

Miran Željko

Evroterm in Evrokorpus – terminološka baza in korpus prevodov

Evroterm and Evrokorpus – Terminology Database and Corpus of Translations

Povzetek

Prevajanje zakonodaje Evropske unije je zelo obsežno delo, zato pri tem sodeluje velika skupina prevajalcev. Da bi omogočili uporabo usklajene terminologije pri vseh prevajalcih in izboljšali kakovost prevodov, smo podatke, ki nastajajo pri prevajanju (terminološka zbirka in pomnilniki prevodov), prenesli na spleto (<http://www.gov.si/evroterm>). Tam so podatki prosto dostopni, na voljo pa so v dveh oblikah: kot slovar in korpus, pri čemer sta slovar in korpus medsebojno povezana.

Abstract

Translation of European Union legislation is organised by the Translation Unit of the Government Office for European Affairs (GOEA). Although there are 25 in-house translators employed at the GOEA, the majority of translations are done by translation agencies and freelance translators. The translators at the GOEA Translation Unit have created a terminology database which is continuously maintained and updated. In order to make this database available to translators outside the GOEA we have developed our own software for web access (<http://www.gov.si/evroterm/e/>). A concordance function shows the translator how a particular word is used in context.

1. Uvod

Preden bo Slovenija postala polnopravna članica Evropske unije (EU), bodo morali naši prevajalci prevesti ogromno pravnih dokumentov (Krstič, 1999), večinoma iz angleščine v slovenščino in v nasprotni smeri. Pri tem nastaja več težav: pojavljajo se novi izrazi, ki jih ni v slovarjih, splošni slovarji, ki so na voljo, so pomanjkljivi in zastareli (npr. Veliki angleško-slovenski slovar je nastajal že pred 30 leti!), v novejših pravnih slovarjih (npr. Pravni terminološki slovar (ur. Humar in Torkar), Pravni slovensko-angleški prevajalski slovar (Longyka) in Slovenski pravni leksikon (Apovnik, Primožič, Feri)) terminologija še ni povsem usklajena (Velka-

verh, 2001), terminologija posameznih strok pa je razdrobljena v množici malih terminoloških slovarjev (če ti sploh obstajajo; seznam na internetu je npr. na <http://www.gov.si/slovar.html>). Pri tako obsežnem projektu torej lahko uskladimo terminologijo le tako, da zgradimo svojo terminološko bazo.

Zakonodajo EU v slovenščino prevajajo v prevajalskem oddelku Službe Vlade RS za evropske zadeve (SVEZ – <http://www.gov.si/svez>). Tam je zaposlenih 25 prevajalcev, dve terminologinji in tri lektorice, poleg njih pa pri prevajanju pogodbeno (samostojno ali prek prevajalskih agencij) sodeluje še okoli 150 prevajalcev. Povsem razumljivo je to-

rej, da zakonodajo z istega področja prevajajo različni prevajalci – pri tem pa je zelo pomembno, da vsi uporabljajo enotno terminologijo.

Pri prevajalcih, zaposlenih v SVEZ-u, je ugodno to, da so vsi na isti lokaciji in si pri reševanju težav lahko med seboj pomagajo. Raba enotne terminologije je pri njih zagotovljena z uporabo skupne terminološke baze – v ta namen prevajalci uporabljajo program MultiTerm nemškega podjetja Trados (<http://www.trados.com>). Drugi pomemben prevajalski pripomoček so pomnilniki prevodov – v SVEZ-u uporabljajo Tradosov program Translator's WorkBench (TWB). Nakup teh programov je za SVEZ-ove prevajalce financirala služba za tehnično pomoč pri EU (TAIEX). Programi podjetja Trados so bili izbrani, ker jih uporabljajo tudi prevajalske službe v ustanovah EU.

Z zunanjimi prevajalci je več težav. Pri prevajanju zakonodaje EU je poleg pravnega izražja treba dobro poznati terminologijo ene

ali več strok – tovrstno izkušenih prevajalcev pa je malo. Ker na mnogih strokovnih področjih nimamo ustreznih slovarjev in podobnih pripomočkov, se novi prevajalci težje usposobijo za prevajanje novega področja. Poleg tega so ti prevajalci bistveno slabše povezani med sabo in so praviloma tudi slabše opremljeni z računalniškimi pripomočki. Za naše razmere so Tradosova orodja še vedno dokaj draga (blizu 150.000 SIT za osnovno verzijo programa), zato lahko nakup teh orodij ekonomsko upraviči le malo prevajalcev. (Res je za 50 evrov možno kupiti program WordFast (<http://www.champollion.net>), ki omogoča delo s Tradosovimi bazami in ga marsikateri prevajalec s pridom uporablja, ima pa tudi ta program svoje omejitve.) V zadnjih letih se je zelo povečala izobraženost prevajalcev pri uporabi interneta, zato smo že leta 2000 začeli prenašati terminologijo na internet (Krstič in Željko, 2000), leta 2002 pa smo terminologijo dopolnili še s korpusom prevodov (Željko in Krstic, 2002).

