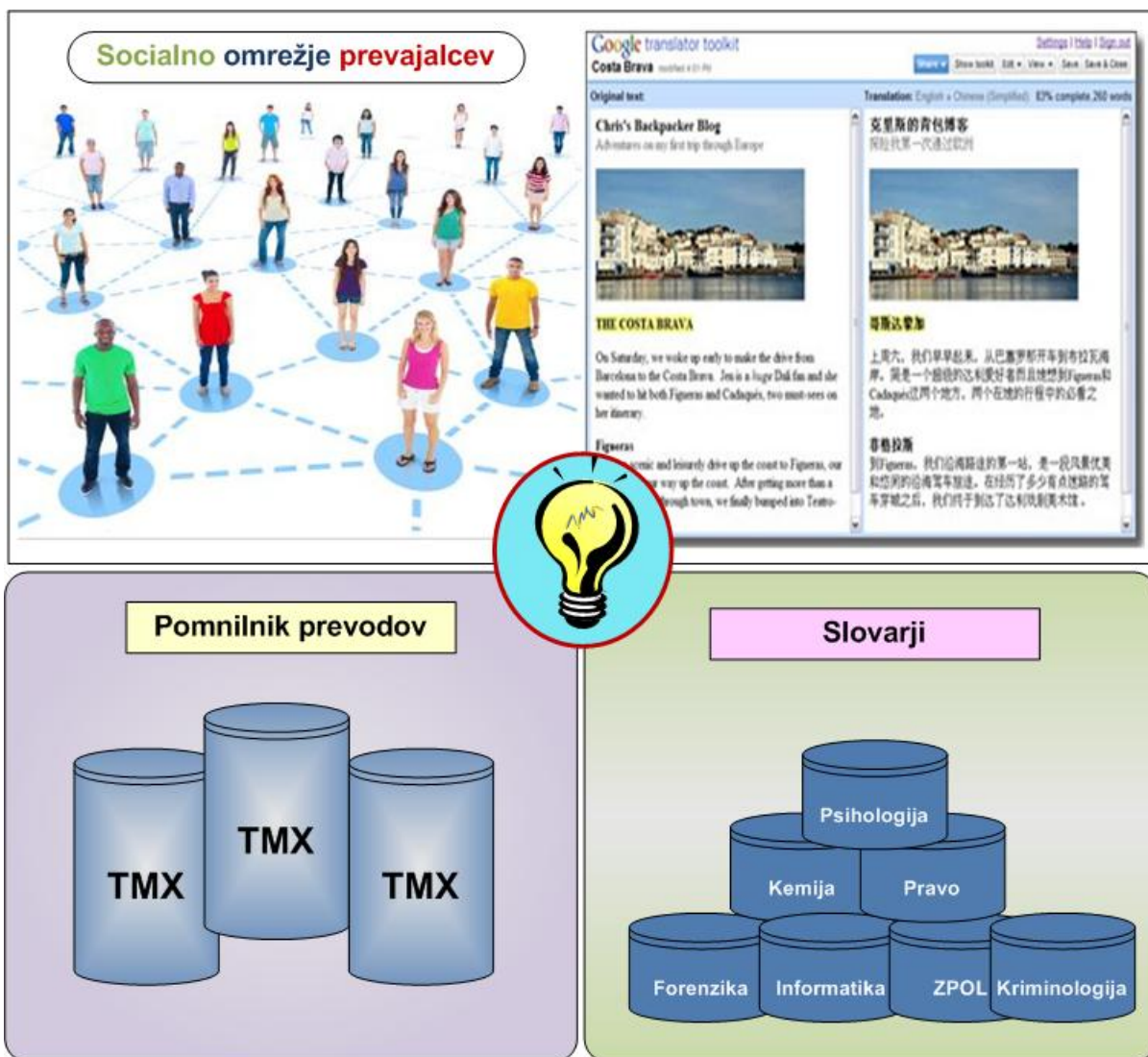


MINISTRSTVO ZA NOTRANJE

ZADEVE RS



Spletno prevajalsko orodje Google translator toolkit

Avtor dr. Karl Petrič

V Ljubljani, 2013

Kazalo vsebine
1 Uvod
1.1 Namen in cilj dela
2 Google translator toolkit
2.1 Prijava v sistem in nadzorna plošča Google translator toolkit
2.1.1 Slika 1: Prijava v sistem
2.1.2 Slika 2: Nadzorna plošča
2.1.3 Slika 3: Prevajalsko okno dokumenta
2.1.4 Slika 4: Obrazec za povabilo uporabnikov k sodelovanju
2.1.5 Slika 5: Vmesnik za nalaganje zbirk slovarjev in pomnilnikov prevodov na strežnik
3 Izdelava zbirk pomnilnikov prevodov in slovarjev
3.1.1 Preglednica 1: XML stavki in pomen označb
3.1.2 Slika 6: Del pomnilnika prevoda za menedžment in informatiko
3.2 Izgradnja slovarjev
3.2.1 Slika 7: Del spletnega slovarja s področja analize protestov
4 Zaključek
5 Viri

1 Uvod

Spletni prevajalnik [Google translator](#) poznajo mnogi uporabniki spleta. Besedilo, ki je namenjeno za prevod, se zgolj preslika v levo okence prevajalnika, nakar se določita izvorni in izhodni oziroma prevedeni jezik. Že po nekaj sekundah se v desnem okencu spletnega prevajalnika prikaže prevedeno besedilo, ki navadno za povsem osnovno razumevanje besedila že zadošča. Po drugi strani so včasih tovrstni strojni prevodi precej zavajajoči in je potrebno biti ob branju prevoda zelo pazljiv. Podobnih spletnih prevajalnikov je na spletu zelo veliko in se od Google translatorja mnogokrat bistveno ne razlikujejo. Znotraj paketa storitev [Google labs](#) (npr. Google analytics, Google search, Google history, Google docs) obstaja zanimiva storitev v obliki spletnega prevajalskega orodja, ki omogoča učinkovitejše in kakovostnejše prevajanje v številnih svetovnih jezikih. Omenjeno orodje se imenuje [Google translator toolkit](#), ki poleg strojnega prevajanja omogoča izgradnjo slovarjev, zbirk prevedenih fraz in celo socialno mreženje med različnimi uporabniki tovrstne storitve. Edini pogoj