

Darko Simeršek

Prevajanje z računalniki*

Človek si že od nekdaj želi, da bi namesto njega prevajal stroj, ni mu vsaj pomagal pri prevajanju. Prevajanje iz enega jezika v drugega je poseben proces, kjer se spremeni le zapis pomena, pomen sam pa ostane nespremenjen. Kako človek prevaja, še ni povsem razjasnjeno. Najverjetneje tako, da najprej razbere pomen iz jezika, ki ga prevaja, nato pa poizkuša isti pomen izraziti v drugem jeziku.

Računalnik danes sicer uporabljamo za prevajanje besedil, vendar prihaja zaradi zamotanosti naravnih jezikov pri tem opravilu do različnih težav. Strokovnjak mora vedno pregledati prevedeni tekst ter odstraniti morebitne nejasnosti, preden ga posreduje naprej.

Naravni jeziki, ki jih uporabljamo za sporazumevanje, so po večini nastajali v dolgih stoletjih. Zaradi dolgega in neusklajenega razvoja imajo številni jeziki mnogo pomanjkljivosti. Na primer: ista beseda ima vrsto različnih pomenov; poznamo različne besede, ki pa imajo isti pomen; skladnja besed je zamotana; slovnica je obsežna, ima številna pravila in hkrati izjeme, ki odstopajo od teh pravil.

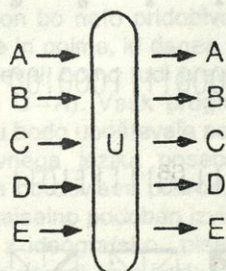
Če želimo, da računalnik opravi pomen-sko pravičen prevod, mora upoštevati vse možne pomena določene besede in jih posamično primerjati z drugimi besedami v stavku. Pri tem mora seveda upoštevati stavčno skladnost in druge posebnosti jezika, v katerega prevaja.

Tudi če se pri tem opravilu omejimo le na zanesljiv prevod smisla besedila in se odpravimo leporečju, je naloga, ki naj jo opravi računalnik, tako zahtevna, da postaja izvedljiva šele v zadnjem času.

V začetku prejšnjega desetletja so razvili prevajalni sistem Systran, kasneje Euro-Tra, nadaljnje raziskave pa nakazujejo še druge obetavne sisteme za prevajanje

Predstavljajmo si, da smo naročeni na publikacije, ki jih tiskajo – na primer – založniki držav: Finske, Japonske, Alžirije in Grčije. Ko jih dobimo po pošti, jih vtaknemo v posebno napravo z optičnim čitalnikom in preberemo tekst na zaslonu – v slovenščini.

Utopija?



Slika 3

B → A	A → B	A → C
C → A	C → B	B → C
D → A	D → B	D → C
E → A	E → B	E → C

A → D	A → E
B → D	B → E
C → D	C → E

Slika 4 E → D D → E

* Življenje in tehnika, oktober 1986.

Novi sistemi za prevajanje

Zgoraj navedeni primer ni tako utopičen, kot je videti na prvi pogled. Izvedba je seveda možna le z uporabo ustrezne tehnologije, računalnika in programske opreme. Če bi želeli uporabiti programe, ki bi bili sposobni prevajati iz drugih jezikov (v našem primeru iz finskega, japonskega, arabskega in grškega jezika), bi najprej naleteli na problem, kje dobiti te programe. Zataknilo bi se že pri finskem jeziku, saj doslej ni izšel pri nas niti navadni finsko-slovenski slovar. Zasnova prevajalnega programa pa je precej zapletena naloga, še posebej če upoštevamo, da ima vsak od omenjenih jezikov svojo pisavo.

Omenili smo že, da računalnik najlažje razbere tekst, če je ta zapisan v kodirani obliki. Tekst lahko obdelamo tako, da vsakemu geslu (pojmu, predmetu itd.) ustreza določena mednarodno dogovorjena koda, ter ga zapišemo v softstripu. Pri prebiranju teksta bomo uporabili le en prevajalni program, ker je tekst v softstripu že zapisan v dogovorjenem jeziku (imenujemo ga univerzalni jezik U).

Na sliki 3 vidimo v poenostavljeni obliki, kako poteka tako prevajanje. Črke A, B, C, D in E pomenijo naravne jezike, ki jih govorijo v posameznih državah. V državi, kjer govorijo jezik B, bodo uporabili prevajalni program B→U ter zapisali tekst v dogovorjenem univerzalnem jeziku U. V državi, kjer govorijo jezik C, bodo uporabili program C→U itd. Če sami govorimo jezik A, bomo za razumevanje teksta uporabili en sam prevajalni program: U→A. Tekst bomo prebrali v svojem jeziku, čeprav je bil zapisan v naravnem jeziku B, C, D ali E.

Iz slike 4 je razvidno, da zahteva bolj neposredno prevajanje iz enega v drug naravni jezik več različnih prevajalnih programov (pri petih jezikih jih imamo 20). Prevajanje iz štirih jezikov (B, C, D, E) v našega (A), zahteva štiri različne prevajalne programe. Nekdo, ki govori jezik B, bo potreboval še dodatne štiri programe (za prevod iz jezikov A, C, D, E) itd.