2. Zasnova rešitve

V uvodu je bila omenjena osnovna naloga, ki jo je treba izpolniti pri prevajanju zakonodaje EU: zagotoviti uporabo enotne terminologije pri vseh sodelujočih prevajalcih. Rešitev za prevajalce, zaposlene v SVEZ-u, je opisana v (Krstič, 1999) in (Krstič in Belc, 1999), v tem članku pa so opisane rešitve, ki smo jih razvili za pomoč zunanjim prevajalcem.

Prevajalci v SVEZ-u so povezani v lokalno računalniško omrežje, zaradi varnosti pa zunanji prevajalci prek interneta nimajo dostopa v to omrežje; glede na običajno hitrost dostopa tak način dela tudi ne bi bil učinkovit. Bolj smiselno je podatke preoblikovati in shraniti na spletni strežnik.

Leta 2000, ko smo se odločili za prenos terminologije na splet, je podjetje Trados omogočalo dostop do terminologije prek interneta, a je njegov program deloval le na strežnikih, temelječih na operacijskem sistemu Win-

dows NT (ta omejitev je še vedno veljala tudi v času pisanja članka). V Centru Vlade RS za informatiko (<http://www.gov.si/cvi/>) smo prvi spletni strežnik pognali leta 1994 (<http://www.matkurja.com/slo/new/1994/12/>). Takrat smo za osnovo izbrali operacijski sistem unix, dosedanje izkušnje pa so pokazale, da je bila to dobra odločitev, saj spletni strežniki na tej osnovi delujejo zanesljivo, stabilno in varno. Dobro utečenega sistema nismo želeli spreminjati samo zaradi terminološke zbirke, ki v celotni ponudbi podatkov državne uprave zavzema razmeroma majhen delež – to pa je pomenilo, da bomo morali sami razviti rešitev za prevajalce.

Naloge smo se lotili v treh korakih:

- terminologijo smo prenesli na internet,
- pomnilnike prevodov smo preoblikovali v poravnan dvojezični korpus in ga prenesli na internet,
- povezali smo terminologijo in korpus.

3. Terminologija

3.1 Terminološka baza SVEZ-a

Za urejanje in iskanje terminologije SVEZ-ovi prevajalci uporabljajo program MultiTerm. Program ima podatke shranjene v lastni bazi, če želimo te podatke sami obdelovati, jih lahko izvozimo v tekstno obliko. Ta je že prirejena za nadaljnjo obdelavo: v vsaki vrstici je med znakoma »« in »» napisano ime parametra, temu pa sledi vrednost parametra. Različni vpisi imajo lahko različne parametre, obvezna sta samo dva: zaporedna številka vpisa (to dodeli program MultiTerm ob vnosu novega podatka) in beseda ali pojem v enem jeziku (z oznako jezika), seveda pa ima vse skupaj smisel šele, ko je naveden prevod te besede vsaj v enem jeziku. Pri vsakem vpisu je možno imeti še veliko drugih podatkov; primer vpisa za enega od pojmov iz SVEZ-ove terminološke zbirke je takle:

**

<Creation Date>16.06.1997 - 11:45:44
 <Created By>super
 <Change Date>01.09.1999 - 12:46:56
 <Changed By>super
 <Entry Class>1
 <Graphic>
 <Entry Number>4
 <Subj>tehnologija
 <Project>terminološka skupina pri Medresorski komisiji za razvrščanje nevarnih snovi
 <SourceDoc&Lang>Directives 92/32/EEC, 88/379/EEC <SL>pospeševalnik
 <TermTyp>Term

<TermRef>Bela knjiga, Aneks: s. 177

<Note>(če je naprava)

<SL>pospeševalo

<Note>če je snov

<SL>pospeševalec

<Note>če je oseba

<EN>accelerator

<TermTyp>Term

<TermRef>Annex: p. 222

<Definition>Apparatus or installation emitting ionizing radiation with an energy higher than 1 Mev.

<TermRef>Amended proposal for a Council Directive laying down the basic safety standards for the protection of the health of workers and the general public against the dangers arising from ionizing radiation, COM(93) 349 final, p. 26

<DE>Beschleuniger

<TermTyp>Term

<FR>accélérateur

<TermTyp>Term

**

Iz tega zapisa vidimo, da je zasnova baze zelo fleksibilna, saj različni vpisi lahko zajemajo različne podatke (npr. možen je prevod v enega ali več jezikov, za različne jezike so možni različni komentarji), pri posameznem vpisu se nekateri podatki lahko ponovijo (npr. kadar za besedo v enem jeziku obstaja več različnih prevedkov v drugi jezik) in nekatera polja so medsebojno povezana (npr. komentarjev je lahko več, vsak pa se nanaša na polje pred sabo). Posamezna polja imajo tele pomene:

