veći broj složenih sintaktičkih odnosa u odnosu na ljudske prijevode,

koji su jednostavniji i pokazuju veći broj parataktičkih odnosa. Izvorni tekstovi sadrže više klauzalnih subjekata i klauza koje modificiraju imenice nego strojni prijevodi, dok ljudski prijevodi ne sadrže klauzalne subjekte. U strojno prevedenim tekstovima dominiraju klauzalni komplementi glagola i pridjeva, kao i klauzalni komplementi bez subjekta, koji su značajno prisutniji nego u ljudskim prijevodima i izvorima. Ovi rezultati ukazuju na to da su strojni i izvorni tekstovi složeniji od ljudskih prijevoda, koji su najjednostavniji među analiziranim podkorpisima.

6.3 Primjena strojnog prevođenja u obrazovanju

Linden Wang (2024) je proveo istraživanje o automatiziranom prevođenju edukacijskih videozapisa sa YouTube kanala „Khan Academy“. Jedan od izazova predstavlja izravno prevođenje govora, koje se suočava s problemima poput niske kvalitete, vremenskih kašnjenja i nedovoljne točnosti zbog nedostatka konteksta. Češći pristup uključuje korištenje automatskog prepoznavanja govora (ASR) za pretvaranje zvučnog zapisa u pisani tekst, koji se potom prevodi. YouTube videozapisi su prevedeni primjenom dva različita sustava za strojno prevođenje, Google Translate i DeepL, na kineski i španjolski jezik.

Rezultati su pokazali da prijevodi na španjolski daju veću točnost od onih na kineskom, bez obzira na to koristite li Google Translate ili DeepL. Izvedba ostaje stabilna u različitim jezičnim modelima, temama i razinama sadržaja kada ciljate na isti jezik. Na primjer, Google Translate dosljedno nadmašuje DeepL za španjolske prijevode, dok DeepL odgovara ili malo nadmašuje Google Translate za kineski. Kako se složenost materijala povećava, postupno se smanjuje točnost prijevoda za čitanje videa. To je u skladu s autorovim očekivanjima, budući da zahtjevniji materijal može komplicirati proces prevođenja. Tijekom ručne analize prijevoda za čitanje uočene su određene situacije u kojima je sadržaj niže razine doveo do smanjene točnosti prijevoda, često zbog sadržaja koji se razlikuje u nastojanju da se privuče mlađa publika. Očito je da, kada su u pitanju matematički videozapisi prevedeni na kineski, Google Translate pokazuje značajno poboljšanje u točnosti prijevoda u odnosu na DeepL unutar kategorije linearne algebre. To ukazuje da bi u ovom konkretnom slučaju Google prevoditelj mogao poslužiti kao učinkovitiji model prevođenja. Za španjolske prijevode postiže se stopa točnosti od približno 86%-95%. Međutim, postotak točnosti za kineske prijevode je relativno niži, u rasponu od 62% do 77%.



Slika 9 Prikaz postupka prevođenja edukacijskih videozapisa (Wang 2024)

Nadalje, Celia Rico i Diana Gonzalez Pastor (2022) su proveli istraživanje o učinkovitosti automatiziranog strojnog prevođenja, u kontekstu obrazovanja i primjene u učionici. Istraživanje je provedeno kako bi se razumjeli stavovi edukatora prevođenja prema strojnim prijevodima, njihovoj primjeni u nastavnom procesu te kako bi se istražile mogućnosti poboljšanja kvalitete prijevoda kroz korištenje suvremenih tehnologija. Cilj je bio procijeniti kako strojno prevođenje može unaprijediti obrazovne metode i olakšati proces učenja jezika, kao i identificirati izazove i prednosti koje ova tehnologija donosi. U konačnici, istraživanje je imalo za cilj razvoj boljih pristupa u obrazovanju prevoditelja i integraciji strojnog prevođenja u nastavi.

Istraživanjem je otkriven važan element u vezi sa stavovima predavača o uključivanju strojnog prevođenja (MT) u njihovu nastavnu praksu: značajan broj nastavnika nema formalnu obuku za MT, što izravno utječe na njihovo razumijevanje i stajališta o ovoj tehnologiji. Njihovo razumijevanje MT-a uvelike je oblikovano osobnim iskustvima povremenih korisnika, što ometa njihovu sposobnost da postanu kritični iiskusni korisnici alata za strojno prevođenje. U idealnom slučaju, nastavnici bi trebali funkcionirati kao "praktičari-istraživači", spajajući praktično iskustvo profesionalnog prevoditelja s teorijskim uvidima i metodologijama. Ovo spajanje omogućuje usklađivanje nastavnih strategija sa zahtjevima tržišta i razvojem krajolika visokog obrazovanja. Istraživanje je ukazalo na to da su edukatori, iako su neformalni korisnici strojnog prevođenja (MT), pozitivno nastrojeni prema ovoj tehnologiji. Oni su svjesni mogućih

nedostataka i izražavaju zabrinutost kako bi MT mogao utjecati na ulogu prevoditelja, pa čak i dovesti do njihovog nestanka. Iako postoji opće slaganje o značaju ljudskog faktora u procesu prevođenja, teško je precizno odrediti njegovu poziciju prema mišljenju sudionika.

