

STROJNO PREVAJANJE SYSTRAN

KAZALO

| | |
|--|----|
| UVOD | 1 |
| 1 RAZPOZNAVANJE GOVORA | 2 |
| 1.1 Potencialne aplikacije za razpoznavanje govora | 4 |
| 2 SINTEZA GOVORA | 4 |
| 2.1 Zgradba sistema za sintezo govora | 5 |
| 2.1.1 Pretvorba teksta v enote sinteze | 6 |
| 2.1.2 Pretvorba teksta v metrične parametre | 6 |
| 2.1.3 Pretvorba sinteznih enot v govor | 7 |
| 3 NUANCE | 8 |
| 3.1 Program za sintezo govora | 9 |
| 3.2 Program za razpoznavanje govora | 10 |
| 4 STROJNO PREVAJANJE | 12 |
| 4.1 Systran | 12 |
| 4.2 Zgodovina | 12 |
| 4.2.1 Izdelava strojnega prevajanja | 13 |
| 4.2.2 Systran – strojno prevajanje | 14 |
| 4.2.3 Systran – takrat in danes | 14 |
| 4.3 Strojno prevajanje | 16 |
| 4.4 Človeško prevajanje | 16 |
| 4.5 Razlika med strojnim in človeškim prevajanjem | 17 |
| 4.5.1 Kapaciteta in hitrost | 17 |
| 4.5.2 Povrnitev investicije | 17 |
| 4.5.3 Produktivnost | 17 |
| 4.5.4 Natančnost | 17 |
| 4.5.5 Prilagoditev | 18 |
| 4.6 Raziskave in študije | 18 |
| 4.6.1 Metodologija | 18 |
| 4.6.2 Modularnost | 19 |
| 4.7 Slovar | 19 |
| 4.7.1 Osnovni slovar (Stem dictionary) | 19 |
| 4.7.2 Izrazoslovni slovar (Expression dictionary) | 19 |
| 4.7.3 Slovar za specifičnega uporabnika (Customer Specific Dictionary - CSD) | 19 |
| 4.8 Sistemski program | 20 |
| 4.9 Jezikoslovni program | 20 |
| 4.9.1 Analizator | 20 |
| 4.9.2 Ciljni prevajalski jezikovni moduli | 20 |
| 4.9.3 Strnitveni modul (Synthesis modul) | 21 |
| 4.10 Izdelava dodatnih jezikov | 21 |
| 4.11 Systran – tehnologija | 21 |
| 4.12 Standardno strojno prevajanje proti prirejenemu | 22 |
| 4.13 Stvari, ki jih je dobro upoštevati za prevajanje | 23 |
| 4.14 Prevajalski program za izobraževalne ustanove | 25 |
| 4.15 Prevajalski program za vladne ustanove | 26 |
| 4.16 Systran – produkti | 26 |
| 4.17 SystranLinks | 27 |
| 4.18 SystraNet | 28 |
| 4.19 SystranBox | 28 |
| 4.20 Odjemalec – strežnik aplikacije | 29 |
| 4.21 Systran – orodna vrstica | 32 |
| 4.22 Systran – Translation Project Manager (STPM) | 32 |
| 4.23 Systran – Dictionary Manager | 33 |
| 7 ZAKLJUČEK | 34 |
| 8 VIRI IN LITERATURA | 35 |

UVOD

Kaj je jezikovna tehnologija

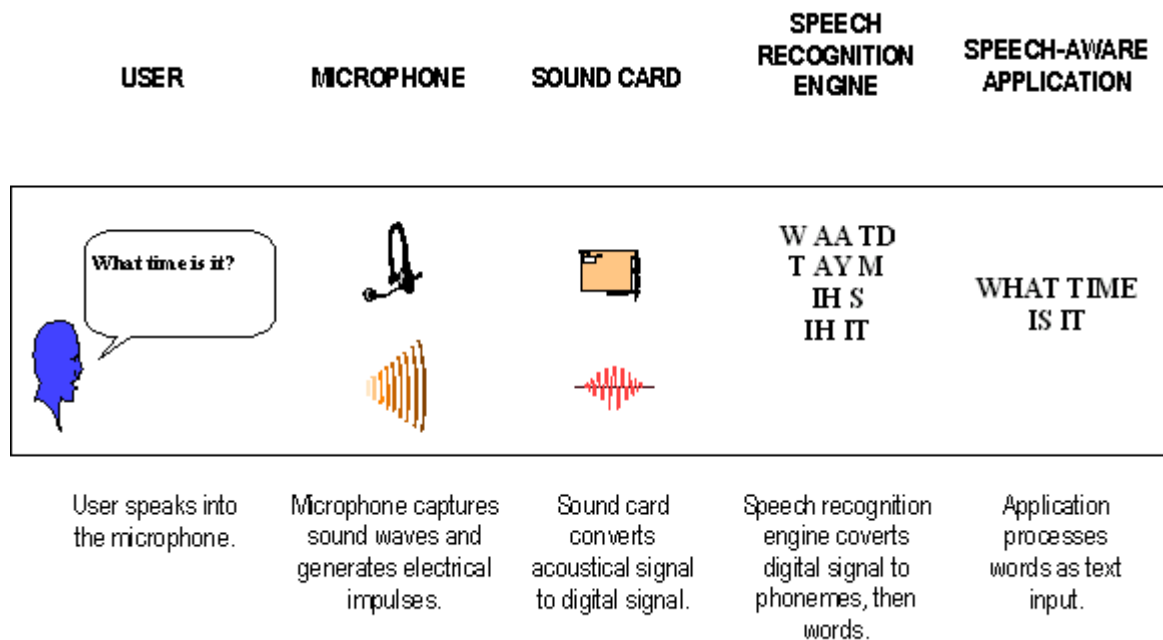
V sredini 90-ih let so osebni računalniki postajali dovolj močni, da govorijo z ljudmi in ljudi njim. Kakorkoli že jezikovna tehnologija danes omogoča veliko realnih prednosti v aplikacijah.

Na primer:

Zelo veliko večjih podjetji dodaja razpoznavanje govora v njihove sisteme. Samo s klicem in govorom lahko uporabnik kupuje in prodaja delnice, preveri letalske informacije ali naroči dobrine. Sistem se odzove na način z uporabo kombinacij naprej posnetih sposobnosti in umetno generiranih glasov. Napredek se kaže tudi v sistemih za sintezo govora, s pomočjo katerih stvari, ki so podane v tekstovni obliki spremenimo v govor. Ti nudijo predvsem pomoč slepim ali slabovidnim lahko pa tudi služijo za učenje jezika, itd... Zelo pomembno mesto pa imajo sistemi za strojno prevajanje. Njihov namen je prevajanje oz. pretvarjanje iz enega jezika v drugega. Pomembnost se kaže v tem, ker se vse več držav z različnimi jeziki združuje v skupnosti (npr. EU) in tako je pomembna hitrost oz. način komunikacije med njimi.

1 RAZPOZNAVANJE GOVORA

Razpoznavanje govora ali govor-v-tekst vključuje zajemanje in digitaliziranje zvočnih valov, njihovo pretvorbo v osnovne jezikovne enote in foneme, gradnjo besed iz fonemov in analizo besed, da se zagotovi slovnična pravilnost za besede, ki zvenijo podobno. Spodnja slika prikazuje ta proces.



Slika 1: Delovni proces razpoznavanja govora

Stroji za razpoznavanje govora so programski gonilniki, ki pretvarjajo akustične signale v digitalni signal in omogočijo prikaz prepoznanega jezika v tekstovni obliki. Večino teh sistemov omogočajo kontinuiran govor, kar pomeni, da lahko govorimo v mikrofonsko hitrostjo običajnega pogovora. Izolirani ali diskretni sistemi za razpoznavo govora zahtevajo od uporabnika, da med vsako besedo naredi premor in so sedaj nadomeščeni z sistemi za kontinuiran govor. Ti sistemi podpirajo dve obliki razpoznavanja govora:

1. **Diktiranje, kjer uporabnik vnaša podatke z branjem računalniku.**
2. **Ukazovanje in kontrola, kjer uporabnik vpeljuje akcije z govorjenjem komand in zastavljanjem vprašanj.**

Diktiranje je način, ki omogoča uporabniku da diktira zapiske, pisma ter elektronska sporočila, kot tudi vnos podatkov z uporabo orodja za diktiranje razpoznanega jezika. Možnosti, kaj je lahko razpoznano je omejeno z velikostjo slovnice ali slovarja besed. Večino sistemov za razpoznavo v tem načinu so odvisni od govorca, kar pomeni, da je sama natančnost odvisna od osnove uporabnikovega vzorca in naglasa. Da se zagotovi natančna razpoznavo, mora aplikacija izdelati ali dostopati do govorcevega profila, ki vključuje natančen zemljevid uporabnikovega vzorca, ki se uporablja v procesu med razpoznavo.

Način ukazovanje in kontrole pa nudi razvijalcem najlažjo vgradnjo govornega vmesnika v obstoječo aplikacijo. V tem načinu je lahko slovnica (sklop prepoznanih besed) omejena na sklop obstoječih komand, veliko bolj je omejen tudi namen kot pa pri kontinuiranem diktiranem načinu, kateri mora dostopati skoraj do celega slovarja. To zagotovi boljše natančnost in zmogljivost ter zmanjša procesorsko moč, ki jo potrebuje aplikacija. Omejena slovnica prav tako omogoča od govorca neodvisen proces, izključi potrebo po govorcevem profilu ali učenje sistema za razpoznavo govora.