za uporabo tega prevajalskega orodja je v tem, da mora imeti uporabnik Googlov račun oziroma Gmail poštni predal, ki je brezplačen. Še zlasti Gmail poštni predal je zelo pomemben, ker lahko samo tovrstne uporabnike vabimo k sodelovanju v socialnih mrežah profesionalnih in ljubiteljskih prevajalcev.

1.1 Namen in cilj dela

Predstavitev spletnega prevajalskega orodja Google translator toolkit in njegovih gradnikov.

2 Google translator toolkit

V tem poglavju bo predstavljeno spletno prevajalsko orodje Google translator toolkit vključno z njegovimi gradniki kot so slovarji (*angl.: glossaries*), pomnilniki prevodov (*angl.: translated based memories*), statistike in socialno mreženje oziroma povezovanje (*angl.: invite people*).

2.1 Prijava v sistem in nadzorna plošča Google translator toolkit

V primeru, da še nimate Google račun se morate najprej [registrirati](#), pri čemer je potrebno izpolniti obrazec (npr. ime, priimek, e-pošta, uporabniško ime, geslo). Po uspešni registraciji se je možno [prijaviti v sistem](#), kar prikazuje naslednja slika.

Translator Toolkit

Translate your documents faster

Google Translator Toolkit is a powerful and easy-to-use editor that helps translators work faster and better.

Upload and translate documents

Use documents from your desktop or the web.

Download and publish translations

Publish translations to Wikipedia™ or Knol.

Chat and share translations online

Collaborate online with other translators.

Use advanced tools

Use features like translation memories and multilingual glossaries.

[Take a video tour »](#)



Sign in Google

Email

Password

Stay signed in

[Can't access your account?](#)

2.1.1 Slika 1: Prijava v sistem

Slika 1 prikazuje prijavo v sistem, natančneje na spletno prevajalsko orodje Google translator toolkit. Na desni strani okenca vtipkate vaš e-naslov in veljavno geslo, nakar sprožite gumb "Sign in". Po uspešni prijavi se bo prikazala nadzorna plošča.