<u>Creation Date</u>	datum vpisa v originalno bazo pri SVEZ-u
<u>Created by</u>	ministrstvo ali vladna služba, kjer je bil prevod opravljen (če to ni znano ali je bilo virov več, je oznaka <i>super</i>)
<u>Entry number</u>	zaporedna številka vpisa v originalno bazo
<u>Subject</u>	področje (v angleščini – koda)
<u>Subj</u>	področje (v slovenščini)
<u>Project</u>	projekt, pri katerem so prevodi nastajali ali so se preverjali izrazi (neobvezen podatek)
<u>Source Doc&Lang</u>	naslov dokumenta, ki se je prevajal, ali vir izrazja (neobvezen podatek); iz tega podatka je mogoče sklepati, ali se je dokument prevajal iz tujega jezika v slovenščino (npr. akti ES kot direktive, uredbe, odločbe, včasih označeni s številko Celex) ali iz slovenščine v tuj jezik (slovenska zakonodaja, državni program za prevzem pravnega reda ES)

Definition	opredelitev pojma (priporočen podatek)
TermRef	vir prevedka (dokument, ministrstvo ali oseba; priporočen podatek)
Note	opombe
Reliability	zanesljivost prevoda (od 1 do 5); te podatke imajo vpisi, ki so v bazi od leta 2001. Pomen ocen: 1 – nepreverjeno, nezanesljivo; 2 – preveril samo en strokovnjak; 3 – zanesljiv izraz na posameznem področju, a ni preverjeno na različnih ravneh (npr. v drugi ustanovi morda uporabljajo drug izraz); 4 – preverjeno na različnih ravneh, a še obstajajo manjši pomisleki; 5 – preverjeno na različnih ravneh, doseženo soglasje (strokovnjaki, slovarji) – popolnoma zanesljiv izraz.

3.2 Prenos terminologije na splet

Da bi zagotovili dovolj hitro iskanje ob uporabi preproste baze, smo iz osnovne baze odstranili podatke, ki niso nujno potrebni (npr. nekateri vpisi so v vseh jezikih EU, pri čemer je prevod v italijanščino in španščino za nekatere prevajalce verjetno še zanimiv, gotovo pa nihče nima potrebe po grščini in finščini). Indeksne datoteke so narejene

za iskanje po štirih najuporabnejših jezikih: slovenščini, angleščini, francoščini in nemščini. Program za iskanje po terminološki zbirki na spletu smo poimenovali Evroterm in je dostopen na naslovu <http://www.gov.si/ev-roterm>.

Uporabniški vmesnik programa je iz treh delov (slika 1): zgoraj so stalno dostopen opis slovarja, navodila za iskanje, preklap med slovenskim in angleškim uporabniškim

Slika 1: Spletni uporabniški vmesnik Evroterm

vmesnikom ter preklap med terminologijo in korpusom; levo vpišemo iskano besedo in označimo jezik, dobimo pa po abecedni urejen seznam zadetkov, in ko kliknemo enega od zadetkov, se vsi podatki o tem pojmu izpišejo desno, v osrednjem delu zaslona. Teh podatkov je lahko veliko, zato so prevodi v štirih najpomembnejših jezikih napisani v štirih različnih barvah, da jih prevajalec hitreje najde med drugimi (praviloma manj pomembnimi) podatki. Uporabniški vmesnik je za zdaj v slovenščini in angleščini, če bi bilo potrebno, je možno hitro narediti vmesnik še v kakem drugem jeziku.

3.3 Iskanje

Pri iskanju program najde vse besede ali nize besed, ki se začnejo z vpisanim nizom znakov. Ni treba napisati celotne besede, saj je to uporabnejše pri iskanju besede z vsemi možnimi obrazil: če želite hkrati dobiti vse oblike besede »evropski« (evropska, evropske, ..., evropskega ...), v iskalno polje vpišite: **evropsk**

Uporaba velike ali male začetnice v iskalnem okencu ne vpliva na rezultate iskanja – program vedno izpiše vse, kar najde, ne glede na to, ali je podatek v bazi zapisan z velikimi ali malimi črkami.

Z uporabo regularnih izrazov lahko opravimo kompleksnejša iskanja. Če npr. iščete vse besedne zveze z »energy«, boste v izpisu dobili:

energy
energy advisory network for households
energy advisory network for large energy consumers
energy cycle
itd.

Če želite dobiti tudi besedne zveze, pri katerih se »energy« pojavlja na desni (npr. »nuclear energy«), v iskalno okence vpišite: ***energy**

Če je zadetkov preveč in želite najti le pojme, ki se končajo na izbrano besedo, dodajte na konec besede znak \$; če želite npr. dobiti vse besedne zveze, ki se konča-

jo na »energy«, napišite iskalni izraz takole: ***energy\$**

Nekatere besede se v različni delih sveta pišejo različno, npr. »organisation« in »organization«. Če želite vse take zadetke dobiti z enim iskanjem, lahko uporabite znak ».« kot nadomestni znak za katerikoli znak; iskalni pogoj bi bil v našem primeru torej takle: **organi[sz]ation**