6.4 Primjena strojnog prevođenja u poslovanju

Strojno prevođenje omogućuje tvrtkama brže i učinkovitije prevođenje velikih količina promjenjivih tekstova. Ovo donosi nekoliko ključnih prednosti:

- Brže stavljanje proizvoda na tržište: Proces prevođenja se ubrzava, što skraćuje ukupno vrijeme potrebno da proizvod stigne na tržište.
- Bolja interna komunikacija: Zaposlenici u različitim zemljama mogu lakše pristupiti važnim dokumentima na svom jeziku, što poboljšava razmjenu informacija unutar tvrtke.
- Poboľšano korisničko iskustvo i učinkovitost zaposlenika: Omogućuje se brzo i učinkovito prevođenje u stvarnom vremenu, što postaje standardni dio poslovanja, olakšavajući rad zaposlenicima i poboljšavajući iskustvo korisnika. (Sakre 2019)

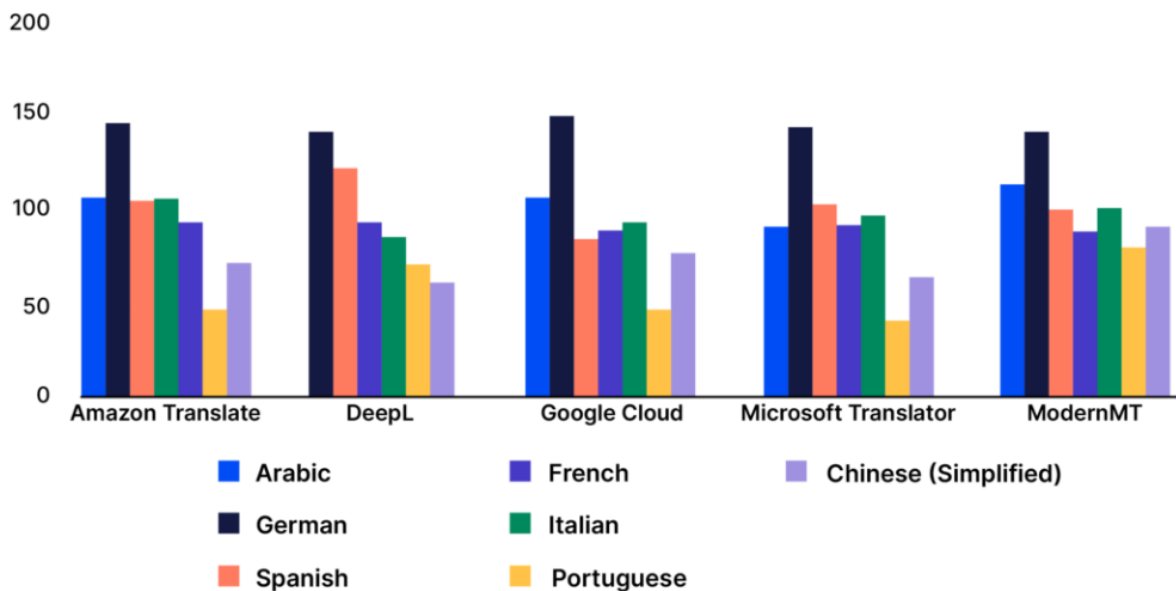
Internet je omogućio ogroman porast objavljivanja članaka na različitim jezicima. Kako bi se nosila s ovim izazovom, industrija je razvila softver za prevođenje koji omogućava:

1. Ubrzano širenje informacija: Strojno prevođenje može brzo prevesti članke, čime se smanjuje vrijeme potrebno da informacije postanu dostupne na drugim jezicima. Iako automatski prijevodi nisu uvijek potpuno točni, mogu se lako ispraviti i omogućiti brži pristup informacijama.
2. Povećanje dostupnosti informacija: Ljudi mogu lakše čitati i razumjeti informacije na stranim jezicima. Tražilice s ugrađenim alatima za prevođenje mogu pretraživati sadržaje na više jezika, čime se olakšava pronalaženje informacija na globalnoj razini (Sakre 2019).

Često se strojno prevođenje koristi u marketingu. Weglot (blog) i jezični konzultanti iz Nimdzi Insights proveli su istraživanje o točnosti softvera za strojno prevođenje (MT) i njegovom utjecaju na prevođenje marketinškog sadržaja, uključujući web stranice. Cilj studije bio je razjasniti uobičajene mitove o MT-u i utvrditi njegovu ulogu u projektima lokalizacije (Pokorny 2024).

U istraživanju koje je provela Elizabeth Pokorny (2024), uspoređeno su pet vodećih pružatelja strojnog prevođenja – Amazon Translate, DeepL, Google Cloud, Microsoft Translator i ModernMT. Testirano je 168 segmenata s 145 dijelova koji nisu trebali nikakve izmjene od strane profesionalnih lingvista nakon prevođenja, dok je portugalski imao samo 58 dijelova bez izmjena. Prevodilo se više od 1,000 riječi s američkog engleskog na nekoliko jezika koje možemo vidjeti na slici 10. Profesionalni lingvisti ocijenili su 85% prijevoda kao 'Vrlo dobro' ili 'Prihvatljivo', dok nijedan nije bio ocijenjen kao 'Vrlo loš'. Njemački jezik pokazao se kao najtočniji, dok je talijanski bio najizazovniji za prevođenje.

Rezultati sugeriraju da strojno prevođenje postaje korisno sredstvo za marketing stručnjake koji žele proširiti svoje poslovanje na globalnoj razini. Weglot integrira MT u svoje projekte prevođenja web stranica, a korisnici su zadovoljni brzinom i točnošću alata. Zaključak studije je da MT sada dopunjuje rad ljudskih lingvista.