V svetu govorijo danes okoli 3000 jezikov. Če bi sestavili programe za te jezike v vseh možnih kombinacijah, bi bilo devet milijono

nov različnih prevajalnih programov premalo. Za razumevanje teksta pri prevajanju s posrednikom (v našem primeru je to jezik U) pa zadostuje en sam prevajalni program. posrednik bi lahko bil tudi eden od najbolj razširjenih naravnih jezikov, vendar je v tem primeru večja nevarnost, da se spremeni smisel sporočila med prevajanjem. Kode, ki jih ima jezik U, so namreč zelo obsežne in veliko bolj natančno ponazorijo pravi pomen določenega gesla, kot jih lahko zajame kombinacija črk v naravnem jeziku. Predstavljajmo si besede v naravnem jeziku, ki sicer zelo natančno določijo neki pojem, vendar imajo večje število znakov – na primer – 20 ali več črk. Pri sporazumevanju bi nas tako dolge besede motile, računalnika pa dolge in zapletene kode niti najmanj ne ovirajo.

Zasnova novega jezika seveda ne bo potekala brez težav. Določitev gesel in ustreznih kod bo morala temeljiti na mednarodnem dogovoru. Programiranje bo težavno, vendar si lahko pomagamo v našem primeru s simboli.



000000|000000|00 = 1



|||000|||000|00 = A



00|||0|||0|||0|| = *



Slika 5: številka 1, črka A in znak so predstavljeni s 16-bitnim zlogom. V spodnji vrsti je zapisano geslo. Enak simbol spredaj in zadaj predstavlja ločnico med gesli, ko si sledijo drugo za drugim

Unikon

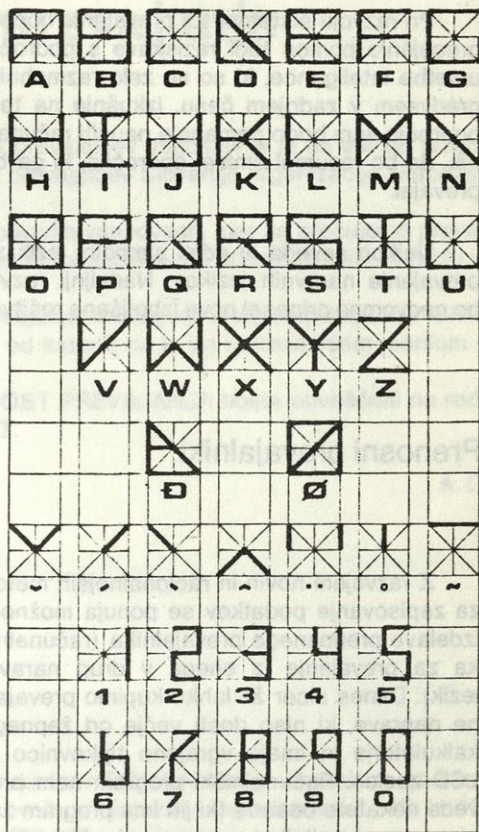
Idejo programiranja s simboli bomo najlažje pojasnili, če povzamemo idejo, ki je bila v naši reviji že predstavljena (v št. 11, 12/1983 str. 125). Namesto da bi geslo zapisali z binarno kodo, jo lahko ponazorimo 16-segmentni lik unikon. Na sliki 5 vidimo, kako tvorijo posamezni segmenti 16-bitno računalniško besedo ter nato nekaj primerov simbolov, ki skupaj sestavljajo geslo. Vsi dogovorjeni simboli skupaj tvorijo simbolni jezik unikon – ime je izpeljano iz besede unik – ki je namenjen računalniškemu prevajanju iz enega v drug naravni jezik.

Simbole uporabljamo predvsem pri zasnovi unikona, ker je takšno programiranje enostavnejše. Namesto simbolov bi sicer lahko zapisali tudi številčno kodo, vendar ima programiranje s simboli nekatere prednosti.

Gesla bomo tvorili z enim od 65535 simbolov oziroma z njihovimi medsebojnimi kombinacijami. Upoštevali bomo tudi različne slovnične oblike posameznih gesel. Zapisali bomo dodatni simbol za ednino, nekoliko drugačen za množino itd. Pri zapisovanju imen, nazivov in številčnih informacij pa bomo uporabili simbole, ki ponazarjajo črke in številke (slika 6).

Unikon bo nato pridobival nova gesla za predmete in pojme, ki danes še niso poznani. Izpopolnjevali bomo tudi prevajalna programa (A→U in U→A). Vsak program bo vseboval pravila, ki bodo upoštevala posebnosti vsakega naravnega jezika posebej. Od zasnove programa bo odvisen prevod, ki naj bi bil kar najbolj smiselno podoben izvorniku. Tisočletne izkušnje z ideografsko pisavo (kitajsko, japonsko in drugo) nam bodo pomagale pri programiranju v unikonu.