Tehnologija za razpoznavo govora omogoča razvijalcem da vključijo naslednje značilnosti v njihove aplikacije:

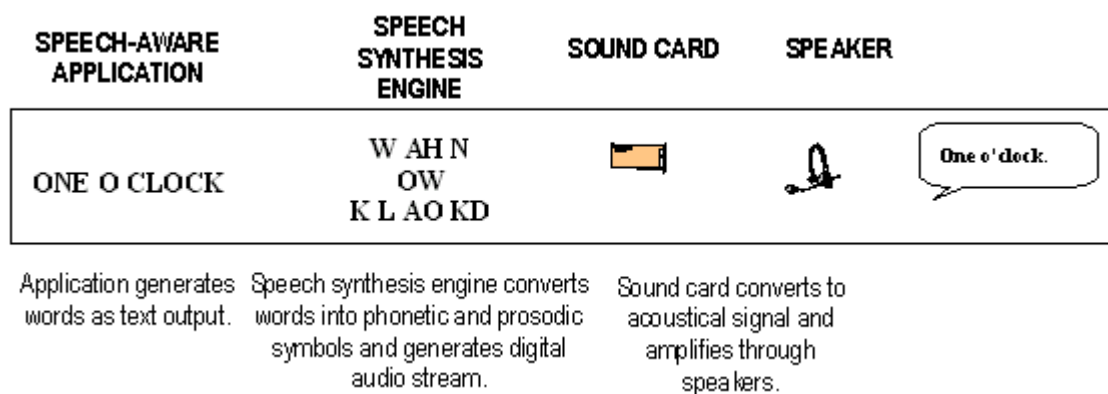
- Uporaba računalnika brez rok (alternativa tipkovnici) ali uporaba aplikacij, kjer je tipkovnica nepraktična (mobilne naprave, telefoni,...)
- Več »človeških« računalnikov, s katerimi lahko uporabnik komunicira in s tem naredi izobraževalne ter zabavne aplikacije bolj prijazne in realistične.
- Jezikovne odgovore na sporočila in čarovniške zaslone se lažje implementira v aplikacije.
- Neovirani dostopi do aplikacijskih kontrol in velikih knjižnic omogočajo uporabniku katerokoli obliko iz knjižnic ali katerokoli komando iz potencialne velike oblike komand, brez da bi bilo potrebno iti skozi nekaj oken in menijev.
- Makro-ji omogočajo uporabniku, da govori naravne besede ali fraze, kot pa da bi uporabljal tipkovnico ali komande za aktiviranje makro-ja.
- Situacijski dialogi so možni med uporabnikom in računalnikom, kjer računalnik sprašuje uporabnika in obdeluje podatke glede na odgovor.

1.1 Potencialne aplikacije za razpoznavanje govora

- Igre in razvedrilo
- Vnos podatkov
- Urejanje dokumentov
- Govorni avtomati
- Pozivni centri
- Dajanje glasovnih komand
- Pomoč slepim in slabovidnim

2 SINTEZA GOVORA

Sinteza govora ali tekst-v-govor je proces kjer pretvorimo tekst v govor. To vključuje razbijanje besed v foneme; analiza za posebno rokovanje teksta kot so: številke, valut, sklanjatev in postavljanje ločil ter generiranje digitalne avdio oblike za predvajanje. Obstaja nekaj funkcij, ki jih sintetizator izvaja in so prikazane na spodnji sliki.



Slika 2: Delovni proces tekst-v-govor

Programski gonilniki imenovani sintetizatorji ali tekst-v-govor glasovi izvajajo jezikovne sinteze, rokujejo z kompleksnostjo pretvorbe teksta in generirajo govorni jezik. Ti sistemi generirajo zvoke podobne tistim, ki jih govorijo ljudje in vključujejo različne filtre, ki simulirajo dolžino vratu, velikost ustne votline, obliko ustnic in pozicijo jezika. Kljub temu da

je lahko razumljiv, zvok ki je narejen s to tehnologijo zveni manj človeško kot pa zvok narejen z digitalnim snemalnikom.

Kljub temu pa so tekst-v-govor aplikacije boljša alternativa v situacijah kjer je digitalno snemanje nepraktično. Ponavadi se uporablja tekst-v-govor aplikacije v sledečih primerih:

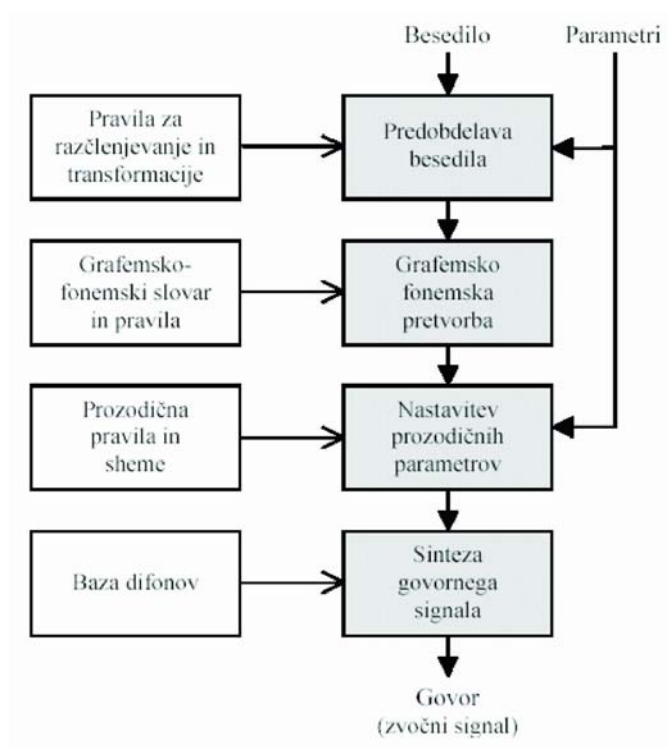
- Avdio snemanje je preveliko za hranjenje na disku ali predrago.
- Aplikacijski odgovori potrebujejo kratke fraze.
- Aplikacije ne morejo predvideti, če bo potrebna komunikacija ali odgovor.
- Uporabniki imajo rajši avdio odgovor ali realističen. (1)

2.1 Zgradba sistema za sintezo govora

Postopek sinteze govora lahko razdelimo na tri glavna področja:

- Pretvorba teksta v enote sinteze
- Pretvorba teksta v metrične parametre
- Pretvorba sinteznih enot v govor

Shematičen prikaz pretvorbe teksta v govor:



Slika 3: Proces sinteze govora

2.1.1 Pretvorba teksta v enote sinteze

Ker večina metod sinteze deluje na osnovi fonemov so enote sinteze fonemi, lahko pa bi uporabljali kakšne druge enote, vendar se taka izbira pokaže kot najboljša.

Kljub temu, da so ponavadi vhod ASCII znaki, je pretvorba večstopenjska operacija katere optimum še ni odkrit. Prva stopnja pri pretvorbi je normalizacija. Z normalizacijo interpretiramo odstavke, ločila, Zaporedja števk se razvijejo v ustrezne številke - glavne ali vrstilne. Okrajšave in akronimi se zapišejo v polni obliki. Zaporedje besed se nato pretvori v zaporedje enot sinteze (fonemov). Pri pretvarjanju besed lahko najprej iščemo v slovarju pogostih izgovorjav, ki ga je predhodno potrebno zgraditi za vsak jezik posebej. Če besede ne najdemo v slovarju ali ga nimamo, ji moramo izgovorjavo določiti sami. Drugi del pretvorbe predstavlja napovedovanje mesta naglasa. Mesto naglasa je zlog, na katerem ima beseda jakostno in tonsko izrazitost. Za slovenski jezik je značilno prosto mesto naglasa, saj se ta lahko pojavi na prvem, zadnjem, predzadnjem ali predpredzadnjem zlogu. Prav tako ima lahko posamezna beseda več mest naglasa. Zadnja faza je pretvorba besed v fonetični zapis. Za to pretvorbo lahko uporabljamo produkcijska pravila za posamezne glasove, ki jih moramo predhodno sami določiti in so odvisna od jezika.

2.1.2 Pretvorba teksta v metrične parametre

Določaje metričnih parametrov spet lahko razdelimo na tri dele. Prvi del je določanje intonacijskih mej. Za idealno določanje mej je potrebno poznavanje celotne sintaktične analize stavka, kar pa je težko izvesti avtomatično. Pogosto pomanjkanje intonacijskih mej povzroči, da je glas bolj grob. Preveliko število intonacijskih mej pa povzročajo zmedo in motnje. Najenostavnejša metoda je postavljanje mej, kjer le te določajo ločila. Malo bolj izpopolnjena rešitev je rešitev v kateri imamo shranjeno listo funkcijskih besed pri katerih nastopajo meje. Poleg teh dveh enostavnih metod obstaja še veliko bolj izpopolnjenih metod. Drugi del je določanje dolžin posameznih segmentov. Dolžina segmentov se ponavadi določa z množico pravil. Pravila so sestavljena iz velikega števila faktorjev, ki bi lahko vplivali na dolžino segmenta. Nekateri izmed teh faktorjev so: frekvenca besede, sintaktična kategorija, fonetičen kontekst in še mnogi drugi.