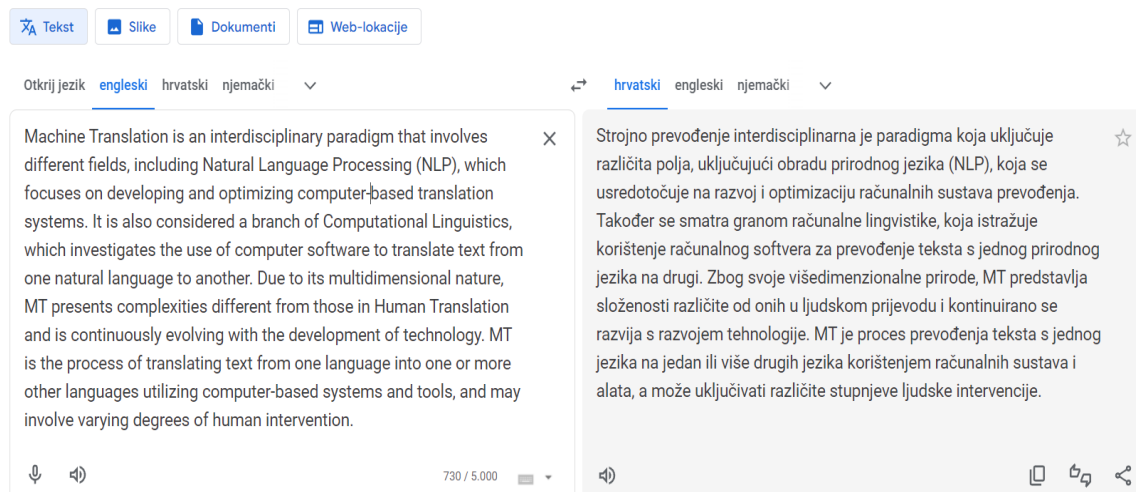
Slika 10 Prikaz jezika na kojima je provedeno istraživanje (Pokorny 2024)

7 USPOREDBA ALATA ZA STROJNO PREVOĐENJE U DISEMINACIJI INFORMACIJA

U ovom poglavlju analizirat ćemo usporedbu prijevoda pomoću četiri popularna online prevoditelja: Google Translate, Microsoft Bing Translator, Yandex Translate i Tilde MT. Svaki od ovih alata nudi jedinstvene značajke i pristupe prevođenju, a njihova učinkovitost može varirati ovisno o jeziku, kontekstu i vrsti teksta koji se prevodi.

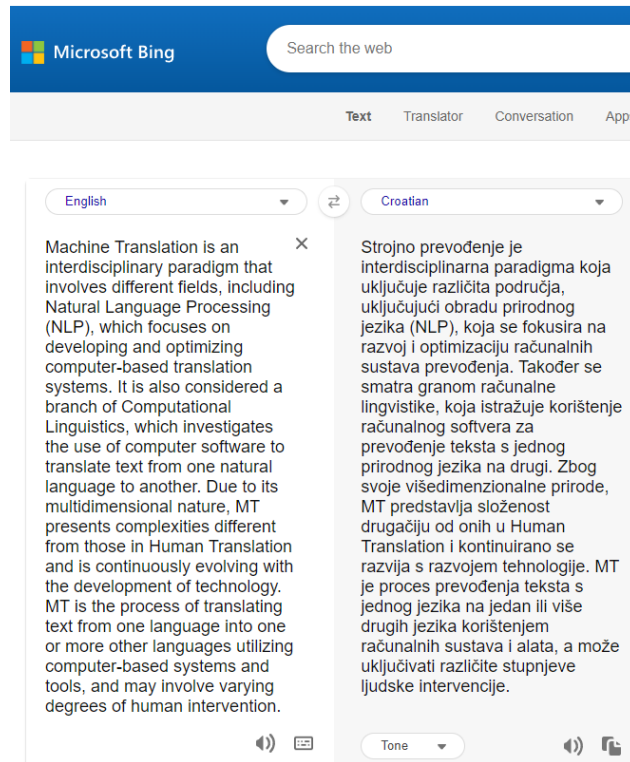
U ovoj analizi, provest ćemo testiranje svakog od ovih prevoditelja na istom izvornom tekstu, kako bismo procijenili kvalitetu prijevoda, točnost i upotrebljivost rezultata. Osim toga, razmotrit ćemo prednosti i nedostatke svakog alata te pružiti pregled rezultata. Ova usporedba ne samo da će prikazati razlike među alatima, već će također pružiti uvid u to kako tehnologija oblikuje rečenice na druge jezike. U ovoj analizi će se prevoditi sa engleskog jezika na hrvatski jezik. Izvorni tekst koji ćemo koristiti u ovoj analizi je:

"Machine Translation is an interdisciplinary paradigm that involves different fields, including Natural Language Processing (NLP), which focuses on developing and optimizing computer-based translation systems. It is also considered a branch of Computational Linguistics, which investigates the use of computer software to translate text from one natural language to another. Due to its multidimensional nature, MT presents complexities different from those in Human Translation and is continuously evolving with the development of technology. MT is the process of translating text from one language into one or more other languages utilizing computer-based systems and tools, and may involve varying degrees of human intervention (Fakih et al. 2024, 215)."