Danes težko napovemo, v kolikšni meri je sploh izvedljiva ideja prevajanja po sistemu unikon. Pri takem prevajanju gotovo ne bo šlo brez težav. Ne smemo namreč pozabiti na po-



Slika 6: črke, diakritična znamenja in številke so predstavljene s 16-segmentnim likom

sebnosti naravnih jezikov, ki bodo trd oreh za vsak računalnik – na primer pri prevajanju fraz. Prevodi leposlovja se bodo verjetno smiselno bolj oddaljili od izvornika kot prevodi tehnične in znanstvene literature.

Dobro zasnovan prevajalni program bo tudi popravljaval slovnične napake in pravilno uredil besedni red v stavku. Uporabnik bo le zapisal tekst in uporabil programa A→U in U→A. Če se bo pri tem spremenil smisel zapisa, bo to pomenilo, da prevajalni program še ni dobro zasnovan ali pa geslom še nismo povsem natančno določili njihovega pravega pomena.

Pri razvoju sistemov za prevajanje bodo v precejšnji pomoč tudi raziskave s področja umetne inteligence, ki so se zelo razmahnile predvsem v zadnjem času. Izkušnje na tem področju nam bodo pomagale naučiti računalnik, da bo razumel smisel sporočila, ki ga bo prevajal.

Unikon seveda ni edini simbolni jezik za prevajanje naravnih jezikov. Nadaljnji razvoj bo nedvomno prinesel nove izboljšane rešitve.

Prenosni prevajalniki

Z razvojem novih in racionalnejših metod za zapisovanje podatkov se ponuja možnost izdelave prenosnega prevajalnika (računalnika za prevajanje iz enega v drug naravni jezik). Danes sicer že lahko kupimo prevajalne naprave, ki niso dosti večje od žepnega kalkulatorja in imajo vgrajeno tipkovnico in LCD zaslon. Računalniški program nam prevede nekatere besede (ki jih ima program zapisane v pomnilniku) v enega ali več jezikov. Vendar ima program precejšnje omejitve, saj je računalniški pomnilnik majhnega obsega, poleg tega pa ne zna prevedi pravilno in smiselno prav vsak zapisani stavek.

Z zapisom na optično ploščo, ki je enakih oblik in velikosti kot znana CD, pa bo sposobnost računalnika precej večja. Danes že zapišemo nanjo za 552 Mb podatkov, kar ustreza 270.000 tipkanim stranem teksta formata A4. Pri sistemu CD-ROM DRIVE firme Hitachi je sleherna informacija dosegljiva v sekundi.

Najmanjše prenosne naprave za predvajanje CD plošč imajo volumen komaj polovico kubičnega decimetra. Novi prevajalnik bo seveda nekoliko večji, saj bo poleg elektronskih in mehanskih delov vseboval ploskovni zaslon, mikrofon in mali zvočnik.

Prevajanje bo zelo enostavno. Uporabnik bo izrekel nekaj besed v mikrofon, preveril na zaslonu, ali jih je računalnik pravilno zapisal, ter nato izdal ukaz za prevod. Računalnik bo tekst zapisal na zaslonu ali sporočil prevod preko zvočnika. Če bomo potovali po svetu in ne bomo znali jezika države, kjer bomo na obisku, si bomo pri komuniciranju z domačini pomagali s prevajalnikom.

Danes težko napovemo, ali lahko pričakujemo opisani prevajalnik že proti koncu tega stoletja ali šele v naslednjem. Pravzaprav posegamo pri napovedovanju novih naprav in sistemskih rešitev že na področje znanstvene fantastike. Čim dlje segajo naše napovedi v prihodnost, večja je verjetnost, da se bomo zmotili. Kljub temu poskušajmo razmišljati o nadaljnjem razvoju recimo v 21. stoletju ali kasneje.

Če bomo takrat že izpopolnili vesoljska vozila ter razvili druge ustrezne tehnologije, potovanja s človeško posadko zunaj sončnega sistema ne bodo nekaj nemogočega. Morda se bomo srečali na teh potovanjih s pripadniki druge razumne civilizacije. Novi prevajalni sistemi nam bodo pomagali pri sporazumevanju. Primerjali bomo naše znanje in izkušnje z njihovimi in pričakujemo lahko, da se bomo pri tem marsikaj novega naučili. Zmenjava informacij bo potekala seveda postopoma. Civilizaciji si bosta izmenjali pomembnejše podatke šele takrat, ko si bosta zaupali. Ker pa zupanje lahko nastane le pri komunikaciji, kjer ni pomenskih nesporazumov, si bomo pomagali pri tem z računalniki.

Računalniško prevajanje je zanimivo tudi za današnji čas. V svetu, kjer ljudstva govorijo veliko različnih jezikov, si ne želimo le, da bi bila sporočila čim bolj racionalno zapisana; bolj pomembno je, da jih razumemo in da se pri prevajanju v naš jezik ne spremeni smisel informacije.

Ali nam bo množična uporaba prevajalnikov v prihodnosti prinesla tudi boljše razumevanje med narodi?