UPLOAD		NAME	WORDS	LAST MODIFIED	SHARING
▼ Translations	<input type="checkbox"/>	OCR-Karl2 6% complete	7863	May 28	me, johnny.modri
Active	<input type="checkbox"/>	OCR-Karl2_slikaprevod 20% complete	6654	May 28	me
Hidden	<input type="checkbox"/>	Scan_Review_of_Less_lethal_Tec 100% complete	16655	May 23	me, johnny.modri
Trash	<input type="checkbox"/>	AnalizaZRD_RKK2013 100% complete	5815	May 23	me, teodor.petric, vladislav.rajkovic
▼ Requests	<input type="checkbox"/>	Analiza publiciranja s pomocjo spletne storitve COBISS 100% complete	4294	May 22	me, matjaz.mravljaja2010, teodor.petric, vladislav.rajkovic
Active	<input type="checkbox"/>	Kalkulacija_stres 100% complete	6802	May 22	me, teodor.petric, vladislav.rajkovic
Hidden	<input type="checkbox"/>	Inovativne ideje 100% complete	7970	May 22	me
Trash	<input type="checkbox"/>	White-collar_crime 100% complete	2170	May 21	me
► Labels	<input type="checkbox"/>	E-publikacije knjižnice 100% complete	360	May 21	me
► Shared with...	<input type="checkbox"/>	august riots in england analyses 100% complete	24575	May 21	me
▼ Tools					
Translation memories					
Glossaries					