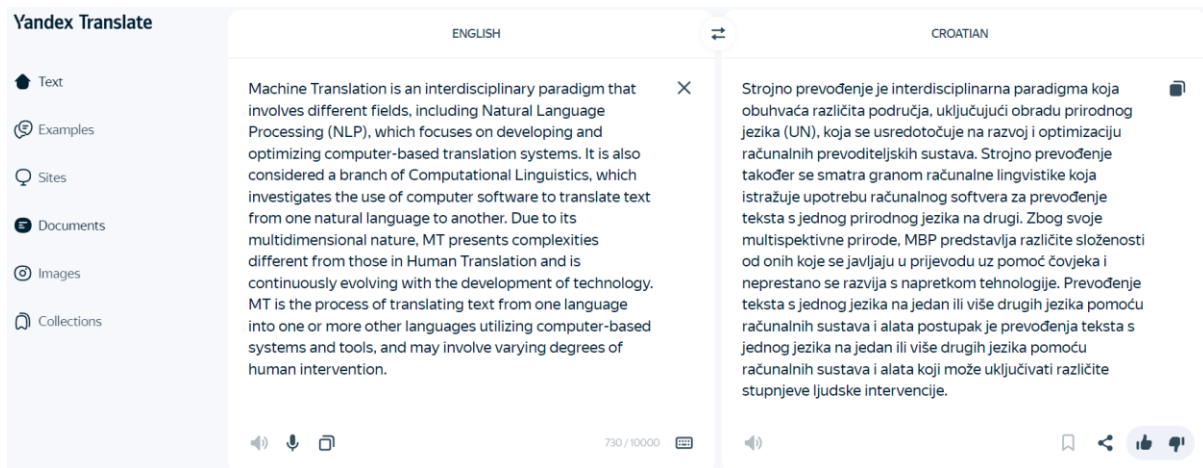
Slika 11 Prikaz prijevoda teksta - Google Prevoditelj

Kako je prikazano na slici 11, Google prevoditelj optimizira tekst na način da ga prijevodom čini kompaktnijim. Bira riječi koje su kraće, iako su istog korijena. Moguće je da nema alternativnih prijevoda za određene izraze u bazi podataka, jer koristi anglicizme kada je to izbježno radi principa uštede prostora, što rade i sistemi Microsoft Bing, te Tilde MT. Vrijedi pretpostaviti da je zbog slabo dostupnih rječnika hrvatskog jezika u odnosu na široko dostupni izvor engleskog prijevod ograničen, te je izbor riječi mnogo uži. Google prevoditelj također, za razliku od većine svojih konkurenata, pravi neobičnu inverziju glagola. Iako bi pojedinci smatrali korištenu inverziju formalnijom i stoga prikladnijom za tekstove, takva inverzija se kosi s tempom čitanja teksta i stoga oduzima od jasnoće i pamtljivosti rečenice. Istu inverziju čini i Tilde MT čime oba alata ukazuju na nerazumijevanje ljudskog procesiranja informacija. Pitanje interpunkcije, za razliku od Yandexa i Tilde MT-a, Bing i Google prevoditelj rješavaju "lijepljenjem" rečeničnih znakova na ista mjesta kao u engleskom tekstu, iako se zarezi u hrvatskom tu ne bi koristili, te ne mijenjaju točke u rečenicama u uskliknike unutar Hrvatskog teksta. Jedna od prednosti Google prevoditelja, ali i Binga je opcija pretvaranja teksta u govor što omogućava prevoditelju auditornu verziju prijevoda, no kvaliteta varira ovisno o dostupnosti govora na odabranom jeziku.



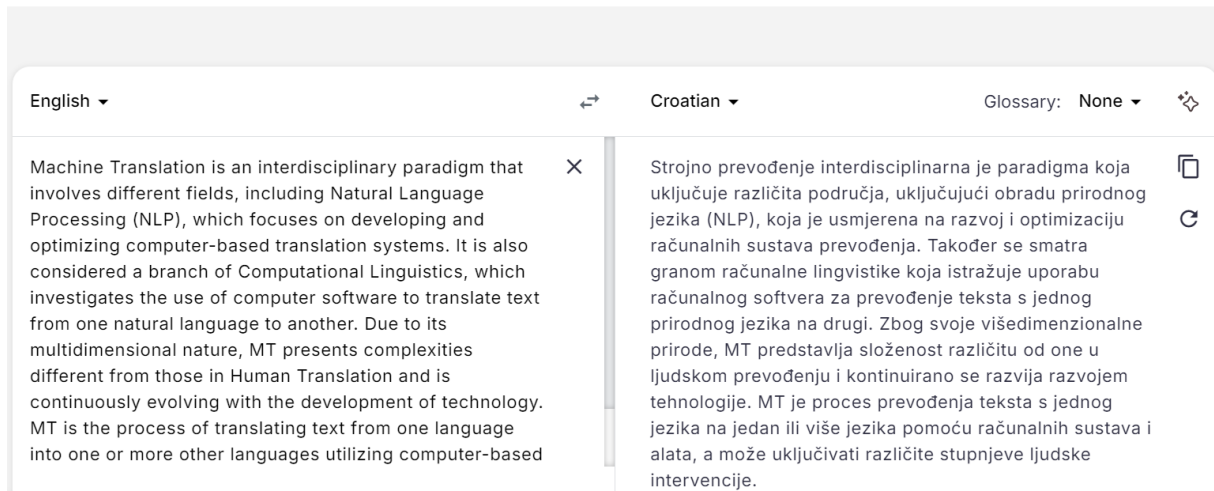
Slika 12 Prikaz prijevoda teksta - Microsoft Bing

Kako možemo vidjeti na slici 12, Microsoft Bing isto kao i Google prevoditelj, Bing ima opciju pretvaranja teksta u govor, ali po pitanju interpunkcije upada u iste zamke. Točke i zarezi se prepisuju iz originalnog teksta i ne mijenjaju se ni onda kada je to potrebno. Microsoft Bing zaostaje za konkurencijom u kvaliteti prevođenja, što se posebno očituje u njegovoj nesposobnosti da ispravno označi fraze koje počinju velikim slovom kao vlastita imena. Ova slabost rezultira time da sustav ne može pravilno obraditi ili prevesti takve izraze, što dovodi do nepotpunih ili netočnih prijevoda. Jednako kao i Google prevoditelj i Tilde MT, Bing se koristi principom uštede prostora, te ponavlja slične riječi u rečenici.