Tretji del je določanje frekvenčnih skic. Osnovna frekvenca človeškega govora je določena kot kombinacija velikega števila faktorjev. Sistemi za sintezo govora ne upoštevajo vsebine

teksta, ki ga sintetizirajo, ker bi morali poznati vsebino teksta. Prav tako ne upoštevajo nekaterih drugih lastnosti človeškega glasu, zato je govor bolj zglajen kot človeški.

2.1.3 Pretvorba sinteznih enot v govor

Ko imamo določene vse parametre - stavčno intonacijo, naglašene zloge besed, znakovni zapis pretvorjen v zaporedje glasov izgovorjave moramo glasove še pravilno realizirati. To v grobem lahko naredimo na dva načina:

1. Z generiranjem glasov s pomočjo simulacije človeškega govornega trakta

- artikularni: fizični model je zasnovan na podlagi opisa fiziologije in akustike govornega trakta
- formantni sintetizatorji: prenosna funkcija je opisana s formantnimi frekvencami in formantnimi amplitudami - umetno rekonstrukcijo formantnih značilnosti dosežemo z vzbujanjem niza resonatorjev in antiresonatorjev (periodično vzbujenje za zveneče glasove in vzbujaje s šumom za nezveneče glasove)

Za dobro generiranje potrebujemo dober matematični model govornega trakta, poraba pomnilnika pa je majhna. Slabosti pa sta kompliciran matematični model pri artikularnih sintetizatorjih in težavno nastavljanje formantnih parametrov pri formantnih sintetizatorjih.

2. Z združevanjem predhodno posnetih enot

Predhodno posnete enote so vzete iz naravnega govora - na primer: alofoni, difoni, trifoni večji deli pogostejše uporabljenih besed, celih besed ali kombinacijo le-teh. Potrebujemo torej bazo osnovnih enot, s katerimi bomo tvorili govor. Pri manjših osnovnih enotah je poraba pomnilnika manjša, vendar pa so pri večjih enotah rezultati boljši. Zaradi spektralnih nezveznosti, ki nastanejo pri lepljenju osnovnih enot, uporabljamo postopke, ki te nezveznosti zgladijo. Najbolj razširjeni postopki, ki se uporabljajo pri obdelavi signala, so:

- večpulsno linearno napredovnanje (LPC)
- obdelovanje signalov sinhrono z osnovno periodo (PSOLA in WSOLA)
- kombiniranje harmonskih in šumnih modelov (HNS, MBR-PSOLA)

Najpogosteje se uporablja algoritem TD-PSOLA (time domain PSOLA - združevanje osnovnih enot poteka v časovnem prostoru). Obstaja pa tudi algoritem FD-PSOLA (frequency domain PSOLA), pri katerem poteka združevanje v frekvenčnem prostoru in daje nekoliko boljše rezultate kot TD-PSOLA, a je procesorsko zahtevnejši. (2)

3 NUANCE

Nuance je vodilni proizvajalec jezikovnih in slikovnih rešitev za podjetja in domače uporabnike širom sveta. Njihove tehnologije, aplikacije in usluge omogočijo uporabniku, da izkusi kako ljudje komunicirajo oz. delajo z informacijami in kako izdelujejo, izmenjujejo in uporabljajo dokumente. Vsak da milijon uporabnikov in tisoč podjetij izkusijo Nuance z klicem avtomatskega telefonskega imenika, pridobivanjem uporabniških informacij, potrpežljivim diktiranjem posnetkov, ukazovanjem navigacijskim sistemom njihove destinacije ali digitalnim reproduciranjem dokumentov, ki se jih da izmenjevati in iskati. Področja kjer se uporablja Nuance-ova tehnologija:

- Informacijski/kontaktne centri
- Telekomunikacije
- Avtomatski dostop do telefonskega imenika
- Avtomatičnost (navigacijski sistemi, mobilnost na višjem nivoju)
- Mobilni aparati
- Igre
- Dostopnost (zmanjšanje ovir za invalide)
- Zdravstvo

Nuance-ov tehnologije pomagajo na tisoče podjetjem, da dosežejo njihove uporabniške usluge in finančne cilje vsak dan. Osnova uspešnih jezikovno podprtih aplikacij, paket jezikovnih in dialognih tehnologij omogočijo uporabnikom, da izdelajo in slišijo govorjene informacije z

uporabo najboljših naravnih jezikovnih dialogov naslednje generacije. Prav tako so njihove tehnologije podprte z najbolj obširnimi industrijskimi odprtimi platformami in podpirajo večino industrijskih standardov. RealSpeak že podpira 24 jezikov in narečij z več kot 30 glasovnimi različicami. Ustanovitvene tehnologije:

- Razpoznavanje govora (Nuance 8 in OpenSpeech Recognizer 3.0)
- Vgrajeno razpoznavanje govora (VoCon za avtomatičnost, mobilni aparati, igre,...)
- Tekst-v-govor/sinteza govora (RealSpeak več jezikovni program za sintezo govora)
- Potrditev govorca z namenov zavarovati prenose
- Pogovorne tehnologije

3.1 Program za sintezo govora

RealSpeak je program, ki spremeni tekst v zelo kvaliteten moški ali ženski govor. Njegova dobra lastnost je, da omogoča uporabo v avtomobilskih navigacijskih sistemih, branje zaslonov za slepe in slabovidne ali pa poveča klicne usluge. RealSpeak je zato dobra odločitev ker je prijazen, izrazit, visoko kakovosten jezikovni vmesnik.

1. **REALSPEAK TELECOM** je vodilni program, ki spremeni tekst v čist človeški govor v več kot 20 jezikov in 30 glasov v različnih naglasih in stilih govora.
2. **REALSPEAK SOLO 4.0** je rešitev, ki je narejen posebej, da poveča kakovost vgrajenih pogovornih aplikacij. Z njim se zagotovi izjemen visoko kakovosten jezik.
3. **REALSPEAK WORD** uporablja izjemen pristop k tekst-v-govor tehnologiji, da doseže največjo kvaliteto govora iz slovarja besed in jezikovnih posebnosti ter omogoča učencem jezika, da slišijo kako morajo biti besede izgovorjene. Zasnovan je na arhitekturi, ki ne potrebuje nobenega snemanja in zasede 3MB prostora za 60,000 besed.
4. **REALSPEAK MOBILE** je družina tekst-v-govor izdelkov, katerih cilj so mobilni aparati, ki vsebujejo SMS in elektronsko sporočanje.

3.2 Program za razpoznavanje govora

Dragon NaturallySpeaking je de facto standard za glasovno razpoznavanje. Trenutno je to najbolj natančen proizvod za razpoznavo govora – do 99% natančnost. Je idealna rešitev za PC navdušence in domače office uporabnike. Samo z neprekinjenim govorjenjem besed in tekst se brez napak prikaže na Microsoft Word in Excel, Corel WordPerfect oz. vseh Windows zasnovanih aplikacijah. Poslušanje prejetih elektronskih sporočil in branje dokumentov naglas. Iskanje po spletu z govorjenjem URL-ja in linkov. Z diktiranjem okrajšav, ki omogočijo vnos bloka teksta kot so ime, naziv ali pa podpis z eno samo glasovno komando. Dragon NaturallySpeaking je narejen v več različicah z edinstvenimi izdelki za medicino in industrijo.

Prednosti:

- **Delo z vsemi priljubljenimi programi;** Microsoft Word, Microsoft Outlook Express, Microsoft Excel, Microsoft Internet Explorer, America Online in Corel WordPerfect
- **Hitrejša kot tipkanje;** večina ljudi govori hitreje kot 160 besed na minuto, vendar tipkajo manj kot 40 na minuto. To pomeni, da lahko ustvarimo pisma, elektronska sporočila približno trikrat hitreje z uporabo tega programa.
- **Natančnost;** natančnost se je povečala za 25% in znaša 99% in je posledično bolj natančna kot pri tipkanju. Ta program ne naredi nikoli slovničnih napak in postaja pametnejši bolj kot ga uporabljamo.
- **Preprosta uporaba**
- **Mobilnost**

Različice:

1. **DRAGON NATURALLYSPEAKING DESKTOP** 99% natančnost, lahka uporaba, fleksibilnost, dostopnost in zagotovi vse posameznikove specifične zahteve.
2. **DRAGON NATURALLYSPEAKING SDK**, če je potreba po implementaciji jezika v windows aplikacijah potem je ta program rešitev za to. Je bolj natančen in robusten. Uporabniške in strežniške verzije omogočajo hitro dodajanje aplikacij za prepoznavanje jezika direktno na strežnik.

3. **DRAGON MT WORKFLOW** avtomatizira transkripcijski proces, da zagotovi optimalno učinkovitost iz nareka skozi distribucijo dokumentov. Iz strateškega vidika bo ta tehnologija zmanjšala stroške, reducirala čas dela z dokumenti in povečala natančnost posnetkov. Uporablja se predvsem v bolnicah in medicinskih skupinah.
4. **DRAGON AUDIOMINING** je revolucionarna rešitev, ki avtomatsko omogoča uporabo tipkovnice in fraz za iskanje avdio in video datotek. To deluje s pomočjo NaturallySpeaking stroja za razpoznavo govora, da izdela XML jezikovni indeks za vsako besedo v avdio in video datoteki ter omogoči uporabniku, da gre na točno lokacijo kjer je bila ta fraza najdena. Pomembne informacije v novih priponkah, vladnih predstavitev, izobraževalnih posnetkih, napovedih in analizah s pomočjo tega orodja je tako lahko iskati, kot po spletnih straneh. (3)

4 STROJNO PREVAJANJE

4.1 Systran

Systran je vodilni ponudnik programske opreme na svetu za prevajane oz. prevajalsko arhitekturo. Njegovo jedrna tehnologija omogoča revolucionarne prevajalske rešitve za Internet, PC in mrežne infrastrukture, ki olajšajo komunikacijo med 36 jezikovnimi pari in 20 specializiranimi domenami.