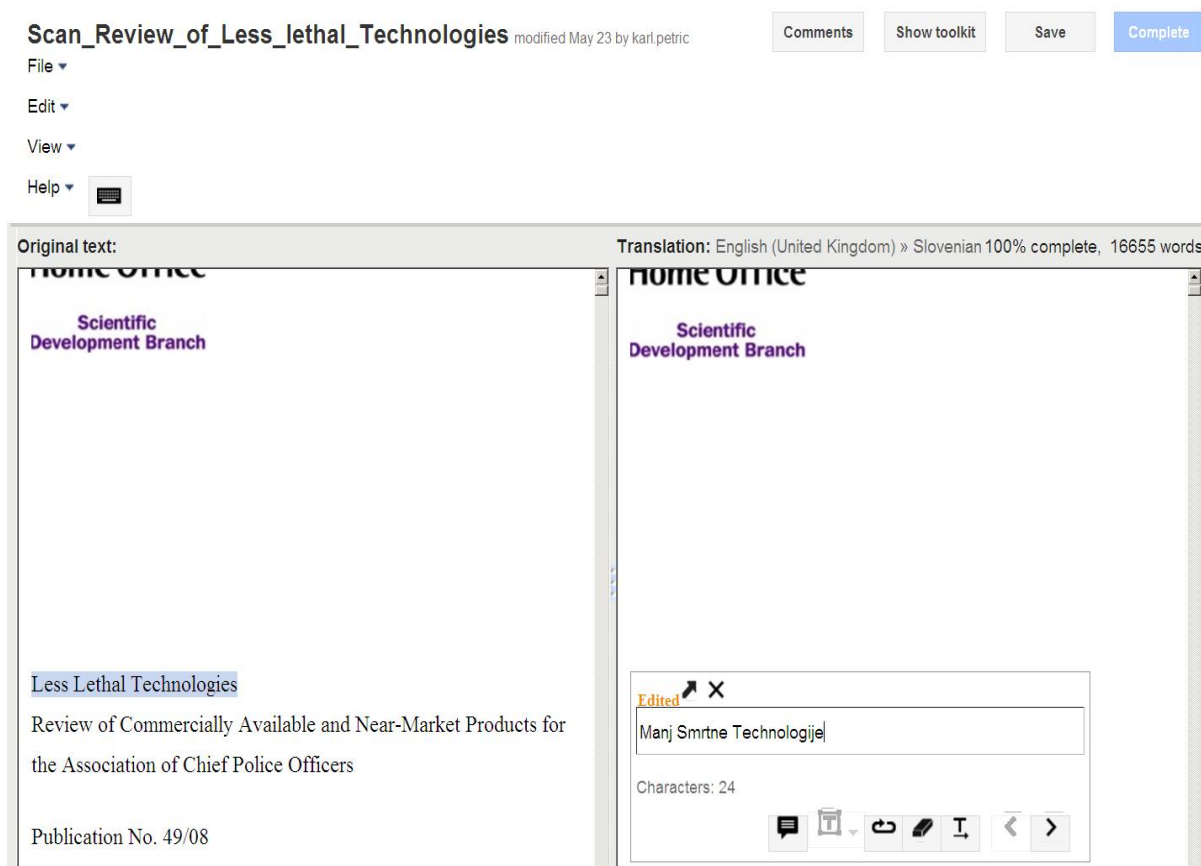
2.1.2 Slika 2: Nadzorna plošča

Slika 2 predstavlja nadzorno ploščo, ki že vsebuje prevedene dokumente, izdelane slovarje, zbirko pomnilnikov prevodov in podatke o sodelujočih uporabnikih. Zaradi večje nazornosti je smiselno po točkah opisati posamezne prvine nadzorne plošče:

a. gumb za nalaganje (*angl.: upload*) dokumentov: po sprožitvi rdečega gumba je možno naložiti naslednje datotečne oblike npr. AdWords Editor Archive (.aea), Android Resource (.xml), Apple iOS Application (.strings), Application Resource Bundle (.arb), Chrome Extension (.json), GNU gettext (.po), HTML (.html), Java Application (.properties), Microsoft Word (.doc), OpenDocument Text (.odt), Plain Text (.txt), Rich Text (.rtf), SubRip(.srt) in SubViewer (.sub). Datoteke v obliki .pdf ni možno naložiti, ampak jih je potrebno pretvoriti v .txt ali .doc. Pretvorba iz .pdf v .doc ali .txt je možna s pomočjo urejevalnika za .pdf datoteke. Podobno velja tudi za skenirane dokumente, ki jih je možno pretvoriti v besedila samo s pomočjo ustreznih programskih orodij (npr. [Abby fine reader 11](#) - komercialno orodje, ki ga lahko koristimo 30 dni brezplačno, [Free Ocr online](#) - brezplačno, [FreeOCR](#) - brezplačno programsko orodje).

b. prevodi (*angl.: translations*): ob kliku na ikonco Active se na desni strani okenca prikažejo imena prevedenih dokumentov, število uporabljenih besed, datum nastanka in imena

sodelujočih pri določenem prevajalskem projektu). V primeru, da kliknemo na dokument z imenom "Scan_Review_of_Less_lethal_Tec" se odpre naslednje okno:



2.1.3 Slika 3: Prevajalsko okno dokumenta

Slika 3 prikazuje prevajalsko okno dokumenta, ki je razdeljeno na dva dela. Prvi del okna na levi strani zaslona prikazuje izvorni dokument, ki je v tem primeru v angleškem jeziku. Drugi del na desni strani zaslona prikazuje izhodni ali prevedeni dokument v slovenskem jeziku (predhodno je bila določena prevajalska relacija angleščina (koda: en - GB) -> slovenščina (koda: sl). Uporabnik lahko desno okno urejuje, kar pomeni, da lahko spremeni besedilo v obliki popravkov, dopolnitev, zamenjav, daje pripombe, barvno označuje dele besedila idr.

c. izmenjava prevodov in sodelovanje (*angl.: shared with*): prevode si lahko izmenjujete z drugimi uporabniki in jih lahko celo povabite k sodelovanju pri prevajanju besedil in izgradnji zbirk slovarjev ter prevedenih fraz. Sicer si lahko prevode izvozite v .rtf (obogateno besedilo - format je podoben .doc) obliki na lokalni disk. Druge uporabnike, ki so lahko iz vrst različnih strok (npr. kriminalist, prevajalec, informatik, menedžer) povabite k sodelovanju tako, da označite ustrezno gradivo npr. "Scan_Review_of_Less_lethal_Tec" in ga kliknete. Odprlo se

bo novo spletno okno, ki je razdeljeno na dva dela t.j. izvorni in prevedeni dokument. Pod opcijo datoteke (*angl.: File*) osvetlite "povabi ljudi" (*angl.: invite people*), nakar kliknete. Na zaslonu se vam bo prikazal naslednji obrazec:

Share with others

Invite people | People with access

Invite: (Any Google account will work)

karl.petric@gov.si

Select access | Can edit

[Choose from contacts](#)

Subject: Prevod znanstvenega članka

Message: Za prevod znanstvenega članka. Prosim do ponedeljka.