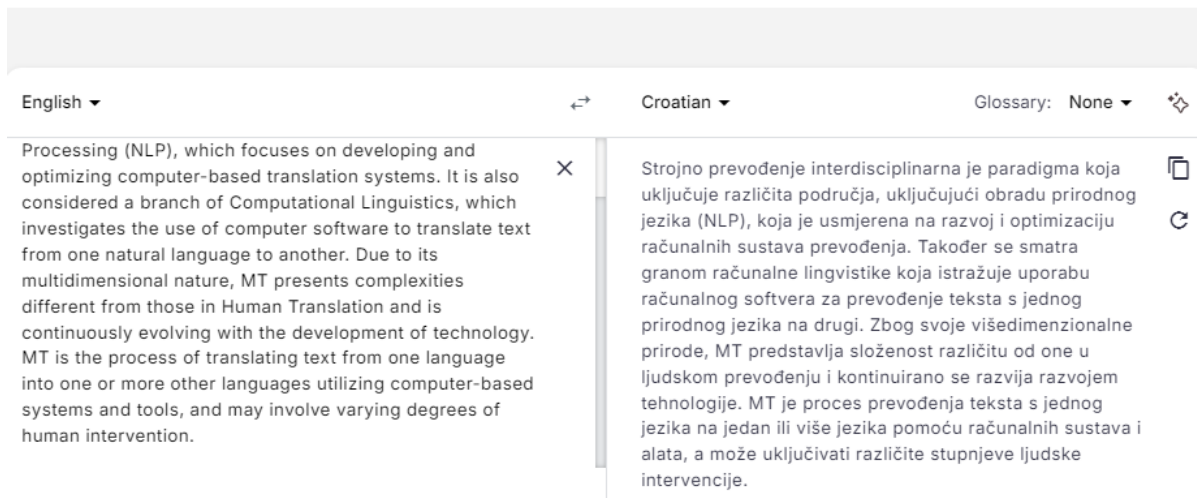


Slika 13 Prikaz prijevoda teksta - Yandex Translate

Yandex, za razliku od svojih konkurenata Google prevoditelja, Microsoft Binga i Tilde MT-a, kako je prikazano na slici 13, ne koristi princip skraćenog prijevoda radi dobivanja kraćeg teksta, niti ima problema s prijevodom interpunkcije. S obzirom na kvalitetniji izbor dužih riječi, smanjeno korištenje anglicizama i izbjegavanje korištenja riječi istog korijena, Yandexov prijevod bliži je ljudskom prevodu. Problem Yandexa je u tome što, za razliku od svojih prethodnika, nije sposoban prenijeti akronime iz teksta, te od njih stvara prijevode koji nisu smisleni. Iako je korištenje dužih riječi i kvalitetnijih izraza poželjno, Yandex uz akronime čini još jednu grešku. Njegove rečenice zahtijevaju puni subjekt. To dovodi do kompliciranih i predugih prijevoda pri kojima se isti izraz višestruko ponavlja u rečenici i ona gubi na kvaliteti i/ili smislu. S obzirom na sve prednosti ovog alata, on se kvalitetom izdiže nad svojom konkurencijom.



Slika 14 Prikaz prijevoda teksta - Tilde MT (1)



Slika 15 Prikaz prijevoda teksta - Tilde MT (2)

Na temelju prijevoda koji se nalazi na slikama 14 i 15, za Tilde MT moglo bi se reći da je poboljšana verzija Google prevoditelja ako uzmemo u obzir to što korektno "prevodi" interpunkcijske znakove. On se s druge strane može smatrati Googlovom lošijom verzijom s obzirom na to da za razliku od svog prethodnika nema opciju pretvaranja teksta u govor, te uz to upada u iste zamke - nezgrapnu inverziju glagola i ponavljanje sličnih riječi unutar rečenice

zbog principa uštede prostora. Jedino čime se još može pohvaliti veći je font u odnosu na google što čini njegov prijevod lakše čitljivim. U usporedbi sa drugim alatima strojnog prevođenja, Tilde MT je pri samom dnu ljestvice.

Analiziramo li riječ „složenosti“, otkrit ćemo da ona ne znači baš ništa. U tom pogledu možemo zaključiti da su sva četiri alata zakazala jer su svoj prijevod utemeljila na pridjevu complex - složeno, te iz toga izvadili imenicu složenost ili složenosti (orig. complexities). Radi se zapravo o složenom problemu ili problematici koja proizlazi iz računalne logike shvaćanja što prijevod jest. U Analizi prirodnog, svakodnevnog jezika ne nalazimo riječ složenosti, dok se pridjev složeno obilno koristi, baš kao i u engleskom jeziku. Ipak, dok je imenica "complexity" na Engleskom rijetka, ali korištena, ona u Hrvatskom ne postoji i strojno je izmišljena zbog manjka ekvivalenta i izbacivanja dužeg no razumljivijeg rješenja, onog složenog problema.

Za počinjene greške nisu nužno krivi alati za strojno prevođenje. Greška je i u samom tekstu. Akronim MT nije stavljen u zagradama pokraj punog izraza na početku prve rečenice, a koristi se u nastavku teksta. Takvo što bi računalu koje analizira daljnje podatke otežalo spajanje kratice sa punim izrazom, ali bi otežalo i razumijevanje čovjeku koji taj prijevod čita bez poznavanja Engleske verzije ili upoznatosti s temom teksta. Još jedan izazov prevođenja hrvatskog jezika, kako za ljude tako i za računalne alate, upravo je ograničena baza podataka hrvatskih riječi i fraza naspram više desetaka riječnika engleskog jezika i fraza koje obuhvaćaju ne samo standardni Britanski i Američki engleski jezik već i dijalektalne varijante Australskog, Irskog, Škotskog, Novo-Zelandskog itd. Dok se i danas stvaraju i izdaju novi riječnici hrvatskog jezika, tek je mali dio digitaliziran i javno dostupan kao potencijalni izvor računalnim alatima, pa stoga alati povremeno izvedu vlastite riječi na bazi postojećih sličnih prijevoda. To najčešće možemo vidjeti ubacujući Engleske izraze u Google gdje se sa desne strane ekrana one prevađaju u najbliži element.