Systranova tehnologija strojnega prevajanja vsebuje več jezikovne funkcije v procesiranju informacij ter zamenjave za aplikacije kot so elektronska trgovina, CRM, vsebinski management, podatkovne baze, »corporate Intranets«, elektronsko pošto, stalno pošiljanje sporočil, SMS, WAP in mnoge druge. Strojno prevajanje omogoča izdelavo dinamične poslovne vsebine z zelo nizkimi stroški.

Njihovo strokovno znanje se razteza že več kot 30 let gradenj prilagojenih prevajalskih rešitev za velike korporacije, portale, ISP, vlade ter javne zavode skozi odprto in robustno arhitekturo. (4)

4.2 Zgodovina

Systran se je prvič pojavil leta 1968 kot program za prevajanje. Nudil je celoten spekter produktov in uslug za avtomatsko prevajanje širši množici uporabnikov.

Systran razvija in dobavlja svetovno najbolj obsežen prevajalski program narejen posebno za vlade, večje korporacije, večja in manjša podjetja, posamezne uporabnike, večje medmrežne aplikacije in spletne portale. Systran-ova tehnologija lahko prevede več kot 36 jezikovnih parov in se jo da integrirati v več namiznih programov. Systran je eden izmed začetnikov izdelava tehnologij in aplikacij za naravno jezikovno prevajanje. Njihove proizvode uporabljajo »U.S. intelligence community«, obrambno ministrstvo, evropska unija in več kot 500 velikih podjetij.

Ena izmed najbolj opaznih dosežkov je bila tvegana združitev z Altavisto leta 1997 in rezultat je bil BABELFISH, na svetu prvi spletni prevajalski sistem. BABELFISH je omogočil prvo brezplačno prevajalsko uslugo milijon uporabnikom. BABELFISH se je pojavil kot vodilni v

globalni internetni skupnosti in milijoni po celem svetu so bili izpostavljeni njegovim prednostim avtomatskega prevajanja.

V 21. stoletju se Systran še vedno smatra kot vodilni na področju industrije in kontinuirano proizvaja najnovejše tehnologije jezikovnega prevajanja.

Dejevniki ki jih je dobro razmisliti:

- ▶ SYSTRAN-ova tehnologija prevede več kot **4 milijone spletnih strani dnevno**.
- ▶ Več kot **300.000 strani** prevedejo z njihovo tehnologijo.
- ▶ Več kot **300.000 strank** je prijavljenih na njihovo SYTRANNET uslugo.
- ▶ Več kot **300.000 podjetij** je prijavljeno na njihovo SYSTRANLINK uslugo.
- ▶ U.S. obrambno ministrstvo, vladne agencije, podjetja ter več narodnostne grupe uporabljajo Systranovo tehnologijo, da opravijo na milijone prevajanj dnevno.

4.2.1 Izdelava strojnega prevajanja

Na začetku se je zdela ideja za dekodiranje jezikov skozi matematične tehnike nedosegljive, vendar je kmalu postala resnična po drugi svetovni vojni. V petdesetih letih prejšnjega stoletja, raziskave na izvajanju avtomatskih prevodov, danes znanih programi za prevajanje ali strojno prevajanje, je zavzelo obliko v smislu literarnega prevajanja oz. prevajanje besedo za besedo. Iz tega je bilo razvidno, da se bo to tehnologijo naprej razvijalo tako, da bo uporabljala bolj kompleksna navodila in da bo program prepoznal slovnična pravila za vsak jezik.

4.2.2 Systran – strojno prevajanje

Peter Toma, Ph. D., jezikoslovec za področje strojnega prevajanja, je začel njegovo delo leta 1957 v Kalifornijskem inštitutu za tehnologije. Nekaj let pozneje, Dr. Toma postane del projekta za razvoj Ruskega v Angleški jezik programa za strojno prevajanje na Georgetown univerzi, ki naj bi bil največji projekt v ZDA tistega časa. Leta 1968 Dr. Toma ustanovi podjetje v San Diegu, z izdelkom SYSTRAN sinonim za »System Translation«.

Kmalu za tem je bila podjetju zaupana naloga izdelava Ruskega v Angleški jezik programa za strojno prevajanje za ameriško letalstvo. Prvi SYSTRANOV sistem je bil preizkušen na začetku leta 1969 v Wright-Patterson zračni bazi v Daytonu. Od leta 1970 naprej sistem kontinuirano omogoča prevode za zračne sile tujih oddelkov tehnologij.

4.2.3 Systran – takrat in danes

Koncept dešifriranja jezikov z matematičnimi tehnikami se je pojavil po drugi svetovni vojni. Raziskave in študije avtomatskega prevajanja, danes znano kot prevajalski program, je bila v teku. Rast ni bila vzpenjajoča, brez težav in prav tako ni bila poceni. V resnici so ti poizkusi peljali do ubogega prevajanja, posledično rezultat ustavitve financiranja vlad v raziskave. Kljub temu pa so se raziskave nadaljevale.

Danes moč osebnih računalnikov ter povečana kapaciteta in hitrost omogočajo zelo impresivno kvaliteto prevajanja s katero se strinjajo tudi strokovnjaki, da je veja prihodnosti.

Preglednica razvoja Systrana:

| | |
|------------------|---|
| 1968 | Systran odpre svoje prostore v San Diegu in je najet da izdela cel Ruski v Angleški jezik prevajalski sistem za ameriško letalstvo. |
| 1968 | Prvi systran je testiran v začetku 1969. ta sistem še naprej nudi usluge za ameriško letalstvo. |
| 1974-1975 | NASA uporabi SYSTRAN prevajalsko tehnologijo pri združitvi US-USSR Apollo-Soyuz vesoljske oprave. |

| | |
|-------------|--|
| 1975 | »Commission of the European Communities« (CEC) zaposli SYSTRAN, da razvije veliko število jezikovnih parov evropske unije. Danes so SYSTRAN-ovi prevajalski sistemi v celoti integrirani v CEC sisteme. |
| 1989 | SYSTRAN je originalni razvijalec strojnega prevajanja, ki predstavi idejo o specifičnih uporabniških slovarjih; uporabnik definira slovarje, ki vsebujejo uporabnikove besede in terminologijo in jih vključi v prevajalsko proceduro, ki izboljša kvaliteto prevajanja za uporabnike. |
| 1995 | Systran Professional za Windows je lansiran in nudi cenovno učinkovito uporabo prevajanja za podjetja. |
| 1996 | SYSTRAN sprejme pogodbo v vrednosti \$10 milijonov od US NAIC (National Air Intelligence Center at Wright-Patterson Air Force Base) za razširitev kar velikega števila vzhodno evropskih jezikovnih sistemov ter prvi srbo-hrvaški v angleški sistem. |
| 1996 | SEIKO Instruments vključi SYSTRAN-ovo tehnologijo v njihov elektronski ročni prevajalnik. |
| 1997 | BableFish , prvi online prevajalnik, podprt s SYSTRAN-ovo prevajalsko tehnologijo, je lansiran na internet. Avtomatsko prevajanje postane dostopno vsem uporabnikom brezplačno. |
| 1998 | Electronic Art pooblasti SYSTRAN-ovo prevajalsko tehnologijo za online igralne izdelke. |
| 2000 | OracleMoblie.com , izbere SYSTRAN-ovo prevajalsko tehnologijo za njihov brezžični servisni portal. |
| 2001 | Autodesk ščuva prvi več jezikovni online prevod za tehnično dokumentirano podporo z namenom SYSTRAN-ove običajne prevajalske rešitve. |
| 2002 | PricewaterhouseCoopers in SYSTRAN se združita z namenom dodajanja več jezikovnih tehnologij XBRL-u »Financial Reporting Format«. Zagotovi se XML, ki podpira prevajalski program sploh prvič. |
| 2003 | SYSTRAN tehnologija podpira Atomica's več jezikovni »fact-finder« in je partner z francoskimi spletnimi portali Voila in Wanadoo. |

| | |
|-------------|--|
| 2004 | Systran Professional Premium 5.0 dobi 1. nagrado na evropskem natečaju. Ta prestižna čast je nagrajena za odlične izdelke, ki predstavljajo najboljšo Evropsko inovacijo iz področja informacijski družbenih tehnologij. |
|-------------|--|

Systran se neprestano bori, da potiska meje in presega njegove obstoječe standarde v informacijah in prevajalskih tehnologijah. Danes je znano, da je systran vodilni prevajalski operator na medmrežju z velikim številom odjemalcev kot so Altavista, Lycos, AOL, Compuserve, Terra, Google, Apple,... (5)

4.3 Strojno prevajanje

Strojno prevajanje je proces, ki uporablja računalniški program za prevod teksta v enem jeziku v drugega. Ta definicija se nanaša na slovnično strukturo vsakega jezika in uporablja različna pravila in predvidevanja za prenos slovnične strukture izvirnega jezika (tekst, ki se prevaja) v drug jezik (preveden tekst).