Send a copy to myself

Send invitat | Skip sendin | Cancel

2.1.4 Slika 4: Obrazec za povabilo uporabnikov k sodelovanju

Slika 4 prikazuje obrazec za povabilo uporabnikov k sodelovanju pri prevajalskih in svetovalnih dejavnosti. Pri tem lahko glavni upravljaivec dokumentov drugim uporabnikom (z Gmail računom) določi privilegij, npr. da lahko ureja (*angl.: Can edit*). Obstajajo še drugi privilegiji, ki jih v tem delu ni smiselno dodatno izpostaviti. Na koncu je potrebno še klikniti modri gumb "pošlji vabilo" (*angl.: Send invitation*). Oseba, ki je bila povabljena k sodelovanju pri prevajanju določenega ali določenih dokumentov bo prejela v e-poštnem predalu obvestilo, ki bo poleg besedila vsebovala še spletno povezavo do spletnega prevajalskega orodja. Oseba lahko aktivira spletno povezavo, nakar se s svojim uporabniškim imenom in geslom prijavi v sistem.

d. orodja (*angl. tools*): prevajalsko orodje omogoča tudi izdelavo zbirk pomnilnikov prevodov (*angl. translated based memories*) in slovarjev (*angl.: glossaries*). Na strežnik je možno tovrstne zbirke naložiti s pomočjo vmesnika za nalaganje zbirk podatkov.

Translator Toolkit

Add a translation memory

You can create and upload translation memories for use with your translations

[« Back to Translator Toolkit](#)

Add a translation memory

Select a file: (optional)

Prebrskaj ...

Translation memory name

Sharing

If a translation memory is shared with everyone, its translations are searchable by all users. Otherwise, its translations are searchable only by you and users with whom you explicitly share access. Once you create a translation memory, you cannot change this setting. [Learn more about how Google uses your translation data.](#)

Shared with everyone Not shared with everyone

Add TM

Types of files that you can upload: (total up to 50MB)

Translation Memory eXchange (.tmx)

- A TMX file is an XML file that contains translations from one language into another language. To be accepted, TMX files must conform to the specifications for TMX 1.0 or higher.

Translator Toolkit

Upload a glossary

You can create and upload glossaries for use with your translations

[« Back to Translator Toolkit](#)

Upload a glossary

Select a file:

Prebrskaj ...

What do you want to call it ?

Upload glossary

Types of files that you can upload: (total up to 1MB)

Comma Separated Values (.csv)

- CSV file should contain translation from one language into another language. To be accepted, CSV files must conform to Translator Toolkit's [glossary format](#)

2.1.5 Slika 5: Vmesnik za nalaganje zbirk slovarjev in pomnilnikov prevodov na strežnik

Slika 5 prikazuje vmesnik za nalaganje zbirk slovarjev in prevedenih fraz na strežnik. Zbirko prevedenih fraz je potrebno pripraviti v obliki .TMX datoteke, ki je ena od številnih različic .XML formata. (vsebuje prevode iz enega v drug jezik). Omejitev pri tem je, da datoteka ne sme biti večja kot 50 MB. Za zbirko slovarjev je potrebno ustrezno konfigurirano .CSV

datoteko, kajti sicer jo bo Google translator toolkit zavrnil. Omejitev za obseg slovarjev se giblje okoli 1 MB.

Kratka predstavitev nadzorne plošče Google translator toolkit je v bistvu izpostavila najbolj pomembne prvine. Naslednje poglavje bo posvečeno izdelavi zbirke pomnilnika prevodov in slovarjev.