Dok bi se većina složila s tvrdnjom da je prijevod ovih alata razumljiv, također bi potvrdila i da je mjestimice nejasan što možemo vidjeti iz dosad navedenih primjera prevoda teksta. Spojimo li Yandex-ovu bazu riječi, mogućnost čitanja akronima i audiozapis Google prevoditelja, interpunkciju Tilde MT-a i veći izraženi font Microsoft Binga, možda jednog dana dobijemo potpun prevoditeljski alat, barem za tekstove lišene metafore. Također, važno je naglasiti da Yandex ipak zasad ostaje najkorisniji alat za strojno prevođenje, a zatim ga po kvaliteti slijede Google prevoditelj, Microsoft Bing i Tilde MT.

8 ZAKLJUČAK

Strojno prevođenje ima značajan potencijal u unapređenju diseminacije informacija u suvremenom globalnom društvu. Napredak u području dubokog učenja i neuronskih mreža doveo je do znatnog poboljšanja kvalitete strojnog prevođenja, čineći ga sve prikladnijim za upotrebu u različitim kontekstima. Ova tehnologija ne samo da omogućava prevladavanje jezičnih barijera, već i doprinosi bržem i učinkovitijem prijenosu informacija među različitim korisnicima, što je od esencijalne važnosti u današnjem brzom svijetu.

Primjeri upotrebe strojnog prevođenja u znanstvenoj komunikaciji, zdravstvu, obrazovanju i poslovanju pokazuju da ova tehnologija može olakšati, a može i onemogućiti širenje informacija među jezično različitim skupinama korisnika kao što je to prikazano u primjerima primjene strojnih prevoditelja u zdravstvu. U znanstvenoj komunikaciji, strojno prevođenje omogućava istraživačima da dijele svoja otkrića s globalnom publikom, čime se povećava dostupnost znanstvenih rezultata i potiče suradnja među istraživačima iz različitih zemalja.

Međutim, izazovi poput točnosti prijevoda, prilagodbe specifičnom kontekstu i etička pitanja i dalje zahtijevaju pažljivu integraciju strojnog prevođenja u procese diseminacije informacija te su prisutni u gotovo svakom istraživanju. Na primjer, kvalitetni prijevodi u medicinskom kontekstu mogu imati izravan utjecaj na zdravlje pacijenata, stoga je važno osigurati da strojno prevođenje ispunjava visoke standarde točnosti i pouzdanosti.

Daljnji razvoj strojnog prevođenja, uključujući unapređenje neuronskih mreža i prilagodbu specifičnim domenama, imat će ključnu ulogu u budućnosti diseminacije informacija. Interdisciplinarna suradnja stručnjaka iz područja jezikoslovlja, informacijskih znanosti i računarstva bit će ključna za ostvarivanje punog potencijala strojnog prevođenja u službi učinkovitije diseminacije znanja i informacija. Osim toga, potrebno je razviti etičke smjernice koje će regulirati korištenje strojnog prevođenja kako bi se osiguralo da se tehnologija koristi odgovorno i s poštovanjem prema kulturnim razlikama.

U zaključku, strojno prevođenje predstavlja moćan alat koji može značajno unaprijediti način na koji dijelimo i pristupamo informacijama. Njegova sposobnost da prevlada jezične barijere otvara vrata novim mogućnostima za suradnju, inovacije i širenje znanja na globalnoj razini. Kako se tehnologija nastavlja razvijati, važno je da istraživači, praktičari i donosioci odluka zajednički rade na optimizaciji strojnog prevođenja kako bi ono postalo još učinkovitije sredstvo za diseminaciju informacija u svim aspektima ljudskog djelovanja.

LITERATURA

- AI-Salman, Saleh M. 2004. *The effectiveness of machine translation*. International Journal of Arabic-English Studies, 5(1), 145-160.
- Almahasees, Zakaryia, Samah Meqdadi, i Yousef Albudairi. 2021. *Evaluation of Google Translate in Rendering English COVID-19 Texts into Arabic*. Journal of Language and Linguistic Studies 17, no. 4 (2021): 2065–80. <https://search.informit.org/doi/10.3316/informit.228360028176817>.
- Anugu, Anusha i Gajula Ramesh 2020. *A survey on hybrid machine translation*. In E3S Web of Conferences (Vol. 184, p. 01061). EDP Sciences.
- AnyChart. 2022. *What Is Statistical Machine Translation?* Blog.pangeanic.com. Pristupljeno: 18.08.2024. URL: <https://blog.pangeanic.com/what-is-statistical-machine-translation>
- AnyChart. 2023. “Transformer Architecture explained.” Pristupljeno: 05.10.2024. URL: <https://medium.com/@amanatulla1606/transformer-architecture-explained-2c49e2257b4c>
- AnyChart. 2024. “seq2seq Model in Machine Learning“ Pristupljeno: 15.09.2024. URL: <https://www.geeksforgeeks.org/seq2seq-model-in-machine-learning/>
- Blagus Bartolec, Goranka. 2015. *Diseminacija*. Hrvatski jezik 2, br. 2 (2015): 33-0. <https://hrcak.srce.hr/171934>
- Cambridge rječnik. Pristupljeno: 06.10.2024. URL: <https://dictionary.cambridge.org/>
- Chauhan, Nagesh Singh. 2020. *Introduction to Neural Machine Translation(NMT)*. The AI dream. Pristupljeno: 14.09.2024. URL: <https://www.theaidream.com/post/how-ai-is-changing-personal-data-tracking>
- Dahal, Suresh Budha, i Muhammad Aoun. 2023. *Exploring the Role of Machine Translation in Improving Health Information Access for Linguistically Diverse Populations*. Advances in Intelligent Information Systems, 8(2), 1-13.
- Deilen, Silvana, Ekaterina Lapshinova-Koltunski, Sergio Hernández Garrido, Christiane Maaß, Julian Hörner, Vanessa Theel, i Sophie Ziemer. 2024. *Towards AI-supported Health Communication in Plain Language: Evaluating Intralingual Machine Translation of Medical Texts*. In Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health) @ LREC-COLING 2024, pages 44–53, Torino, Italia. ELRA and ICCL.
- Dew, Kristin N., Anne M. Turner, Yong K. Choi, Alyssa Bosold, i Katrin Kirchhoff. 2018. *Development of machine translation technology for assisting health communication: A systematic review*. Journal of biomedical informatics, 85, 56-67.