Ta postopek deluje pri dolgih besedah, pri kratkih pa lahko dobite tudi napačne zadetke (pojavi se besede, ki jih ne želimo). Temu se izognete, če med oglatima oklepajema navedete le dovoljene znake; v prej omenjenem primeru bi torej napisali takole: **organi[sz]ation**

Razlika v zapisu nekaterih besed je še večja; npr. besede, ki se v britanski angleščini končajo na -our, se v ameriški angleščini pišejo z -or (npr. colour – color). Če želite z enim iskanjem dobiti vse zadetke tega tipa, za neobveznim znakom dodajte vprašaj, npr.: **colou?r**

Vprašaj v zapisu pomeni, da se črka »u« lahko pojavi ničkrat (torej iščemo besedo color) ali enkrat (v tem primeru iščemo besedo colour).

Druga možnost, s katero natančno določite število ponovitev, je taka, da za neobveznim znakom med zavitima oklepajema naštejete najmanjše in največje število teh znakov; za besedo colour/color bi torej napisali takole: **colou{0,1}r**

V tem zapisu smo natančno določili, da se črka »u« lahko pojavi ničkrat (iščemo besedo color) ali največ enkrat (v tem primeru iščemo besedo colour).

Posamezne iskalne pogoje lahko seveda poljubno kombiniramo. Več o regularnih izrazih najdete npr. na http://www.uga.edu/~ucns/tti/Computer_Review/Spring95/Regular_expressions.html.

Program zapisuje vse besede, ki so jih v iskalno okence vpisali uporabniki, in jih ločeno zapiše v dve datoteki – posebej najdene in posebej nenajdene izraze. Iz tega se vidi, katera področja prevajalce bolj zanimajo in katere besede bi bilo koristno v prihodnosti vključiti v slovar.

4. Korpus prevodov

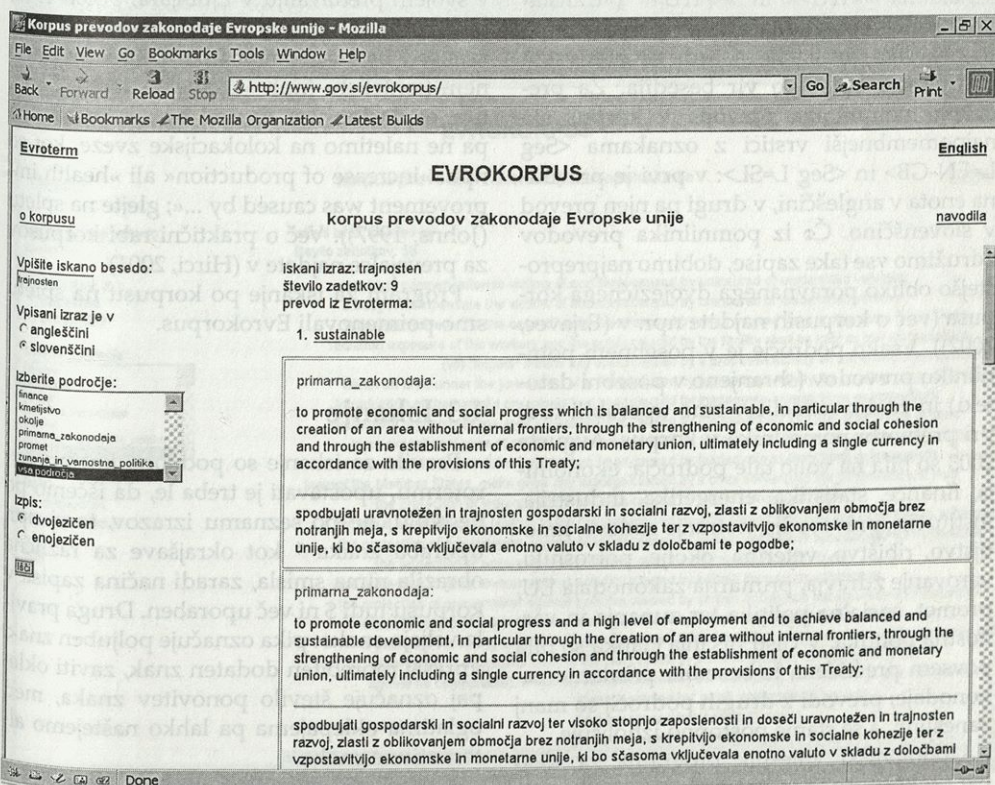
4.1 Pomnilnik prevodov

Za delo s pomnilnikom prevodov (Vintar, 1999) prevajalci v SVEZ-u uporabljajo Tradosov program Translator's WorkBench. Ob namestitvi programa se dopolnijo funkcije urejevalnika besedil MS Word in potem prevajalec večino dela opravi kar v Wordu: ko začne prevajati, dobi v okvirju prevodno enoto (navadno stavek) izvirnega besedila, v okvir pod njim pa napiše prevod. Ko konča prevajanje, se izvirnik in prevod shranita v pomnilnik. Pred prevajanjem naslednjega stavka program pregleda svojo bazo, ali je morda v njej enak ali vsaj zelo podoben izvir-

nik, in ponudi prevajalcu izvirnik in prevod. Prevajalec se potem odloči, ali bo obdržal ponujeni prevod iz baze, ga morda popravil ali začel prevajati povsem od začetka. Pomnilnik prevodov ima dvojno možnost uporabe: če se besedila vsaj delno ponavljajo, gre prevajanje hitreje (to je uporabno npr. pri spremembah veljavne zakonodaje), poleg tega pa je tudi v pomnilniku prevodov skrite veliko uporabne terminologije, ki je v terminološki zbirki morda ni.