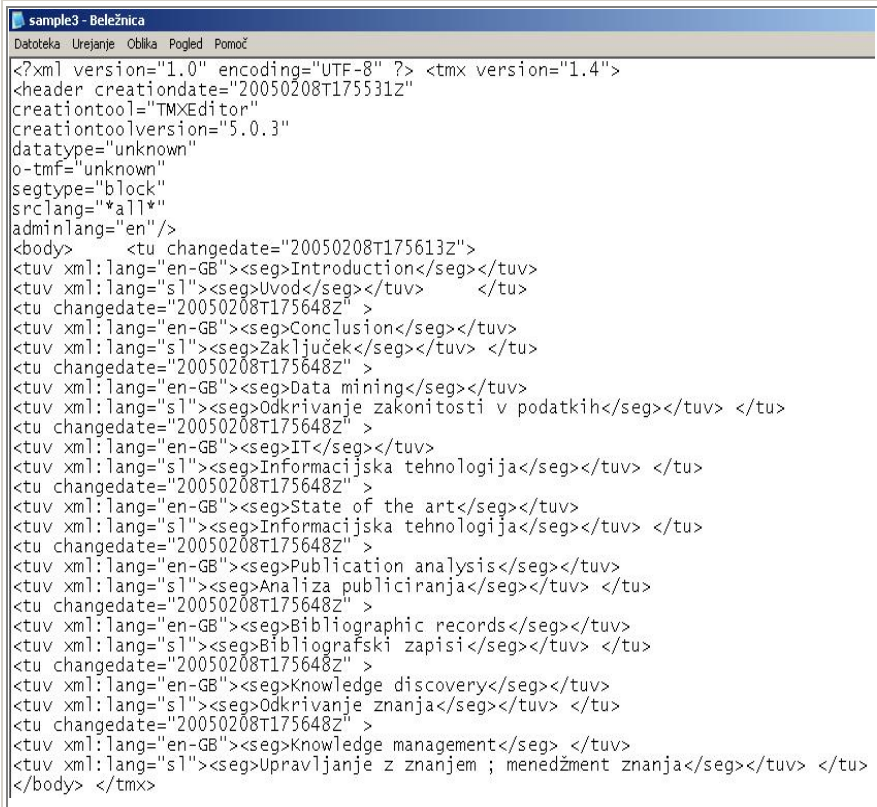
3 Izdelava zbirk pomnilnikov prevodov in slovarjev

Glavni namen izdelave zbirke slovarjev in pomnilnikov prevodov je v tem, da s tem prispevamo k dodani vrednosti oziroma kakovosti strojnih prevodov, ki so mnogokrat premalo natančni. Povrh tega je možno dokaj natančno razlikovati tako med različnimi pomenskimi različicami kot tudi med homonimi. Kot prvo se bo natančneje predstavila izgradnja zbirke prevedenih fraz, nato še slovarjev.

3.1 Izgradnja pomnilnika prevodov

Na spletu obstaja izjemno malo brezplačnih .TMX urejevalnikov, ki občutno olajšajo izgradnjo zbirke pomnilnikov prevodov. Brez tovrstnih urejevalnikov je v bistvu potrebno v XML označevalnem jeziku programirati različne relacije med izvornimi in izhodnimi pojmi, kar je lahko precej zamudno. Večina .TMX urejevalnikov je komercialnih kot npr. [Heartsome TMX editor](#), [MadCap](#), [SDL Trados](#), [Swordfish](#), [Fluency translation suite 2013](#), [CafeTran Espresso](#). Kadar je potrebno naštetih brezplačna programska orodja, ki zmorejo urejevati, uvoziti in izvoziti .TMX datoteke lahko ugotovimo, da je tovrstnih orodij izjemno malo (npr. [Google translate](#), [OmegaT](#), [TMX editor](#), [spletniTrM](#), [TmxPad](#), [Visual localize demo](#)). V danem primeru se je izdelala .TMX datoteka s pomočjo beležnice (*angl.: notepad*), znotraj katere so se pisali XML stavki.

3.1.1 Preglednica 1: XML stavki in pomen nekaterih označb

	<TMX>	Vsebuje glavo in telo dokumenta in vsebuje podatek o verziji.
	<MAP>	Podobno.
	<BODY>	Vsebuje glavne podatke.
	<TU>	Prevajalna enota.
	<TUV>	Varianta prevajalne enote.

Preglednica 1 prikazuje XML stavke pomnilnika prevoda (TMX) in pomen označb. Za namen shranjevanja različnih kosov besedila in prevoda sta zelo pomembna ukaza **<TU>** in **<TUV>**. Gre za prevajalni enoti, pri čemer je prva enota izvorna, medtem ko je druga izhodne narave. Primer: angleško besedo data mining nekateri prevedejo kot podatkovno rudarjenje, čemur v zadnjih letih dokaj strogo oporekajo. Pravi prevod po mnenju prevajalske in informacijske stroke je "odkrivanje zakonitosti v podatkih". Pomnilniku prevodov je možno dodati razne opredelitve, opombe idr., ki še osvetlijo pomen določene besedne zveze. Za posameznika izgradnja pomnilnika prevoda pomeni velik vložek, saj se le-ta lahko razvija leta in leta. Izgradnja pomnilnika prevoda s pomočjo skupin ali še boljše organiziranega socialnega strokovnega mreženja se zdi z vidika časovne porabe precej boljše rešitev. Preden se bo izvedel opis izgradnje slovarjev, bo v tem delu še prikazano, kako takšen pomnilnik prevoda z vidika zunanega uporabnika izgleda.