Prevod ni lahek. To ni prava zamenjava za vsako besedo, temveč je zmožnost vedenja vseh besed v danem stavku ali frazi in kako lahko ena vpliva na drugo.

Človeški jeziki vsebuje oblikoslovje (način kako so besede narejene iz majhnih enot), sintakse (zgradba stavka), semantike (pomena) in številnih dvoumnosti.

4.4 Človeško prevajanje

Posamezniki prevajajo tekst iz enega jezika v drugega in s tem nudijo prevajalske usluge. Ti prevajalci morajo imeti veliko znanja o jezikih, ki jih prevajajo.

V mnogih primerih je težko najti ljudi, ki imajo dovolj znanja o jezikih na visokem nivoju za prevajanje.

4.5 Razlika med strojnim in človeškim prevajanjem

4.5.1 Kapaciteta in hitrost

Prevajalec lahko dnevno prevede od 2,000 do 3,000 besed dnevno. Systranov sistem pa lahko prevede 3,700 besed na minuto in s tem prihrani na stroških in času. Cena za ročno prevajanje se giblje med 20 in 60 centi na besedo. Strošek strojnega prevajanja se povrne v prvem letu uporabe.

4.5.2 Povrnitev investicije

Ponovna uporaba ročnega prevoda ni možna, kakor koli že systran-ov sistem dovoljuje ponovno uporabo in arhiviranje prevedenih dokumentov v celoti ali pa samo po delih. Podjetja, ki uporabljajo systranovo tehnologijo dobro izkoriščajo v svoj prid njihovo investicijo.

4.5.3 Produktivnost

Najbolj zanesljiv prevod in ravnanje z več jezikovnimi dokumenti, ki zahtevajo 100% natančnost, je prevod s sistemom za strojno prevajanje in nato še ročni pregled prevajalca za pravilnost prevoda. Dokumenti, ki potrebujejo 100% natančnost so uradni dokumenti, priročniki, oglasni material in kompleksna poročila. Statistično je dokazano, da tak način prevajanja zagotovi podjetjem prihranek na času in denarju.

4.5.4 Natančnost

Natančnost teksta je pogojena z avtorjem. Oni razumejo končni prevod in postavijo pričakovane ravni. Vendar se natančnost strojnega prevajanja lahko izboljša, vsaj z uporabo systrana.

Prvič obstajajo določene smernice, ki pomagajo uporabnikom kako pripraviti in pisati besedilo za strojno prevajanje. Na primer sistem za strojno prevajanje pričakuje, da so vsi dokumenti slovnično pravilni.

Odvisno od kvalitete dokumenta, lahko uporabniki izdelajo svoje pripadajoče slovarje in s tem vključijo svojo tehnologijo. Ker se da te slovarje vključiti v prevajalski proces, je posledično rezultat tega boljša kakovost prevoda.

4.5.5 Prilagoditev

Natančnost je ključnega pomena v procesu prevajanja in potrebuje prilagoditve. Systran je vključen v celoten proces dokumentacije in komunikacije z odjemalci. Pomagajo izgradnjo tehnologij, stilov, izboljšajo berljivost, reciklirajo dobro preveden material in s tem posledično v naprej predvidijo stvari, ki zadevajo prevajanje.

4.6 Raziskave in študije

Glavno ogrodje, ki ga systran uporablja v vseh svojih sistemih za strojno prevajanje je dokazano koristno in zanesljivo. V svoji dolgi zgodovini je bilo narejenih veliko sprememb v originalnem izdelku kar je posledica velike modularnosti.

Uporaba obstoječih modulov, kot uporaba obstoječih metod na različnih jezikih kadar je možno, omogoča hiter in učinkovit razvoj funkcionalnega sistema za katerikoli nov jezikovni par.

Systranova arhitektura je prav tako zelo fleksibilna in omogoča predstavitev novih metod. V bistvu, z vsakim novim jezikovnim parom dodanim v systranov inventar se s tem preizkusijo nove tehnike, ki jih posledično izzove tisti jezik. Ponavadi se take inovacije kasneje izkažejo kot zelo koristne tudi v drugih jezikovnih parih.

4.6.1 Metodologija

Systranova metodologija je tehnologija približati se besedo za besedo, skoncentrirano na posamezne beseda v knjižnici podatkov slovarja nato na analizi stavkovnih enot, ki sledi prevod analiziranih stavkov.

4.6.2 Modularnost

Systranovo arhitekturo opisujejo tri glavne skupine: slovar, sistemski program in jezikoslovni program. Vsak od teh vsebuje veliko število modulov, ki delujejo skupaj in s tem zagotovijo avtomatski sistem za strojno prevajanje.

4.7 Slovar

Systran vsebuje tri različne vendar združene oblike slovarjev za sistem za strojno prevajanje vseh jezikov.

4.7.1 Osnovni slovar (Stem dictionary)

Osnovno slovar je eno besedi »stem dictionary«. Beseda so vnesena v osnovni obliki kodami, da prikažejo vplivni vzorec, del govora, sintaktično obnašanje, semantične lastnosti in jezikovni pomen skupaj s kodami potrebnimi za generiranje v želeni jezik. Dvourni oblike so vnesene posebej za vsak del govora prirejene osnovnim vnosom in vključene kot dvourni člen besedila.

4.7.2 Izrazoslovni slovar (Expression dictionary)

To je slovar za večbesedne izraze. Ti izrazi vsebujejo »co-occurrence-based » in »rule-based expressions« in se lahko raztezajo od preproste samostalniške fraze do izrazov ki vsebujejo prevajalska pravila bazirana na sintaktičnih ali semantičnih zvezah med posameznimi besedami, ali celi celotne razrede besed. Besede v tem slovarju so podane v osnovni obliki in se lahko z vključitvijo v osnovni slovar omogoči izvajanje pravil, ki veljajo v osnovnem slovarju.

4.7.3 Slovar za specifičnega uporabnika(Customer Specific Dictionary - CSD)

PC/Windows zasnovan slovar dovoljuje uporabniku vnos terminov (besed in skupek prej obdelanih oblik), ki se ne nahajajo v osnovnem slovarju. Uporabnik prav tako lahko popravi ali spremeni pomen besed, ki se nahajajo v osnovnem slovarju. Slovar za specifičnega uporabnika je narejena za posameznika ali pa za industrijsko uporabo z omejenimi potrebami.

4.8 Sistemski program

Telo sistema programa, obstoječ v vseh systranovih jezikovnih parih dela z oblikovanjem, znakovnimi zamenjavami, uporabniškim vmesnikom, besednimi in stavkovnimi mejami, pregled slovarjev in morfologije ter ravnanje z ne najdenimi besedami. Kontrolira pretok jezikoslovnih modulov in izdela končno obliko. Prav tako so podprte številne variante orodij za pripravo slovarja, kvaliteto zaupanja, manipulacijo s korpusom in diagnostičnimi analizami.

4.9 Jezikoslovni program

4.9.1 Analizator

Največji izziv vsakega sistema za strojno prevajanje je analizator, to je modul ki analizira vsak stavek in poskuša narediti primer izvornih stavkov. Systran analizira s številnimi proceduralnimi moduli, kateri rešujejo korak za korakom številne sintaktične in semantične zveze in določi strukturo znotraj stavka. Systranov analizator je odločen po naravi, tako da vsak modul naredi trdni odločitev in odda rezultat na naslednji modul. Prednost je v tem, da vsak stavek, tudi če je nepopoln, bo analiziran in tako preveden. Pomanjkljivost take odločnosti je, da so lahko napačne odločitve prenesene naprej in sestavljene iz modula v modul. Systran je zmožen to zmanjšati z nekaj mehanizmi, ki zmanjšajo netočne odločitve. Systranov končni korak v tem procesu preverjanja je »filter program«, ki razpozna glavne napake analizatorja.

4.9.2 Ciljni prevajalski jezikovni moduli

Po analizi strukture vhodnih stavkov se naredijo algoritmi za strukturo prevoda. Prevajalske informacije za besedno in izrazoslovno raven je narejen med pregledom slovarja in rešitvenih faz prevoda za uporabo dveh modulov (transfer in synthesis). »Transfer« komponenta omogoča specifične glede na situacijo prestrukturiranja, odvisno od stopnje razlik med izvornim in končnim jezikom. Je edini modul, poleg slovarja, ki se nanaša na izvorni in končni jezik in je precej majhen, ko sta dva jezika podobna.

4.9.3 Strnitveni modul (Synthesis modul)

Strnitveni modul generira stringe izvirnega jezika, ki ustreza informacijam preskrbljene z prejšnjimi moduli. Ta modul je neodvisen od izvirnega jezika. Vsebujejo posebne algoritme za izdelavo specializiranih zgradb končnega jezika, kot so negacije, vprašanja, glagoli s celotnim oblikoslovjem, postavitev prislovov in členkov,...

4.10 Izdelava dodatnih jezikov

Izdelava novih jezikovnih prevajalskih parov za katere ima systran že izvirne in končne module je lahko doseči. Samo nov prenos modula in transfer/target slovarje je potrebno narediti.

Izdelava določenih ciljnih jezikovnih zmožnosti za vsak izvorni sistem je mogoča in ekonomična zato ker so systranovi sistemi kot več-ciljni sistemi. Za dodajanje novega ciljnega jezika je potrebna samo izdelava novega prenosnega modula in novega strnitvenega modula ter transfer/target slovarjev.

