

ELRC

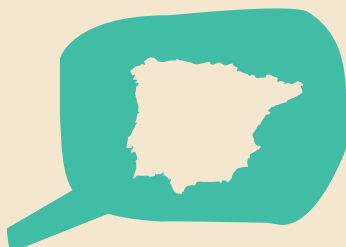
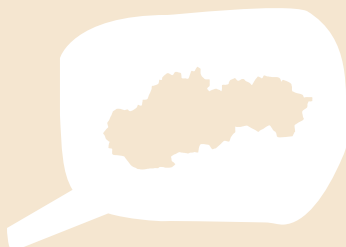
European Language Resource Coordination



Slovenský workshop



14. apríl 2016 Bratislava



CESTA K VIACJAZYČNE PREPOJENEJ EUROPE

Počet úradných jazykov v Európskej únii je v súčasnosti 24 a **jazyková rozmanitosť tak prirodzene predstavuje základný kameň EÚ**. Jazyková rôznorodosť však zároveň bráni občanom jednotlivých krajín naplno využívať online aktivity a obmedzuje ich prístup k mnohým digitálnym výhodám. Nejednen Európan často nemá k dispozícii potrebné informácie v jazyku, ktorému rozumie. Riešením, ako prekonať a odstrániť jazykové bariéry pri vytváraní jednotného digitálneho trhu, sú jazykové technológie.

Jazykové technológie pomáhajú zabezpečiť bezproblémovú komunikáciu medzi jednotlivcami, podnikmi a verejnými inštitúciami bez ohľadu na hranice a jazyky.

Umožňujú totiž malým a stredným podnikom vstúpiť na celoeurópsky trh, podporujú rast podnikania a tvorbu nových pracovných miest. Jazykové technológie sú zárukou, že spotrebiteľia nebudú diskriminovaní pri využívaní online aktivít kvôli jazyku, ktorým hovoria.

Vďaka pokroku v oblasti jazykových technológií, akým je aj **strojový preklad**, môže byť jednotný digitálny trh naozaj viacjazyčný. Komunikácia uskutočňujúca sa aj pomocou strojového prekladu spája občanov, verejné inštitúcie a mimovládne organizácie rôznych národností, podporuje zvýšenú interakciu, angažovanosť a vzájomné porozumenie. Výsledkom nie je len jednotný digitálny trh, ale aj integrovanejšia **viacjazyčná Európa**.

CEF

Nástroj na prepájanie Európy
(NPE)

Connecting Europe Facility

CEF (NPE) financuje projekty, ktorých cieľom je zaplniť biele miesta na mape európskej energetiky, dopravy a digitálnych služieb. Jeho zámerom je aj budovanie ekologickejšieho európskeho hospodárstva podporovaním čistejšieho spôsobu dopravy, vysokorýchlostného širokopásmového pripojenia a ľahšieho prístupu k využívaniu obnoviteľných zdrojov energie v súlade so stratégiou **Európa 2020**.

Financovanie energetických sietí bude mať za následok ďalšie zjednocovanie vnútorného energetického trhu, znižovanie energetickej závislosti EÚ a posilňovanie bezpečnosti dodávok. CEF, ktorý sa zameriava na inteligentné, udržateľné a plne prepojené dopravné, energetické a digitálne siete, pomôže dotvoriť jednotný európsky trh.

CEF



Strojový preklad NPE (NPE.SP)
**Connecting Europe Facility.
Automated Translation**

CEF.AT (NPE.SP) je bezpečná a flexibilná platforma strojového prekladu, ktorá sa dokáže pripojiť na akúkoľvek **infraštruktúru digitálnych služieb**, čím sa táto platforma stáva viacjazyčnou.

CEF.AT nadväzuje na systém strojového prekladu Európskej komisie **MT@EC**, ktorý je v prevádzke od roku 2013 a je k dispozícii inštitúciám EÚ, vládam členských štátov, mnohým informačným systémom EÚ a internetovým službám. CEF.AT chce v budúcnosti ponúkať nástroje a služby, vďaka ktorým by verejné služby jednotlivých štátov fungovali viacjazyčne.

ELRC

Koordinácia európskych jazykových zdrojov (KEJZ) European Language Resource Coordination

ELRC (KEJZ) je iniciatíva, prostredníctvom ktorej plánuje Európska komisia vytvoriť jedinečnú databázu jazykových dát. Ide o prvý krok s cieľom prispôsobiť platformu strojového prekladu CEF.AT potrebám verejných služieb vo všetkých členských štátoch EÚ, na Islande a v Nórsku.