Translator Toolkit

Vsa prevajalska orodja > Pomnilniki prevodov > sample3

Uredi lastnosti Prenos prevodov **Iskanje prevodov**

Iskanje: » knowledge « prevedeno iz angleščina v slovenščina Iskanje

Knowledge discovery
76 % ujemanje
Odkrivanje znanja

☆☆☆☆☆  maj 28 Anon
0 Ratings No rating

Knowledge management
76 % ujemanje
Upravljanje z znanjem ; menedžment znanja

☆☆☆☆☆  maj 28 Anon
0 Ratings No rating

Knowledge menedžment
76 % ujemanje
Upravljanje z znanjem

☆☆☆☆☆  maj 28 Anon

3.1.2 Slika 6: Del pomnilnika prevoda za menedžment in informatiko

Slika 6 zgolj ponazarja možno različico prevajalnika prevoda t.j. del pomnilnika prevoda za menedžment in informatiko. Znotraj prevajalnika prevoda so lahko shranjene tudi večbesedne zveze v številnih svetovnih jezikih. Člani, ki razvijajo takšen spletni pomnilnik prevodov imajo možnost, da ovrednotijo obstoječe prevode.

3.2 Izgradnja slovarjev

Podobno kot pri pomnilnikih prevodov je možno izgraditi številne slovarje z različnih področij znanosti in stroke. Zgradbe tovrstnih slovarjev so različne, kajti lahko so enostavnejše ali pa zapletenejše. Slovarji morajo biti izdelani v obliki besedilnega .CSV in opremljene s kodiranjem UTF-8. V nasprotnem primeru ni možno naložiti slovarje na spletno prevajalsko orodje Google translator toolkit. Prva vrstica obvezno vsebuje kodiranje izvornega in izhodnega jezika in označbo besedne kategorije (*angl.: part of speech*). Opis in beležka glede določene besede sta sicer lahko zelo koristna, vendar nista obvezna.

en-GB,sl,pos,description,notes

silver commander,"**drugi poveljnik**","**person**","", ""

silver suport,"**sekundarna podpora**","**service**","", ""

bronze commander,"tretji poveljnik","person", "", ""
 gold commander,"vrhovni poveljnik","Person", "", ""
 gold support,"glavna podpora","service", "", ""

Z vidika končnega uporabnika je lahko izgled spletnega slovarja naslednji:

Translator Toolkit

Vsa prevajalska orodja > [Glosarji](#) > Glosary_Commander2

Uredi lastnosti

Prenos vnosov

Iskanje vnosov

bronze commander	person	tretji poveljnik (slovenščina)
gold commander	Person	vrhovni poveljnik (slovenščina)
gold support	service	glavna podpora (slovenščina)
silver commander	person	drugi poveljnik (slovenščina)
silver suport	service	sekundarna podpora (slovenščina)

3.2.1 Slika 7: Del spletnega slovarja s področja analize protestov

Slika 7 prikazuje del spletnega slovarja s področja analize protestov. V prvi koloni se nahajajo besede v angleškem jeziku (kodiranje en-GB: velja za Veliko Britanijo). V drugi koloni se nahajajo besedne kategorije npr. oseba (*angl.: person*), storitev (*angl.: service*). Tretja kolona vsebuje prevedene besede v slovenskem jeziku (kodiranje sl: velja za Slovenijo). Slovarje je možno posodabljeni na spletu ali pa tudi naložiti na disk.

Velik pomen slovarjev je v avtomatizaciji določenih prevajalskih dejavnosti in v nadgradnji kakovosti prevodov. Pojmi, ki se nahajajo znotraj slovarjev lahko kot upravljavec prevajalskega orodja povežemo z besedilom, ki ga želimo prevesti. Končni izid tega prenosa

podatkov se kaže v samodejnem kakovostnem prevodu določenih pojmov, s čimer je kakovost strojnega prevoda krepko presežena.