- Fakih, Altaf, Mozghan Ghassemiazghandi, Abdul-Hafeed Fakih i Manjet K. M. Singh. 2024. *Evaluation of Instagram's Neural Machine Translation for Literary Texts: An MQM-Based Analysis*. Gema Online Journal of Language Studies, 24(1).
- Harison, Tovmas. 2023. *Statistical machine translation*. Machinetranslate.org. Pristupljeno: 18.08.2024. URL: <https://machinetranslate.org/statistical-machine-translation>
- Kirchhoff, Katrin, Anne M. Turner, Amittai Axelrod, Francisco Saavedra. 2011. *Application of statistical machine translation to public health information: a feasibility study*. Journal of the American Medical Informatics Association, Volume 18, Issue 4, Pages 473–478, <https://doi.org/10.1136/amiajnl-2011->
- LeCun, Yann, Yoshua Bengio, i Geoffrey Hinton. 2015. *Deep learning*. Nature 521(7553):436–444. <https://doi.org/10.1038/nature14539>
- Liu, Yanchao, Pengzhou Zhang, Lei Shi, i Junpeng Gong. 2023. *A Survey of Information Dissemination Model, Datasets, and Insight*. Mathematics, 11(17), 3707.
- Macken, Lieve, Vanessa De Wilde, and Arda Tezcan. 2024. *Machine Translation for Open Scholarly Communication: Examining the Relationship between Translation Quality and Reading Effort*. Information (2078-2489) 15(8).
- Mahadin, Dana Khalid, i Sameer Naser Olimat. 2022. *Jordanian translators' use of machine translation and glossary of COVID-19 terminology with reference to Arabic*. New Voices in Translation Studies, 26(1), 25-54.
- Maučec, Mirjam Sepesy, i Gregor Donaj. 2019. *Machine translation and the evaluation of its quality*. Recent trends in computational intelligence, 143.
- Merriam-Webster rječnik. Pristupljeno: 06.10.2024. URL: <https://www.merriam-webster.com/>
- Nalluri, Srinivasa Rao, i Brahmaiah Gaddam. 2016. *Mobile library services and technologies: A study*. International Journal of Research in Library Science, 2(2), 59-66.
- Niño, Ana. 2020. *Exploring the use of online machine translation for independent language learning*. Research in Learning and Technology, Vol. 28, 2020.
- Neu, Dominic A., Johannes Lahann, i Peter Fettke. 2022. *A systematic literature review on state-of-the-art deep learning methods for process prediction*. Artif Intell Rev 55, 801–827. <https://doi.org/10.1007/s10462-021-09960-8>
- Oh, Seokjin, Su Ah Lee, i Woohwan Jung. 2023. *Data augmentation for neural machine translation using generative language model*. arXiv preprint arXiv:2307.16833.
- Pokorny, Elizabeth. 2024. "Machine Translation for Marketing Content - A Comparative Study Into Machine Translation Quality." Weglot.com. Objavljeno: 01.02.2024. Pristupljeno:

06.10.2024. URL: <https://www.weglot.com/blog/comparative-study-machine-translation-quality>

PubMed. 2017. *PubMed Mobile*. Retrieved January 21, 2017, from NCBI: <https://www.ncbi.nlm.nih.gov/m/pubmed/>

Rico, Celia, i Diana Gonzalez Pastor. 2022. *The Role of Machine Translation in Translation Education: A Thematic Analysis of Translator Educators' Beliefs*. *Translation & Interpreting: The International Journal of Translation and Interpreting Research* 14, no. 1: 177–97. <https://search.informit.org/doi/10.3316/informit.360710023023812>.

Rosmani, Arifah Fasha, Ariffin Abdul Mutalib, i Siti Mahfuzah Sarif. 2020. *The evolution of information dissemination, communication media and technology in Malaysia*. *Journal of Physics: Conference Series*. Vol. 1529. No. 2. IOP Publishing, 2020.

Sai, Ponnuru Mahesh Datta, et al. N.d. *Advancements in Machine Translation Driven by Sequence-to-Sequence Techniques*. Pristupljeno: 15.09.2024.

Sakre, Mohammed M. 2019. *Machine translation status and its effect on business*. *Journal of the ACS*, 10.

Salam, Khan Md Anwarus, Setsuo Yamada i Nishino Tetsuro. 2017. *Improve example-based machine translation quality for low-resource language using ontology*. *International Journal of Networked and Distributed Computing* 5.3: 176-191.

Saxena, Archana, i R.D. Yadav. 2013. *Impact of mobile technology on libraries: A descriptive study*. *International Journal of Digital Library Services*, 3(4), 1-58.