ELRC tak prispeje k prekonávaniu priepasti medzi schopnosťami dnešných systémov strojového prekladu, ktoré Európska komisia ponúka jednotlivým európskym verejným správam, a medzi skutočnými, každodennými požiadavkami týchto správ.



ELR



ELRC

Platforma CEF.AT sa bude opierať o štatistický model strojového prekladu, ktorý sa trénuje na existujúcich prekladoch vytvorených človekom. Na tvorbu prekladových softvérov, ktoré by mohli poskytovať použiteľné texty orgánom verejnej správy v jednotlivých krajinách EÚ a používateľom verejných online služieb, sú potrebné **jazykové zdroje**.

Iniciatíva ELRC hľadá **jednak voľne dostupné dáta** t. j. údaje, ktoré je možné voľne opakovane použiť a šíriť prostredníctvom príslušných portálov, jednak komerčne dostupné súbory dát.

Na zlepšenie povedomia o dôležitosti databáz a jazykových zdrojov, ktoré sú potrebné pre strojový preklad, sa v každej z 30 zúčastnených krajín Európy konajú informačné **semináre**.

ELRC zriadila **technickú a právnu asistenčnú službu**, ktorej úlohou je odpovedať na otázky týkajúce sa technických a právnych aspektov používania, tvorby, získavania alebo zdieľania jazykových zdrojov. Asistenčná služba je bezplatná a odpoveď na vaše otázky odosieľa v priebehu jedného pracovného dňa.



É

ČO SÚ JAZYKOVE ZDROJE A NAČO SÚ POTREBNÉ?

Zlepšenie kvality služieb, ktoré ponúka platforma strojového prekladu CEF.AT, si vyžaduje, aby sa softvér trénoval na **relevantných jazykových dátach vo všetkých úradných jazykoch** krajín zapojených do programu CEF.

Štátne a vládne inštitúcie, ako aj neziskové a súkromné organizácie v členských krajinách Európskej únie a v ďalších krajinách podieľajúcich sa na platforme CEF.AT každý deň vyprodujú značné množstvo jazykových dát, ktoré predstavujú cenné jazykové zdroje pre platformu CEF.AT.

Medzi databázy a jazykové zdroje, ktoré už v súčasnosti slúžia na zlepšovanie strojového prekladu, patria **paralelné korpusy**, rozsiahle jednojazyčné korpusy (napríklad jednotlivé národné korpusy) a **jazykové zdroje zamerané na špecifické oblasti** vrátane terminologických databáz a taxonómií z oblasti verejnej správy a verejných online služieb. Verejné online služby, ktoré chce podporiť platforma CEF.AT, sa týkajú práv spotrebiteľov, kultúry, práva, sociálneho zabezpečenia, zdravotníctva, verejného obstarávania atď.



É

AKÉ JAZYKOVE ZDROJE SÚ VHDNÉ PRE PLATFORMU CEF.AT?

Ako jazykové zdroje môžu poslúžiť rôzne typy dokumentov: správy, súpisy termínov, letáky, sprievodcovia atď. Niektoré súbory dát pochádzajúce z verejných inštitúcií sa dajú v strojovom preklade využiť priamo: zarovnané korpussy z prekladových pamätí, terminologické zdroje, glosáre a slovníky.

Mnohé ďalšie zdroje sú prístupné ako informačné dokumenty: správy, príručky, záznamy o administratívnych záležitostiach atď. Dôležité je, aby sprístupnené textové dokumenty boli **editovateľné**. Skenovaný dokument vo formáte PDF by sa nedal použiť, ak by bol vytvorený jednoduchým rozpoznávaním znakov (OCR).

STROJOVÝ
PREKLADPREKLADOVÉ/
JAZYKOVÉ
MODEL

GLOSÁRE

KORPUS

TERMINOLOGICKÉ
ZDROJE

GRAMATIKY

PREKLADOVÉ
PAMÄTEAKÉ DRUHY DÁT SU
POTREBNÉ
PRI TRÉNOVANÍ
STROJOVÉHO
PREKLADU?

Pri trénovaní strojového prekladu sa používajú rozličné druhy dát:

prekladové pamäte: databázy, ktoré obsahujú preklady vytvorené človekom, sa môžu využiť pri budúcich prekladoch, ale aj pri trénovaní systémov strojového prekladu;

korpusy: jednojazyčné a viacjazyčné korpusy – porovnateľné, zarovnané a paralelné dokumenty;

terminologické zdroje: štruktúrované súbory odborných pomenovaní z rôznych oblastí obsahujúce relevantné jazykové informácie;

prekladové/jazykové modely: štatistické informácie, ktoré určujú pravdepodobnosť časti nepoznaného/nevideného textu a ktoré sú založené na trénovaných dátach;

glosáre: jednojazyčné a viacjazyčné zoznamy slov, viacslovných pomenovaní, viet a i., všeobecné alebo zamerané na konkrétnu oblasť;

gramatiky: súbory pravidiel zaoberajúcich sa stavbou jazyka.