4 Zaključek

Spletno prevajalsko orodje Google translator toolkit bi lahko bilo izjemno uporabno ne samo za vzgojnoizobraževalne procese znotraj visokošolskih zavodov, ampak bi utegnilo koristiti tudi javnim uslužbencem na Ministrstvu za notranje zadeve in organov v sestavi (policija in inšpektorata). Bistveno dodano vrednost lahko vidimo v možnosti povezovanja in sodelovanja pri različnih prevajalskih dejavnostih. Pri vseh orodjih oziroma aplikacijah, ki spodbudijo socialno povezovanje in izmenjavo znanja velja preprosto načelo, da je potrebna večja množica ljudi, ki so pripravljeni prispevati vsebine. V tem vpogledu se tudi obravnava orodje ne razlikuje, kajti ni možno pričakovati veliko učinkovitost, če vsebine prispeva zgolj posameznik. Uporabno vrednost brezplačnega prevajalskega orodje lahko vidimo na naslednjih področjih delovanja MNZ in organov v sestavi:

- a. prevajalska služba (ta ima sicer profesionalna prevajalska orodja) kot dodaten pripomoček,
- b. specialna knjižnica bi lahko zgradila večjezične slovarje in tezavre s področij, ki so predmet zanimanja MNZ in organov v sestavi. Povrhu tega bi morda lahko uporabnike knjižnice, ki manj obvladajo tuje jezike oskrbovala s prevodi za osnovno razumevanje določene snovi?
- c. na višji policijski šoli bi morda lahko prevajali nekatere za policijo zanimive strani z Wikipedije?
- d. izdelava androidnega spletnega jezikovnega asistenta, ki bi lahko uporabnikom svetoval v tujih jezikih (npr. prevodi, pravilna izgovorjava, kako poizvedovati na tujih spletnih znanstvenih zbirkah podatkov?).

Kadar je govora o razvijanju koristnih, kakovostnih in učinkovitih aplikacijah/aplikacijskih sistemov mora biti tako število razvijalcev kot tudi odjemalcev storitev/izdelkov dovolj veliko, kajti sicer smo obsojeni na počasni razvoj in v nadaljevanju na nakup dragih komercialnih izdelkov in licenc.

5 Viri

[Abby fine reader 11](#) (2013-06-04) - za pretvorbo skenov v navadna besedila

[Anaphraseus](#) (2013-06-04) - orodje za menedžment TMX datotek

[Free Ocr online](#) (2013-06-04) - brezplačno, za pretvorbo skenov v navadna besedila

[FreeOCR](#) (2013-06-04) - brezplačno programsko orodje, za pretvorbo skenov v navadna besedila

[My memory](#) (2013-06-04)

[Heartsome TMX editor](#) (2013-06-04) - za pomnilnike prevodov

[MadCap](#) (2013-06-04) - za pomnilnike prevodov

[SDL Trados](#) (2013-06-04) - za pomnilnike prevodov, odlično profesionalno prevajalsko orodje

[Swordfish](#) (2013-06-04) - za pomnilnike prevodov

[Fluency translation suite 2013](#) (2013-06-04) - za pomnilnike prevodov

[CafeTran Espresso](#) (2013-06-04) - za pomnilnike prevodov

[Google translate](#) (2013-06-04) - brezplačno

[OmegaT](#) (2013-06-04) - brezplačno

[TMX editor](#) (2013-06-04) - brezplačno

[spletniTrM](#) (2013-06-04) - brezplačno spletno orodje

[TmxPad](#) (2013-06-04)

[Visual localize demo](#) (2013-06-04)

[WordFast](#) (2013-06-04)

[Seznam prevajalskih orodij](#) (2013-06-04)

[Seznam programskih in drugih orodij za prevajanje](#) (2013-06-04)

[Prevajalski projekt](#) (2013-06-04)

[Zanata](#) (2013-06-04) spletni sistem za projekte, med drugim tudi prevajalski

[TranslateWiiki](#) (2013-06-04) okolje za prevajalske skupnosti

[Prevajalsko orodje BabelFih](#) (2013-06-04)

[Katalog za prevajalska orodja](#) (2013-06-04)