4.2 Pretvorba pomnilnika prevodov v dvojezični korpus

Tudi bazo programa Translator's WorkBench lahko pretvorimo v besedilno obliko. Oblika zapisa je podobna kot prej:



Slika 2: Dvojezični izpis v korpusu prevodov

```

<TrU>
<CrD>
<CrU>BRANKA
<Att L=Status>Prvi prevod
<Att L=Področje>KMETIJSTVO
<Att L=Podpodročje>varstvo rastlin
<Att L=Vrsta dokumenta>Directive
<Att L=Naročnik>SVEZ-MKGP
<Txt L=Celex številka>32000L0029
<Txt L=Besedilna zbirka – TWB>en_sl_KMETIJ-
STVO
<Seg L=EN-GB>Having regard to the opinion of
the Economic and Social Committee,
<Seg L=SL>ob upoštevanju mnenja Ekonom-
sko-socialnega odbora,
</TrU>

```

Razlika glede na prejšnji zapis je ta, da ločnik med posameznimi zapisi nista več zvezdici, temveč je vsaka prevodna enota med oznakama »<TrU>« in »</TrU>« (»translation unit« – prevodna enota), parametri pa so nekoliko drugačni, a tudi tu predvsem podrobneje opisujejo vir besedila. Za pre-delavo pomnilnika prevoda v korpus sta najpomembnejši vrstici z oznakama <Seg L=EN-GB> in <Seg L=SL>: v prvi je prevodna enota v angleščini, v drugi pa njen prevod v slovenščino. Če iz pomnilnika prevodov združimo vse take zapise, dobimo najpreprostejšo obliko poravnane dvojezičnega korpusa (več o korpusih najdete npr. v (Erjavec, 2002)). Vsako področje je v posebnem pomnilniku prevodov (shranjeno v posebni datoteki) in enako razdelitev smo uporabili tudi pri pretvorbi pomnilnika v korpus. Avgusta 2003 so bila na voljo tale področja: ekonomija, finance, statistika, energetika, industrija, institucije, pravosodje, notranje zadeve, kmetijstvo, ribištvo, veterina, okolje, potrošniki, varovanje zdravja, primarna zakonodaja EU, promet, socialna politika ter zunanja in varnostna politika. V času pisanja članka so bila povsem prečiščena le besedila primarne zakonodaje; prevodi z drugih področij so manj zanesljivi, a se stanje postopno izboljšuje.

Uporabniški vmesnik programa za iskanje po korpusu je podoben kot prej (slika 2): Zgoraj je stalni del z enako navigacijo kot pri

Evrotermu, na levi uporabnik vpiše iskano besedo (ali niz besed), izbere jezik, v katerem je vpisana iskana beseda, področje iskanja in način izpisa (eno- ali dvojezičen), na desni pa se izpišejo rezultati. Če izberete dvojezični izpis, dobite celotno prevodno enoto v izvorniku in prevodu, pri enojezičnem pa se izpiše samo do 50 znakov levo in desno od iskanega pojma. Dvojezični izpis je koristen predvsem takrat, ko prevajalec ne pozna prevoda iskane-ga pojma v danem kontekstu (npr. te besede ni v slovarju) in ga zanima, kako so ta pojem doslej prevajali drugi prevajalci. Enojezični izpis (slika 3) je konkordančnik in je koristen, kadar prevajalec sicer pozna pomen pojma, a ga ob tem zanimajo druge posebnosti glede rabe, ki iz slovarja ali jezikovnih priročnikov niso razvidne – ta podatek je posebej koristen pri prevajanju v tuj jezik. Kot primer: Tim Johns, veteran v korpusnem jezikoslovju, je v svojem predavanju v Ljubljani, poleti 1999, omenil, da je ugotovil, da se v angleščini izraz »caused by« večinoma uporablja v negativnem kontekstu (npr. »accident, death, flood, fire, explosion was caused by ...«; praviloma pa ne naletimo na kolokacijske zveze, kot so npr.: »increase of production« ali »health improvement was caused by ...«; glejte na spletu (Johns, 1997)). Več o praktični rabi korpusov za prevajalce najdete v (Hirci, 2001).

Program za iskanje po korpusu na spletu smo poimenovali Evrokorpus.