Izdelava določenih izvornih jezikovnih zmogljivosti za vsak ciljni sistem je bolj težka, če je potrebno izdelati popolnoma nov analizator. Kakor koli že, če je nov izvorni jezik podoben enemu, ki že obstaja v systranovih izvornih jezikih, izdelava novega analizatorja lahko pomeni prednost pri običajnih pravilih znotraj jezikovne družine z uporabo obstoječega »Trunk analizatorja« (kot je romance Trunk, Slavic Trunk,...) (6)

4.11 Systran – tehnologija

Systran je eden izmed prvih in ostaja eden izmed redkih neodvisnih razvijalcev orodij za strojno prevajanje. V letu 2002 je systran izdal preoblikovano arhitekturo, ki je bila posledica 30 letnega dela. Združen z izboljšanim delovanjem današnjih računalnikov je posledica tega višja prevajalska pokritost.

Njegova tehnologija je narejena pod Linuxom in deluje na vseh Unixovih platformah ter Windowsih. Vsi systranovi sistemi izkoriščajo eno napravo in uporabljajo zadnje tehnologije

naravnega jezikovnega procesiranja (NLP). Sistem prav tako vsebuje tehnologijo ki pospeši dostop do zelo velike jezikoslovne baze podatkov. Dejstvo je, da ima Systran večjo količino več jezikovnih baz, kot jih lahko podpira običajen računalnik. XML, Unicode in http standardi so centralnega pomena v izdelavi, ko se naslavlja več jezikovne zadeve v aplikacijah kot so splet, elektronska pošta, intranet, tisku,... Systran ter R&D se trudita pri neprestanem razvoju in investiranju v jezikoslovje ter NLP. Njihov cilj je, da se v prihodnosti naredi novo generacijo sistemov za strojno prevajanje, ki bodo izboljšali kvaliteto prevajanja ter robustnost »muli-agent« in »multi-treaded« arhitekturah.

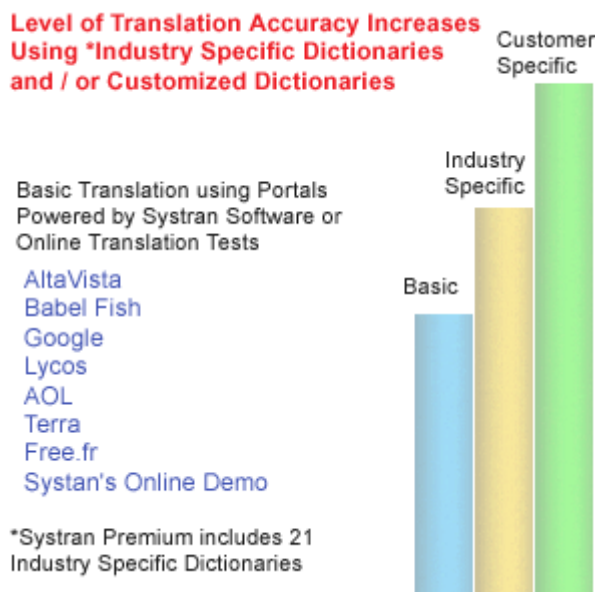
4.12 Standardno strojno prevajanje proti prirejenemu

Zelo pomembno je, da se razlikuje med generičnim strojnim prevajanjem in prirejenim strojnim prevajanjem. Systranova tehnologija, ki oskrbuje večino spletnih portalov je generična prevajalska usluga, ki deluje na osnovi največjih možnih slovarjih. Njen namen je, da zagotovi »gisting level« prevajalsko uslugo kot je vpogled v tujo stran in zmožnost razumevanja osnovnih linij strani. Sistem mora dobiti ne samo pravih sintaktičnih analiz teksta, ampak tudi pravih semantičnih analiz. To je zato, ker imajo nekatere besede različne pomena in se drugače obnašajo v odvisnosti od semantične vsebine v kateri so uporabljene. Celo s posebnimi specializiranimi slovarji kvaliteta prevoda ne more biti vedno popolnoma zadovoljiva, zaradi svobode in produktivnosti vsebine na spletu; izbrani slovarji ne morejo biti točna domena besedišča; nemogoča naloga je imeti celoten opis vseh možnih domen (problem ni samo cena ampak tudi znanje). Ključna točka je, da noben generični sistem za strojno prevajanje danes ne more rešiti vseh problemov.

Kljub temu se Systran smatra kot najboljši prevajalski sistem pa še vedno ne more nadomestiti človeškega prevajanja, ker je jezik še vedno preveč kompliciran, vendar se lahko izboljša natančnost prevoda z izdelavo svojih slovarjev ali prilagoditvijo programa z industrijsko terminologijo.

Prirejanje programa lahko izboljša natančnost prevoda z uporabo posebnih industrijskih slovarjev, kateri so vključeni v boljše systranove verzije ali pa izdelava svojega posebnega slovarja.

Možno je tudi narediti spisek besed katere se ne želi, da se jih prevede (recimo imena podjetij).



4.13 Stvari, ki jih je dobro upoštevati za prevajanje

Kadarkoli se želi narediti dokument, ki ga želimo prevesti z programom za prevajanje je potrebno uporabljati pravilno slovnico ter priporočila proizvajalcev programov in tako izboljšamo kakovost prevoda.

Natančnost prevoda je zelo odvisna od vnosa. Tukaj je nekaj nasvetov za izboljšanje prevoda:

- Vnesen tekst mora biti slovnično pravilen
- Uporabljati je potrebno cele stavke
- Postavitev ločil je zelo pomembna; stavki bodo najboljše prevedeni, če bodo ločeni z vejicami
- Potrebno je uporabljati preproste stavke
- Izogibati se je potrebno dvoumnosti
- Izogibati se je potrebno kratic, okrajšav in pogovornih besed

Potrebno je biti direkten; pisati je potrebno preprosto. Systran lahko dela s kompleksnimi strukturami stavkov, vendar se je kljub temu potrebno izogibati preveč zapletenih stavkov,...

Bodi kratek. Potrebno je uporabljati kratke stavke in ne sme se izpuščati pomembnih slovničnih elementov.

Ne sme se izpuščati pomembnih besed.

Izogibati se je potrebno jezikovnim posebnostim in slenga. Posebnosti in sleng, ki ga želimo prevesti se razlikujejo iz države v državo.

Pravilno je potrebno postaviti ločila. Vedno je potrebno preveriti če so ločila pravilno postavljena. Ti elementi nudijo vodila tako za ljudi kot za računalnike. Z njimi ločimo stavke v logične dele. Brez njih se lahko stavke interpretira na več načinov.

Preveriti je treba, da ni slovničnih napak. Še predno kliknemo na gumb za prevod uporabimo korekturo. Če podamo računalniku napačne informacije, potem bo bil prevod napačno interpretiran.

Uporabljati je potrebno člene če je le mogoče. Člen je beseda, ki pokaže samostalnik in nakaže njegov namen.

Stalna uporaba terminologije in okrajšav. Vedno je potrebno uporabljati isto besedo, frazo ali okrajšavo za objekt ko se pojavi v dokumentu.

Izdelava specifičnih slovarjev s systranom. Izdelava takšnega slovarja lahko izboljša natančnost prevoda.

Ponovno prevedi in potrdi. Ko smo že prevedli svoj tekst se lahko doda besede v slovar. Ko so spremembe shranjene, ponovno prevedemo tekst, da potrdimo efekt novih vnosov.

4.14 Prevajalski program za izobraževalne ustanove

Potreba po strojnem prevajanju v izobraževalne namene je v porastu. Systranov prevajalski program je najbolj pogosto uporabljen program na univerzah, srednjih šolah, osnovnih šolah in tehničnih univerzah širom sveta.

Prevajalski program se uporablja na različne načine; v izobraževalne namene kot lažje sporazumevanje z otroci ali starši tako, da se dokumenti prevede v njihov jezik. Mnogokrat se zgodi, da čeprav otroci govorijo recimo

angleško starši ne znajo in tukaj pride uporabnost programa tako, da se izdelajo dokumenti v angleškem jeziku in nato se jih prevede v drug jezik.

Možen je prevod direktno iz Microsoft Offica, ki vsebuje **word, excel, power point, outlook** in **internet explorer**, ki je najbolj pogosto uporabljen paket v šolah. Prevajanje spletnih strani je tudi lahko z uporabo IE pluginov, brskanje v različnih jezikih je olajšano, odličen je za poučevanje v razredih. Uporaba je zelo lahka, ker ni potrebno nobenega posebnega znanja sej se po integraciji v office pojavi orodna vrstica s katero je preprosto rokovati. Najbolj pogosta uporabljena paketa na področju izobraževanja sta systran professional standard in premium.



4.15 Prevajalski program za vladne ustanove

Za vladne agencije je z leti postal prevajalski program zelo pomembno orodje. Na tem področju ima systran dominantno vlogo.

Že več kot 30 let nudi svojo tehnologijo skoraj vsem ameriškim vladnim agencijam, evropski uniji, združenim narodom. Zaradi vse večje vloge arabskega sveta in dogodkov je systran celo dodal angleško – arabski par. Na listo jezikov, ki so pod razvojem.