Shonhe, Liah. 2017. *A literature review of information dissemination techniques in the 21st century era*. *Library Philosophy and Practice (e-journal)*, 1731.

Singh, Shashi Pal, Ajai Kumar, Hemant Darbari, Lenali Singh, Anshika Rastogi i Shikha Jain. 2017. "Machine translation using deep learning: An overview." U *2017 international conference on computer, communications and electronics (comptelix)*. IEEE, 2017. 162-167.

Slocum, Jonathan. 1985. *A survey of machine translation: Its history, current status and future prospects*. *Computational linguistics*, 11(1), 1-17.

Steigerwald, Emma, Valeria Ramírez-Castañeda, Débora Y C Brandt, Andrés Báldi, Julie Teresa Shapiro, Lynne Bowker, i Rebecca D Tarvin. 2022. *Overcoming language barriers in academia: Machine translation tools and a vision for a multilingual future*. *BioScience* 72.10: 988-998.

Šuman, Sabrina. 2021. *Pregled metoda obrade prirodnih jezika i strojnog prevođenja*. *Zbornik Veleučilišta u Rijeci*, 9 (1), 371-384. <https://doi.org/10.31784/zvr.9.1.23>

Tan, Zhixing, Shuo Wang, Zonghan Yang, Gang Chen, Xuancheng Huang, Maosong Sun i Yang Liu. 2020. *Neural machine translation: A review of methods, resources, and tools*. AI Open, 1: 5-21.

Terumasa, Ehara. 2007. *Rule based machine translation combined with statistical post editor for japanese to english patent translation*. In Proceedings of the Workshop on Patent translation. URL: <https://aclanthology.org/2007.mtsummit-wpt.4.pdf>

The Free Dictionary. Pristupljeno: 06.10.2024. URL: <https://www.thefreedictionary.com/>

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser i Illia Polosukhin. 2017. *Attention Is All You Need*. arXiv preprint arXiv:1706.03762 <https://arxiv.org/abs/1706.03762>

Vohland, Katrin, Anne Land-Zandstra, Luigi Ceccaroni, Rob Lemmens, Josep Perelló, Marisa Ponti, Roeland Samson, Katherin Wagenknecht. 2021. “Communication and Dissemination in Citizen Science.” U *The Science of Citizen Science*, 475–494, Springer International Publishing. https://doi.org/10.1007/978-3-030-58278-4_21

Wang, Haifeng, Hua Wu, Zhongjun He, Liang Huang, i Kenneth Ward Church. 2022. *Progress in machine translation*. Engineering, 18, 143-153.

Wang, Linden. 2024. *Applying automated machine translation to educational video courses*. Education and Information Technologies 29.9 (2024): 10377-10390. <https://doi.org/10.1007/s10639-023-12219-0>

Yang, Shuoheng, Yuxin Wang, and Xiaowen Chu. 2020. *A survey of deep learning techniques for neural machine translation*. arXiv:2002.07526. <https://arxiv.org/abs/2002.07526>

POPIS SLIKA

Slika 1 Prikaz razvoja medija za širenje informacija (Rosmani et al. 2020)	6
Slika 2 Kronološku prikaz razvoja strojnog prevođenja (Maučec i Donaj 2019, 145).....	9
Slika 3 Proces prevođenja japanskog teksta temeljen na RBMT sustavu (Terumasa 2007, 2)	10
Slika 4 Arhitektura EBMT sustava (Anwarus et al. 2017, 178)	12
Slika 5 Prikaz Encoder-Decoder arhitekture (Chauhan 2020).....	14
Slika 6 Prikaz modela seq2seq (AnyChart 2024)	15
Slika 7 Arhitektura Transformatora (AnyChart 2023).....	19

Slika 8 Prikaz prosječnog vrijeme čitanja (Macken, De Wilde i Tezcan 2024)	24
Slika 9 Prikaz postupka prevođenja edukacijskih videozapisa (Wang 2024).....	29
Slika 10 Prikaz jezika na kojima je provedeno istraživanje (Pokorny 2024)	31
Slika 11 Prikaz prijevoda teksta - Google Prevoditelj	33
Slika 12 Prikaz prijevoda teksta - Microsoft Bing.....	34
Slika 13 Prikaz prijevoda teksta - Yandex Translate	35
Slika 14 Prikaz prijevoda teksta - Tilde MT (1)	36
Slika 15 Prikaz prijevoda teksta - Tilde MT (2)	36

POPIS TABLICA

Tablica 1 Značenja riječi "diseminacija"	4
------------------------------------------------	---

MACHINE TRANSLATION OF TEXTS AS A TOOL FOR INFORMATION DISSEMINATION

SUMMARY

The subject of this paper is the analysis of information dissemination through machine translation, with a particular emphasis on the technologies that support this process. Machine translation serves as a crucial tool in modern communication, facilitating rapid and efficient transfer of information across different languages and cultures. The motivation for studying this topic arises from the increasing need for effective communication and information exchange in a globalized world, where the speed and accuracy of information are paramount, especially in the contexts of scientific communication, healthcare, and education. The research methodology involves analyzing a specific text input into existing machine translation tools, followed by a comparison of the results obtained from selected tools. The key findings indicate that modern machine translation tools, particularly those based on deep learning, have significantly improved the quality of automatic translations, thereby opening the door to broader applications across various sectors, including science, healthcare, and education. This technology not only enhances access to information but also fosters global collaboration by reducing language barriers and improving mutual understanding. The main conclusion of this study highlights that the integration of machine translation into information dissemination processes is essential for enhancing access to knowledge and resources, thereby creating new opportunities for collaboration and innovation across different sectors. These implications underscore the necessity for further research and development of technologies that will enhance the efficiency of communication processes.

Keywords: dissemination, information, knowledge, machine translation.