4.3 Iskanje

Pravila za iskanje so podobna kot pri Evrotermu, upoštevati je treba le, da iščemo po besedilu, ne po seznamu izrazov, torej npr. uporaba znaka * kot okrajšave za različna obrazila nima smisla, zaradi načina zapisa v korpusu tudi \$ ni več uporaben. Druga pravila veljajo enako: pika označuje poljuben znak, vprašaj morebiten dodaten znak, zaviti oklepaj označuje število ponovitev znaka, med oglatima oklepajema pa lahko naštejemo alternativne znake.

Možnih je tudi nekaj novih načinov iskanja: če želite z enim iskanjem najti več besed,

jih lahko ločite z znakom »|« (na tipkovnici za PC: AltGr W); če bi npr. želeli dobiti v izpisu vse zapise, v katerih se pojavlja beseda »zadeva« ali »zadevo«, lahko to napišete kot: **zadeva | zadevo**

Poljubno besedo z regularnim izrazom napišite: **[a-z]*** (pojasnilo: beseda lahko vsebuje znake od a do z (s šumniki so lahko težave, zato jih bomo tu preskočili) in v eni besedi je lahko poljubno število znakov). Če vas zanima, kateri zadetki tipa »in case of«, »in respect of« ipd. (v splošnem: »in angl. beseda of«) obstajajo v korpusu, napišite iskalni niz kot: **in [a-z]* of**

Če vas zanimajo poljubno dolge fraze, ki imajo na začetku »in«, na koncu pa »of« (med »in« in »of« je torej lahko ena ali več besed), je treba med dovoljene znake za besedo dodati presledek, iskalni niz pa bi bil takle:

in [a-z]* of

(Razlika glede na prejšnji primer je le ta, da črki »z« sledi presledek.) V tem primeru se število zadetkov poveča na več tisoč (pazite, kaj in kako iščete, če imate počasno linijo!), ker pa sta »in« in »of« zelo pogosti besedi, je veliko zadetkov neuporabnih (v resnici ne dobimo fraz, saj računalnik ne pozna definicije »fraz«), si je pa načelo koristno zapomniti, ker nam lahko kdaj koristi.

Če želite dobiti besede, ki se končajo na določen niz znakov, napišite to npr. takole:

[a-z]*aga

S tem iskalnim pogojem dobite vse besede, ki se končajo na »aga«. (Na začetku in na koncu je treba dodati presledek, sicer dobite vse besede, ki vsebujejo niz »aga«.)

Korpus prevodov zakonodaje Evropske unije - Mozilla

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop http://www.gov.si/evrokorp/ Go Search Print

Home Bookmarks The Mozilla Organization Latest Builds

Evroterm English

EVROKORPUS

o korpusu korpus prevodov zakonodaje Evropske unije navodila

Vpišite iskano besedo: **caused by** Iskan izraz: **caused by** Število zadetkov: **58**

Vpisani izraz je v: ☐ angleščini ☐ slovenščini

Zberite področje: ☐ finance ☐ kmetstvo ☐ okolje ☐ primarna zakonodaja ☐ promet ☐ zunanja in varnostna politika ☐ vsa področja

Izpis: ☐ dvojezičen ☐ enojezičen

providing compensation to victims of accidents caused by uninsured or unidentified vehicles;
 a body to compensate the victims of accidents caused by uninsured or unidentified vehicles;
 Whereas, however, in the case of an accident caused by an uninsured vehicle, the victim is required in radiation exposure of the workers and the public caused by the facility shall be kept as low as reasonably (vii) "Impact" means any effect caused by a proposed activity on the environment including within an area under the jurisdiction of a Party caused by a proposed activity the physical origin of which is introduction of the harmful organisms has been caused by inadequate examinations or inspections. the harmful organism in the area concerned was caused by the movement into that area of one or more is an exceptional case of market disturbance caused by serious surpluses and/or problems of quality. introduction of the harmful organisms has been caused by inadequate examinations or inspections. laws of the Member States, make good any damage caused by it or its servants in the performance of their the protection of waters against pollution caused by nitrates from agricultural of forest ecosystems in relation to the damage caused by atmospheric pollution and other factors influencing the protection of groundwater against pollution caused by certain dangerous substances-☐3>chfnchfn the protection of waters against pollution caused by nitrates from agricultural the protection of groundwater against pollution caused by certain dangerous substances, the protection of groundwater against pollution caused by certain dangerous substances-☐1>chfnchfn Contents of plot file with observed damage caused by meteorologic events/phenomena of plots, the recording of observed damage caused by meteorologic phenomena should be executed on C. observation of damage of trees in the plot caused by extreme weather situations (all plots) should be comparable), and disturbances caused by the monitoring should be kept to a minimum).
 Damage caused by: game and grazing (1 = yes)
 Damage caused by: insects (1 = yes)
 Damage caused by: fungal (1 = yes)

Slika 3: Enojezični izpis v Evrokorp

Če želite poiskati besede, dolge štiri ali več znakov, ki vsebujejo samo soglasnike, napišite iskalni niz takole:

[bcčdfghjklmnprrstvz]{4,}

(Pred iskalnim nizom in za njim mora biti presledek.)