Najbolj pogosta uporabljena paketi na področju vladnih ustanov sta systran Professional standard, Premium ter server software systran enterprise edition. (5)

4.16 Systran – produkti

Systran personal 5.0

Je odličen za domače uporabnike, kateri potrebujejo prevod manjše količine besedila, elektronskih sporočil, krtkih paragrafov, manjših pisem,... Z njim je možno prevesti do 5 kb besedila z različnih aplikacij Windows. Ta tekst se avtomatsko prevede.

Systran office translator 5.0

Prevede do 5 strani dokumenta Microsoft Worda, PowerPointa, Excela ter internet Explorer spletne strani direktno iz Office namizja.

Systran Professional standard 5.0

To je celotna verzija prevajalskega programa. Z njim se prevaja velike dokumente, vgrajen ima uporabniške slovarje, prevaja PDF, DOC, RTF in HTML datoteke. V celoti se ga da

integrirati v Office aplikacije. Vsebuje tudi napredna orodja za bolj natančno prevajanje. Ta verzija vsebuje vse pluggine za Microsoft Office, PDF in internet Explorer.

Systran Professional premium 5.0

Največja natančnost: ta verzija ima lasnosti Professional standard plus dostop do še bolj naprednih orodij, ki lahko povečajo natančnost prevoda in s tem omogočijo še bolj učinkovit potek dela. Vsebuje 21 dodatnih industrijskih slovarjev s tisočimi termini, področij računalništva, medicine, strojništva,... Ta verzija prav tako vsebuje pluggine za Microsoft Office, PDF in internet Explorer.

Systran PDF Translator

PDF postaja standard za izmenjavo velikih dokumentov. S tem programom se lahko prevede do 5 strani PDF datotek. Datoteke se prevedejo v sekundi

Systran WebTranslator 5.0

Prevede spletne strani nemudoma brez da bi bilo potrebno zapustiti Microsoft internet Explorer. Prevede v realnem času medtem ko smo na liniji ali pa shrani strani ko smo izključeni. (7)

4.17 SystranLinks

SystranLink deluje tako, da pravzaprav linka spletno stran na njihov prevajalski server. Z izbiro jezika na spustnem oknu se prevod spletne strani izvrši nemudoma.

- Hitra in lahka nastavitev
- Prevodi se naredijo na systranovih severjih
- 30 jezikovnih kombinacij
- 4 stopnje online prevajalskih uslug

4.18 SystraNet

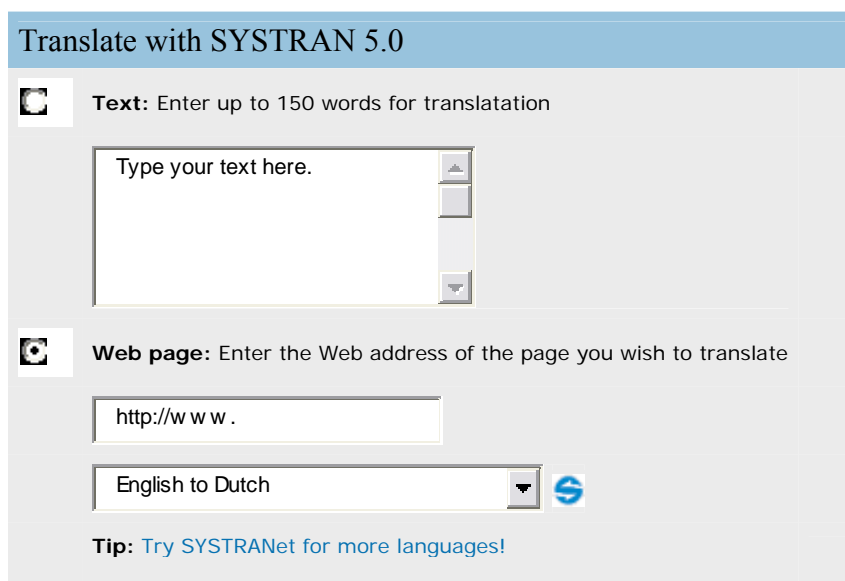
Na uslugo 24/7 systraNet prevede osebne datoteke, spletne strani in tekst in v nekaj sekundah se prevod prikaže na brskalniku ali pa je poslan po elektronski pošti.

- Več kot 35 jezikovnih parov
- Zelo veliko specifičnih slovarjev
- Možna izdelava svojega slovarja ki pomaga izdelati visoko kakovosten prevod
- Ohrani original dokumenta

4.19 SystranBox

SystranBox je idealna rešitev strojnega prevajanja za portale in ponudnike vsebin. S svojim znanjem so izdelali AltaVista BabelFish. Danes vodilni portali kot so AOL, Lycos, Terra, Google, Voila, Wanadoo,... uporabljajo SystranBox.

Z njim lahko dodamo prilagoditveni prevajalski okvir spletni strani in z njim povečamo promet in profit. Uporabniki lahko prevajajo tekst, spletne strani in lahko brskajo po internetu v svojem jeziku.



Translate with SYSTRAN 5.0

Text: Enter up to 150 words for translation

Type your text here.

Web page: Enter the Web address of the page you wish to translate

http://w w .

English to Dutch

Tip: Try SYSTRANet for more languages!

4.20 Odjemalec – strežnik aplikacije

Veliko podjetij potrebuje okolje, ki dovoljuje uporabnikom da dostopajo do prevajalskega okolja z namenom prevajanja dokumentov.

Systran Enterprise omogoča vrhunsko tehnologijo prevajanja za na osnovi na mreži zasnovano odjemalec/strežnik okolje ali medmrežno strežniško rešitev.

Systran Enterprise odjemalec/strežnik je lahko nastavljen, lahko je z njim rokovati ter narejen za uporabo velikemu številu uporabnikov. Zagotavlja na zahtevo dostop do strežnika za prevajanje iz katerega koli uporabnika, ki je priključen na skupno mrežo in ima ta program. Ker vsebuje prevajalske procese ter vse združene uporabniške slovarje, prevajalski spomin in normalizirane slovarje na enem strežniku je rezultat konsistenten prevod ter neverjeten prihranek na času.

Systran prav tako nudi dostop za večje število uporabnikov z medmrežnim strežnikom, ki pošilja preko našega spletnega iskalnika dokument za hiter prevod.

Systran enterprise 5.0

Po meri narejen za intranet, extranet ali LAN okolje. Glavne lastnosti:

- Dve vmesniški opciji (na mreži zasnovano ali Office plugini)
- Prevede direktno in namizja v Word, Explorer, Outlook, Excel, PowerPoint, HTML, PDF, RTF, TXT in XML
- Sodelavci si delijo uporabniške ali lastne slovarje
- 5 specializiranih slovarjev
- Vgrajena so prilagoditvena orodja za izboljšanje kvalitete prevoda
- Vgrajen prevajalski spomin – podpira TMX (Translation Memory eXchange)
- Povečana je produktivnost – prihranek na času in denarju
- Izredno hitri prevodi
- Program se lahko prilagodi specifično podjetju ali domeni
- Rezultat večje kvalitete je dosežen z naprednimi prilagoditvenimi orodji ali prevajalskim spominom
- Ponovna uporaba slovarjev, TM in prejšnjih slovarjev

Dve vmesniške opcije povečajo zmogljivost Enterprise5.0 za posamezno delovno postajo:

- *SYSTRAN Online Tools*

Na mreži zasnovan vmesnik ki omogoča prevod TXT, RTF, DOC, PDF, XML in HTML datoteke. Vsebuje opcije ki pomagajo izboljšati kvaliteto rezultatov tako, da omogočijo dodajanje terminologije prevajalskemu procesu.

- *SYSTRAN Application Pack*

Uporabniški program, ki omogoča direktni dostop do prevajalskih zmožnosti s pomočjo Office pluginov (Word, Outlook, Excel, PowerPoint, in spletne strani na internet Explorer)

Systran WebServer 5.0

To je najbolj obsežna in močna rešitev za prevajanje danes. Če ima podjetje željo po integraciji zelo zmognega programa za prevajanje zmožnostjo iskanja, podjetniško zadovoljivega managementa, pogovora, elektronskega poslovanja, elektronskega učenja ali drugih aplikacij je potem Systranov Webserver najboljši izbor.