KDE HĽADAME JAZYKOVÉ ZDROJE?

Medzi využiteľné jazykové zdroje patria texty verejnej správy v členských štátoch EÚ a v krajinách, ktoré sa zapojili do programu CEF.

Takéto texty pripravujú inštitúcie verejnej správy, a to na štátnej, regionálnej, miestnej, pohraničnej alebo medzinárodnej úrovni, ako aj medzinárodné organizácie na európskej úrovni vrátane prezidentských kancelárií, ministerstiev, parlamentov, regionálnych vlád, miestnych úradov a pod.

Pracujú s nimi aj orgány verejnej správy zodpovedajúce za elektronické služby (napr. práva spotrebiteľov, kultúra, právna oblasť, sociálna starostlivosť, zdravotníctvo, verejné obstarávanie), tlačové centrá ministerstiev, dokumentačné centrá, národné knižnice a pod.



É

TEXTOVÉ A JAZYKOVÉ ZDROJE SLOVENSKEHO NÁRODNÉHO KORPUSU (SNK)

Oddelenie SNK JÚLŠ SAV sa sústreďuje na získavanie a počítačové spracúvanie textov, ich dopĺňanie relevantnými jazykovými informáciami a na tvorbu rôznych druhov korpusov a databáz. Okrem hlavného korpusu slovenských písaných textov budujú pracovníci SNK hovorený korpus, korpus nárečí, historický korpus, morfológickú databázu ako slovník všetkých tvarov slov, terminologickú databázu ako súpis termínov z rôznych oblastí a i. Súčasťou SNK sú aj špecializované korpusy vytvorené z textov

konkrétnych vecných oblastí, príp. podľa osobitých kritérií výskumného projektu, napr. korpusy textov z oblasti práva, ekonómie, náboženstva. V prekladateľskej praxi sa využívajú paralelné korpusy, ktoré obsahujú rovnaké texty v dvoch jazykoch spárované na úrovni viet. V rámci SNK je aktuálne prístupný slovensko-anglický, slovensko-bulharský, slovensko-český, slovensko-francúzsky, slovensko-latinský, slovensko-maďarský, slovensko-nemecký a slovensko-ruský paralelný korpus.

PROGRAM ELRC WORKSHOP

8.00 – 9.00

Registrácia

9.00 – 9.15

Otvorenie

Dušan Chrenek (Zastúpenie Európskej komisie v SR)
Zuzana Komárová (Ministerstvo kultúry SR)
Mária Šimková (Jazykovedný ústav Ľ. Štúra SAV)
Stelios Piperidis (ELRC)
Miroslav Zumrík (Jazykovedný ústav Ľ. Štúra SAV)

9.15 – 9.30

Ciele projektu ELRC
Stelios Piperidis (ELRC)

9.30 – 9.45

Európa a viacjazyčnosť

Nataša Procházková (Zástupkyňa DGT pre SR)

9.45 – 10.15

Panelová diskusia 1

Preklad v štátnej správe na Slovensku

Moderátor: **Nataša Procházková** (DGT SR)

Členovia diskusie:

Marketa Štefková (Filozofická fakulta UK),
Michal Kmeť (Asociácia prekladateľských spoločností Slovenska),
Zuzana Mrvová (Oddelenie medzinárodných vzťahov Národnej banky Slovenska),
Barbora Malliarová (Odbor znaleckej, tlmočnickej a prekladateľskej činnosti Ministerstva spravodlivosti SR)

10.15 – 10.45

Kávoová prestávka

10.45 – 11.15

Strojový preklad – ako funguje?

Ondřej Bojar (Ústav formální a aplikované lingvistiky MFF UK)

11.15 – 11.45

Aké dáta sa používajú v strojovom preklade?