Lahko poiščemo besede določene dolžine; za besede z 20 ali več znaki je iskalni niz npr. takle:

[a-z]{20,}

Zapis pomeni, da se poljubna črka lahko ponovi 20- ali večkrat; pomembno je, da pred iskalnim nizom in za njim napišemo presledek. Če bi želeli dobiti samo besede, ki so dolge natančno 20 znakov, bi v zgornjem zapisu odstranili vejico. Če bi želeli dobiti besede, ki so dolge do 20 znakov ..., tega raje ne počenajte, sicer vam bo strežnik poslal celotno vsebino korpusa!

5. Povezava terminologije in korpusa

5.1. Podatki iz korpusa v terminologiji

V nekaterih primerih se v SVEZ-ovi terminološki zbirki pod enim vpisom pojavlja več besed v enem jeziku – npr. za angleška izraza *facility* in *sustainable* so v slovenščini navedeni kar štirje prevodi. Pri takih zapisih pogosto ne gre za sinonime, temveč je vsak izraz primeren le v določenem sobesedilu. Prevallec, za katerega je neko področje novo, pa nima občutka za izbiro pravega izraza. V slovarjih v knjižni obliki je ob večpomenskih izrazih navedenih tudi nekaj praktičnih primerov, v našem slovarju si lahko pomagamo kar s korpusom: pri podrobnem izpisu podatkov o posameznem izrazu v programu Evroterm so besede v angleščini in slovenščini podčrtane, to pomeni, da jih je možno klikniti; pri tem dobimo na zaslon rezultat iskanja te besede po celotnem korpusu, z navedbo področja, na katerem se je pojavila najdena beseda. Povezava posameznega izraza iz terminološke zbirke s korpusom se vzpostavi šele ob kliku, zato za nekatere besede ne dobimo rezultatov (ali drugače povedano: Evroterm ne

»ve«, kateri izrazi iz terminološke baze so tudi v korpusu in katerih ni).

5.2. Podatki iz terminologije v korpusu

Pri dvojezičnem izpisu iz korpusa je iskana beseda pobarvana in jo zato hitro najdemo na zaslonu, več časa pa porabimo za iskanje ustreznega mesta v drugem jeziku. Pri tem nam program pomaga tako, da uporabi podatke iz terminološke zbirke: če je iskana beseda v slovarju, je v njem tudi prevod v drug jezik in računalnik pri obdelavi besedila barvno označi izvirnik in prevod, prevajalcu pa s tem precej olajša iskanje na zaslonu. Če želimo dobiti podatke o besedi iz Evroterm, lahko kliknemo na kazalec na začetku izpisa. Težave lahko nastanejo pri sinonimih in pregibnih besedah: pri sedanjí izvedbi korpusa se upošteva le prvi pomen, zapisan v Evrotermu, ker pa v korpusu niso zapisane osnovne oblike besed, pri pregibnih besedah program ne najde besed, pri katerih se končnica razlikuje od iskanega izraza (npr. če bi iskali besedo »kri«, program ne bi našel »krvne skupine«).

6. Prednosti programa

Terminološki podatki se na spletu osvežijo hkrati z ažuriranjem podatkov v MultiTermovi bazi na SVEZ-ovem lokalnem omrežju – torej imajo zunanji prevajalci dostop do novih podatkov tako rekoč sočasno z zaposle-

nimi v SVEZ-u. Pri podatkih za korpus v času pisanja članka še ni bilo take avtomatike – to je ena od nalog, ki jih bo treba opraviti v bližnji prihodnosti. V terminološki bazi

je bilo avgusta 2003 več kot 40.000 vpisov, v korpusu pa okoli 7 milijonov besed.

Naša rešitev ima v primerjavi s Tradosovim spletnim programom predvsem tele pomembne prednosti:

- na splet smo prenesli terminologijo, pomnilnike prevodov smo preuredili v dvojezični korpus ter povezali terminologijo in korpus;
- lasten program nam omogoča, da ga postopno nadgrajujemo, pri čemer skušamo obdržati preprost in usklajen uporabniški vmesnik; hitro lahko naredimo novo jezikovno različico, če bi se pokazala taka potreba;
- podatke imamo v dveh ločenih bazah (terminologija in korpus), ki ju lahko neodvisno dopolnjujemo, dopolnitve v eni bazi pa koristijo tudi uporabnikom druge baze;
- podatki so shranjeni v obliki datotek z oznakami besedila (»tagged text files«), to pomeni, da tu predstavljena rešitev ni omejena na Tradosove baze, temveč bi kot vir podatkov lahko uporabili katerikoli drug program s podobno funkcijo, ki zna izvoziti

podatke v podobni obliki, kot je prikazano v točkah 3.1 (terminologija) in 4.2 (pomnilnik prevodov).