Zmožen je podpre neomejenega števila uporabnikov in milijon prevodov dnevno kot je prikazano z vodilnimi iskalniki kot so Google in Yahoo. WebServer uporablja SOAP na TCP/IP protokolu in s tem na lahek način komunicira z aplikacijami, ki upravljajo z tekstovnimi dokumenti in je lahko celo integriran v delovni proces podjetja. Proces se lahko nadgrajuje kadar koli in je sinhroniziran s slovarji podjetja in drugimi uporabniškimi skladišči, ki so pomemben vir za izboljšanje kvalitete prevoda. Glavne značilnosti:

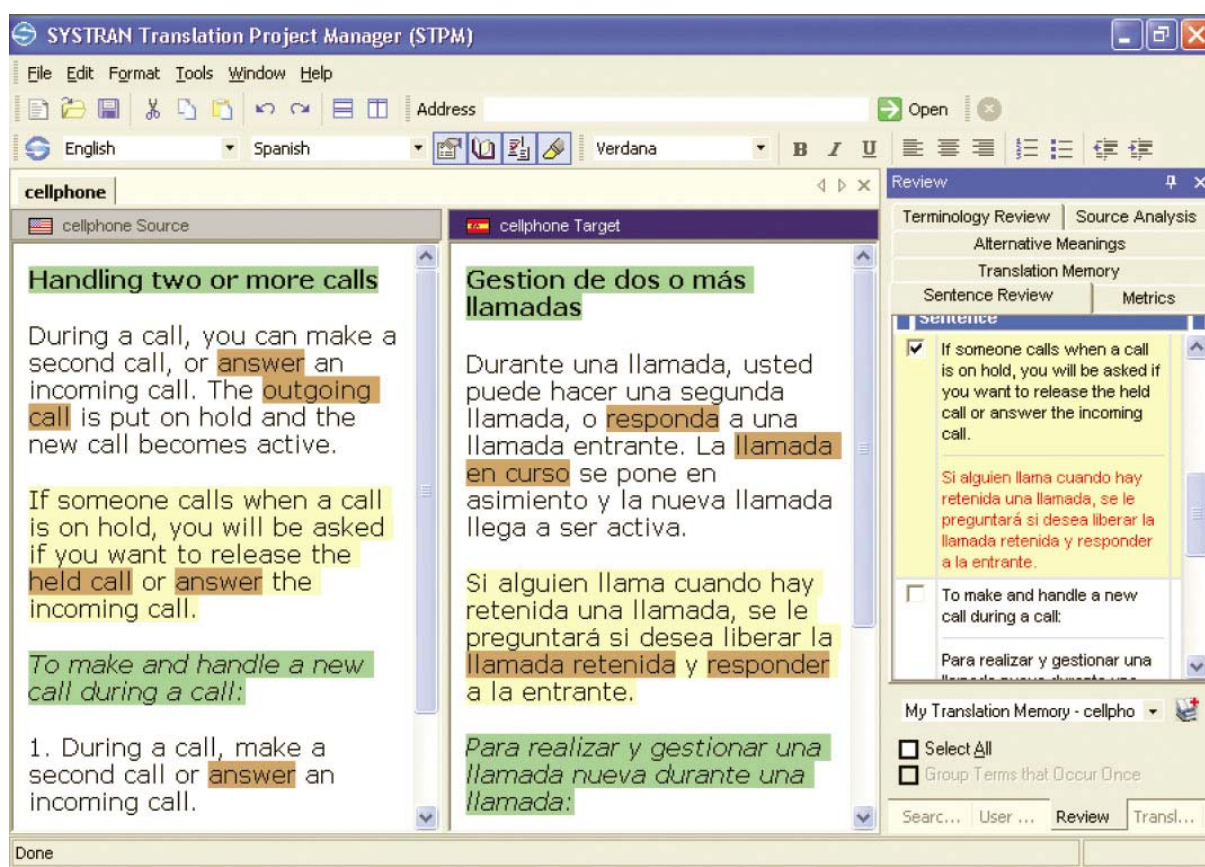
- Obsežna in odprta arhitektura
- Več operacijski sistemi – Linux, Unix, Windows
- API (application program interface) za vgradnjo
- Zmožnost filtriranja za DOC, XLS, PPT, PDF, TXT, RTF, HTML, XML in druge formate
- Vgrajen prevajalski spomin – podpira TMX (Translation Memory eXchange)
- Razumevanja tujih jezikovnih vsebin v realnem času
- Povečana je produktivnost – prihranek na času in denarju

- Prevodi so statični ali dinamično odvisno od naših potreb
- Dodaja več slovnične sloje obstoječim aplikacijam kot so iskanje, podjetniško zadovoljivega managementa, pogovora, elektronskega poslovanja, elektronskega učenja,...
- Lahko se ga prilagodi posebej za podjetje ali domeno
- Efektiven pri kompleksnih prevodih
- Zmožen je odprave večkrat pojavljenih napak enkrat
- Izdelane prilagojene spletne prevajalske usluge; SYSTRANLinks, SYSTRNBox, SYSTRANet
- *SYSTRAN Online Tools*
Na mreži zasnovan vmesnik ki omogoča prevod TXT, RTF, DOC, PDF, XML in HTML datoteke. Vsebuje opcije ki pomagajo izboljšati kvaliteto rezultatov tako, da omogočijo dodajanje terminologije prevajalskemu procesu.
- *SYSTRAN Application Pack*
Uporabniški program, ki omogoča direktni dostop do prevajalskih zmožnosti s pomočjo Officovih pluginov (Word, Outlook, Excel, Powerpoint, in spletne strani na internet explorer) (8)

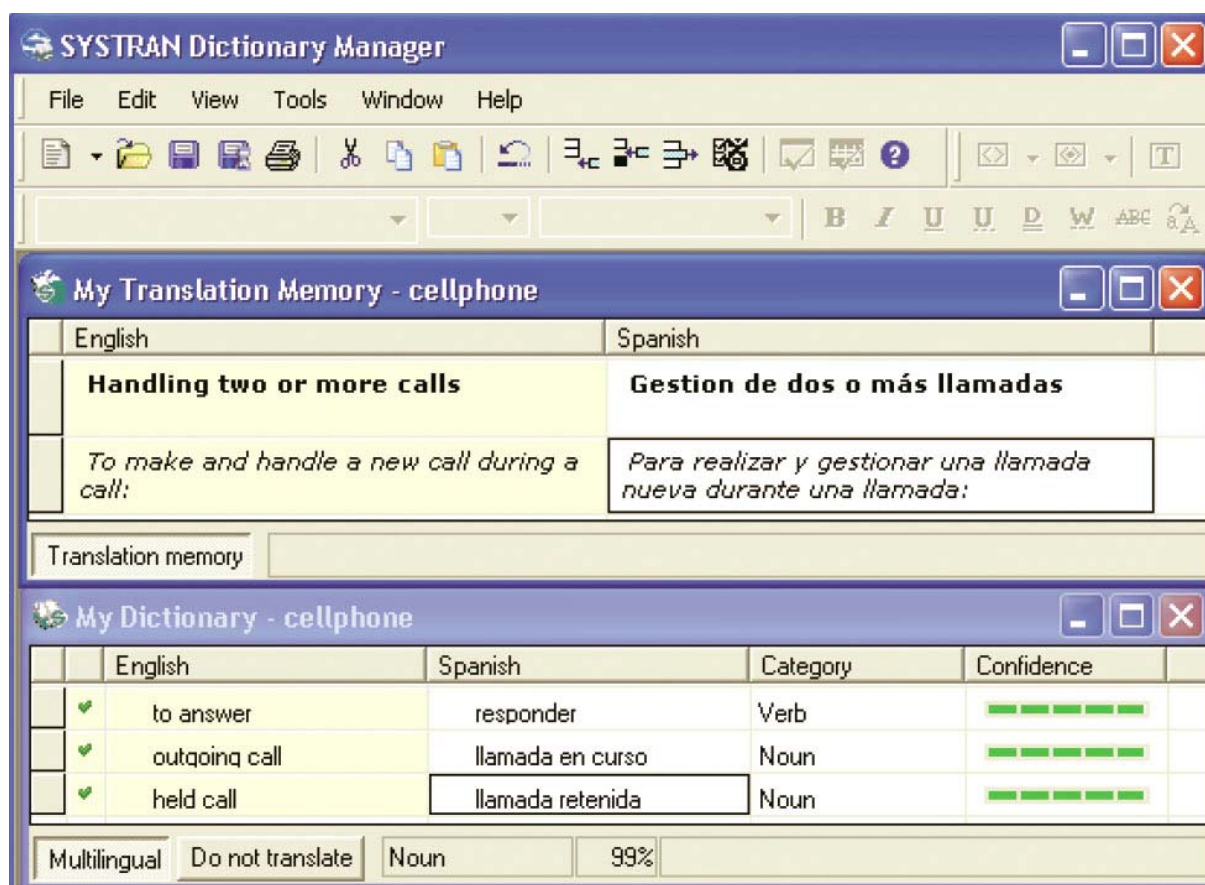
4.21 Systran – orodna vrstica



4.22 Systran – Translation Project Manager (STPM)



4.23 Systran – Dictionary Manager



7 ZAKLJUČEK

V zadnjih letih so tehnologije iz področja jezikovne tehnologije precej napredovale. Tako od sinteze govora, razpoznavanja in strojnega prevajanja. Te tehnologije se smatrajo kot ključne in so v porasti. Današnje družbe si ni več moč predstavljati brez njih, ker omogočajo hitrejšo komunikacijo, izobraževanje, združevanje narodov. Kljub temu pa, da je tehnologije že zelo napredovala in so bili narejeni že revolucionarni dosežki, pa je moč trditi, da je še veliko neraziskanega. Dve vodilni podjetji Systran in Nuance (prej imenovan ScanSoft) se trudita in težita po razvoju novih in vse bolj naprednih tehnologij, da bi zagotovila svojim uporabnikom orodja in okolja, ki bi zadovoljila njihovim potrebam.

8 VIRI IN LITERATURA

- (1) <http://www.microsoft.com/speech/evaluation/techover>
- (2) http://laps.fri.uni-lj.si/dps_arhiv/seminarske/indihar/sinteza.htm
- (3) <http://www.nuance.com>
- (4) <http://www.systransoft.com/index>
- (5) <http://translationsoftware4u.com/translation-software.htm>
- (6) <http://www.omega-first.co.uk/faqs/01.html>
- (7) <http://imagiforce.com/systran-personal.htm>
- (8) <http://wwwsystransoft.com/products/>
- (9) Priročnik v PDF formatu: SYSTRAN Professional Premium 5.0