Ondřej Bojar (ÚFAL MFF UK)

DOM EURÓPY BRATISLAVA 14. 4. 2016

11.45 – 12.15

Jazyky a jazykové technológie na Slovensku

Jozef Juhár (Katedra elektroniky a multimediálnych komunikácií TUKE)

12.15 – 13.15

Obedová prestávka

13.15 – 13.45

CEF.AT – Ako môžu verejné inštitúcie využiť túto platformu

Daniel Klivanec (Zástupca DGT PRE EK)

13.45 – 14.15

Panelová diskusia 2

Jazykové zdroje na Slovensku

Moderátor: **Mária Šimková** (Jazykovedný ústav Ľ. Štúra SAV)

Členovia diskusie:

Ladislav Kluchý (Ústav informatiky SAV),
Jozef Juhár (KEMK TUKE),
Daniel Klivanec (DGT EK)

14.15 – 14.45

Právny rámec poskytovania dát

Róbert Dobrovodský (Ministerstvo spravodlivosti SR)
Magdaléna Miklošová (Ministerstvo kultúry SR)

14.45 – 15.15

Kávoová prestávka

15.15 – 15.45

Dáta a jazykové zdroje: praktické aspekty a „best practice“

Stelios Piperidis (ELRC)

15.45 – 16.15

Ako sa možno zapojiť (diskusia a závery)

Jana Levická (Jazykovedný ústav Ľ. Štúra SAV),
Stelios Piperidis (ELRC)
Diana Šnircová (Ministerstvo kultúry SR)

EU Commission to Pay for Translation When Online Shopping Goes Wrong

by Florian Faes on February 26, 2016, www.slator.com

The European Union is moving forward in eliminating some of the barriers holding back its digital single market. On February 15, the European Commission launched its Alternative / Online Dispute Resolution (ADR / ODR) platform.

In short, the new platform aims to make “consumers and traders more confident in trading online and across borders”. It also wants to avoid protracted legal battles with the Commission’s website saying that “if consumers have a complaint about a good or service they have bought, instead of going to court, they can choose Alternative Dispute Resolution (ADR).”

The procedure comes in four steps: 1) Consumer fills in form; 2) Complaint is sent to trader, who proposes an ADR entity to the consumer; 3) Platform transfers automatically the complaint to that entity; 4) ADR entity handles the case entirely online and reaches an outcome in 90 days. The parties need to agree which one of the more than 100 dispute resolution bodies they are going to use.

In the EU with its more than 20 languages, cross-border dispute resolution will by default involve lots of translation. In fact, as an April 2015 conference on the topic pointed out, “one of the last remaining obstacles to the Digital Single Market is the language barrier.”

The commission promises that “the platform is multilingual. A translation service is available on the platform to assist disputes involving parties based in different European countries.”

Pôvodný text



Európska komisia zaplatiť za preklad pri nakupovaní na internete

ktorú predkladá Florian FAES v februára 26, 2016, www.slator.com

Takto dnes MT@EC Dokáže Prekladať z angličtiny

S pomocou vašich zdrojov zvládne MT@EC prekladať lepšie

Európska komisia zaplatiť za preklad pri nakupovaní na internete

ktorú predkladá Florian FAES v februára 26, 2016, www.slator.com

Európska únia napreduje odstránením niektorých prekážok, ktoré brzdia jej jednotného digitálneho trhu. V februára 15 Európska komisia spustila alternatívne a online riešenie sporov (ARS/RSO) platformy.

V skratke, nová platforma je „dôvery spotrebiteľov a obchodníkov v cezhraničné obchodovanie online“. Zároveň chce vyhnúť závažných súdnych sporov na webovej stránke Komisie, v ktorom sa uvádza, že „ak si spotrebiteľia nechajú sťažnosť na tovar alebo službu kúpili, namiesto súdneho konania, majú možnosť vybrať si alternatívneho riešenia sporov (ARS).“

Postup spadá do štyroch krokov: 1) spotrebiteľ vyplní formulár; 2) sťažnosť odoslaná obchodníkovi, ktorý navrhuje subjekt ARS pre spotrebiteľov; 3) presunov platforma automaticky sťažnosť uvedenému subjektu; 4) Ak subjekt ARS sa zaoberá výlučne online a dosiahne výsledok v roku 90 dní. Je potrebné, aby sa strany dohodli, ktoré viac ako 100 subjektov riešenia sporov, ktoré budú používať.

EÚ so svojimi viac ako 20 jazykov, cezhraničné riešenie sporov bude automaticky znamenať veľa prekladu. Keďže v apríla 2015 konferenciu na túto tému zdôraznil, „jedným z posledných zostávajúcich prekážok jednotného digitálneho trhu je jazyková bariéra“.

Komisia sľubuje, že „platforma je viacjazyčná. prekladateľské služby je k dispozícii na platforme pri sporoch medzi stranami so sídlom v rôznych európskych krajinách.“



THE EUROPEAN LANGUAGE RESOURCE COORDINATION



KONZORCIUM ELRC



ORGANIZÁTORI WORKSHOPU



Viac informácií
o Slovenskom národnom
korpuse a iniciatíve ELRC
môžete nájsť na:





<http://korpus.sk>
info@lr-coordination.eu
www.lr-coordination.eu