Za prevajalski oddelek SVEZ-a ima prenos podatkov na splet tele koristi:

- zunanji prevajalci imajo dostop do istih podatkov kot prevajalci, ki so zaposleni v SVEZ-u;
- terminologija v prevodih je bolj usklajena in prevodi zunanjih prevajalcev so boljši;
- prevode na internetu vidi več ljudi, zato hitreje najdejo napako in ta se lahko hitreje odpravi;
- zaradi boljših prevodov je delo lektorjev, strokovnih in pravnih urednikov hitrejš in lažje;
- vsi podatki so javno dostopni, in to je najboljše zagotovilo, da se bo v Sloveniji uporabljala enotna terminologija v zvezi z zakonodajo EU – pomembno je, da je terminologija na voljo tudi novinarjem, prevajalcem, ki tovrstna besedila prevajajo za druge naročnike, in študentom prevajanja.

7. Odziv prevajalcev

Ob takšnih rešitvah je seveda jasno, da so te vredne toliko, kolikor jih ljudje uporabljajo v praksi. Statistika kaže, da so spletne strani Evroterma (med vsemi stranmi državnih ustanov na strežniku www.gov.si) med tistimi, pri katerih število obiskov najhitreje raste, skupno pa je bilo v zadnjem letu več kot 900.000 poizvedb v slovarju (podrobnosti so na: <http://www.gov.si/evroterm/stat.html>); na lestvici dostopov do vseh strani na strežniku www.gov.si (podrobnosti so na: <http://www.gov.si/stat.html>) pa je julija 2003 najbolj obiskana stran Evroterma na 2. mestu (pred njo je le vstopna stran www.gov.si; vseh datotek na strežniku pa je več kot 50.000!). K

temu rezultatu je gotovo nekaj prispevalo to, da je DZS leta 2002 umaknil spletno različico Velikega angleško-slovenskega slovarja (nekateri pač ne razumejo pomena tako preproste promocije, čeprav verjetno precej vlagajo v druge oblike oglaševanja) in je zato Evroterm postal najboljšežnejši prosto dostopen spletni slovar, ki vključuje slovenščino. Ti rezultati ponovno potrjujejo velikokrat napisano in izrečeno trditev, da ljudje na spletu iščejo predvsem konkretno uporabno vsebino, take pa na spletnih straneh državnih ustanov, kljub množici podatkov in slik, žal, obupno primanjkuje.

8. Sklep

V članku je opisan prenos Tradosove terminološke baze in dvojezičnega korpusa (ki smo ga pripravili iz pomnilnika prevodov) na spletni strežnik, ki teče pod operacijskim sistemom unix. Ker razvijalec in prodajalec programov Trados ni ponudil ustrezne rešitve, smo jo razvili sami. S prenosom podatkov na splet smo omogočili prost dostop do ter-

minološke zbirke in korpusa prevodov vsej zainteresirani javnosti. Prevajalski oddelk Službe Vlade RS za evropske zadeve največje koristi pričakuje pri poenotenju terminologije, če bodo te podatke dosledno uporabljali njegovi zunanji sodelavci – prevajalci zakonodaje EU.

9. Opomba za uporabnike

Članek opisuje stanje, kakršno je bilo v začetku avgusta 2003. Ker se Evroterm in Evrokorpus stalno razvijata, so morda v času, ko smo Mostove pripravljali za tisk, nastale take

spremembe obeh programov, da navedbe v članku ne veljajo več – tedaj velja stanje, kot je opisano na spletnih straneh.

10. Literatura

- Erjavec, T. (2002). Compilation and Exploitation of the IJS-ELAN Parallel Corpus, *Zbornik konference Informacijska družba 2002 – Jezikovne tehnologije*, str. 86–93.
- Hirci, N. (2001). Nov pripomoček pri prevajalskem delu: računalniški korpusi, *Mostovi 2001*, str. 46–59.
- Johns, T. (1997). Kibbitzer 24. Cause v. lead to v. bring about. <http://web.bham.ac.uk/johnstf/revis024.htm>
- Krstič, A. (1999). Prevajanje zakonodaje Evropske unije, *Mostovi 1999*, str. 55–59.
- Krstič, A., Belc, J. (1999). Prevajanje pravnih aktov Evropske unije. *Uporabno jezikoslovje 1999* (7–8), str. 191–205.
- Krstič, A., Željko, M. (2000). Evroterm – terminologija EU na internetu, *Zbornik referatov posvetovanja INDO 2000*, str. 112–116. <http://www.gov.si/cvi/slo/indo/indo2000/referati/referat14.html>
- Velkaverh, G. (2001). Pravni slovensko-angleški prevajalski slovar. *Mostovi 2001*, str. 109–112.
- Vintar, Š. (1999). Računalniška orodja za prevajanje, *Mostovi 1999*, str. 47–54.
- Zeljko, M., Krstic, A. (2002). Web-based Trados Databases – An Alternative Approach, *Translation: New Ideas for a New Century (Proceedings of the XVI FIT World Congress)*, str. 303–308.
- Željko, M. (2000). Pomnilniki prevodov v praksi, *Mostovi 2000*, str. 75–90.