

Štatistický preklad textu

Dokumentácia k inžinierskemu dielu

Tím č. 4 – aDictIT

Bc. Róbert Horváth

Bc. Peter Jurčík

Bc. Peter Macko

Bc. Vladimír Ruman

Bc. Peter Sládeček

Bc. Maroš Ubreži

Bc. Matúš Vacula

Obsah

1	Úvod.....	4
2	Účel dokumentu	4
2.1	Forma a rozsah dokumentu.....	4
2.2	Použité technológie.....	4
2.3	Slovník pojmov.....	4
2.4	Zoznam skratiek	4
3	Prvý šprint – káva	6
3.1	Návrh používateľského rozhrania.....	6
3.1.1	Analýza existujúcich rozhraní.....	6
3.1.2	Návrh.....	8
3.1.3	Implementácia.....	9
3.1.4	Testovanie	10
3.2	Preklad slova.....	11
3.2.1	Analýza existujúcich slovníkov	11
3.2.2	Návrh slovníka	12
3.3	Detailný popis tabuliek.....	12
3.3.1	Implementácia.....	13
4	Druhý šprint – pivo	15
4.1	Preklad vety	15
4.1.1	Analýza	15
4.1.2	Návrh.....	21
4.1.3	Implementácia.....	26
4.1.4	Testovanie	28
5	Použitá literatúra	30
5.1	Príloha A – Preklad viet.....	I
A.1.	Vysvetlenie označenia v prílohách	II
A.2.	Z angličtiny do slovenčiny a češtiny.....	II
A.3.	Zo slovenčiny a češtiny do angličtiny.....	III
5.2	Príloha B – Preklad paragrafov	I
B.1.	Vysvetlenie označenia v prílohách	II

B.2. Z angličtiny do slovenčiny a češtiny.....II
B.3. Zo slovenčiny a češtiny do angličtiny..... IV

1 Úvod

2 Účel dokumentu

Obsahom predkladaného dokumentu je dokumentácia k vývoju softvérového systému. Úlohou vyvíjaného systému je prekladať používateľom zadaný text z jedného jazyka do iného jazyka a späť. Text sa prekladá po vetách alebo skupinách slov. Spôsob prekladania je založený na štatistickej podobnosti výskytu prekladanej časti v korpuse. Tento softvérový systém je výsledkom študentského tímového projektu z predmetu Tímový projekt. Vedúcim a zároveň zadávateľom témy je Ing. Dušan Zeleník.

2.1 Forma a rozsah dokumentu

Projekt je vyvíjaný pomocou agilnej metodiky SCRUM, ktorá má viaceré špecifiká. Vývoj v nej prebieha v diskrétno oddelených častiach nazývaných šprinty, ktoré trvajú spravidla dva týždne. Táto skutočnosť sa odráža aj na dokumentácii, ktorej obsahom je podrobný opis vyvinutej práce v prvých dvoch šprintoch. Každý šprint pozostáva z používateľských príbehov (User story), ktoré majú jasne formulované rozdelenie na analýzu, návrh, implementáciu a testovanie. V rámci príbehov sú identifikované úlohy, ktoré v dokumente štruktúrujú jednotlivé časti príbehov.

Na konci dokumentu sa vo forme príloh nachádzajú výstupy analýzy prekladu viet a odstavcov.

2.2 Použité technológie

Pre prácu na projekte sme sa rozhodli pre využitie nasledujúcich technológií:

- Java
- MySQL
- PHP
- HTML5
- Tomcat
- Apache
- Systém pre verziovanie: Git
- Systém pre správu úloh: Redmine
- IDE: Eclipse, NetBeans

2.3 Slovník pojmov

Korpus – súhrn všetkých skúmaných textov, komunikátov daného jazyka

Paralelný korpus – korpus obsahujúci totožné texty v rôznych jazykoch

Webová služba – je to služba prístupná na webe podľa zadanej požiadavky, prostredníctvom XML súboru vracia odpoveď s výstupom v XML formáte

2.4 Zoznam skratiek

WS – webová služba

SQL - Structured Query Language, je počítačový jazyk na manipuláciu (DML) (výber, vkladanie, úpravu a mazanie) a definíciu dát (DDL).

3 Prvý šprint – káva

V súvislosti s názvom nášho tímu sme sa jednotlivé šprinty rozhodli pomenovávať návykovými látkami. V prvom šprinte je to káva, ako jedna z najneškodnejších návykových látok. V tomto šprinte sme riešili nasledujúce úlohy:

1. Analýza existujúcich prekladačov
2. Navrhnutie používateľského prostredia
3. Výber slovníka pre preklad
4. Vytvorenie webovej služby

Výsledkom prvého šprintu je teda funkčné prostredie pre prácu s prekladačom, takisto slovník, ktorý obsahuje dostatočné množstvo slov pre preklad. V konečnej fázy sme spustili službu a prepojili ju s grafickým rozhraním tak, aby si náš zákazník dokázal prekladať slová.

3.1 Návrh používateľského rozhrania

Ako používateľ chcem *pracovať s používateľským rozhraním*, pomocou ktorého budem *prekladať text*.

3.1.1 Analýza existujúcich rozhraní

V kapitole sa venujeme rozboru rozhraní populárnych webových aplikácií a to konkrétne Google Translate a Yahoo! Babel Fish.

Google Translate

Google Translate poskytuje používateľom prehľadné a elegantné rozhranie, ktorého použitie je veľmi intuitívne. Používateľ má možnosť nastavenia zdrojového a cieľového jazyka prekladu pomocou dvoch tlačidiel, ktoré sa kliknutím rozvinú do zoznamov ponúkajúcich kompletnú jazykovú ponuku.

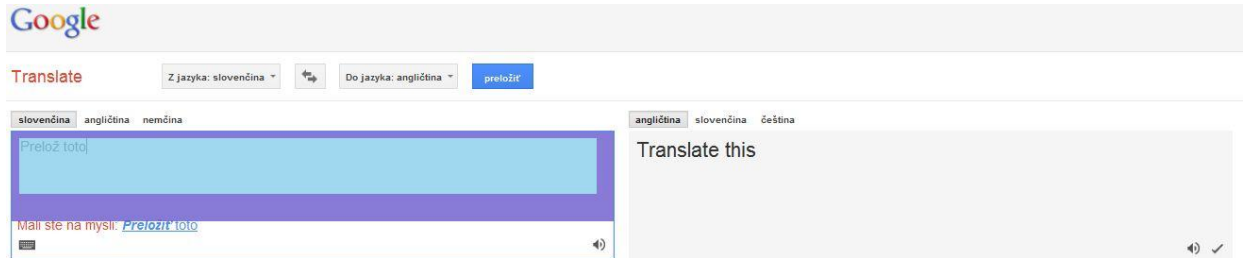
Do výrazného textového poľa v ľavej časti obrazovky máme možnosť vpisovať text, ktorý si želáme prekladať. Preklad je ponúknutý v podobnom elemente na opačnej strane obrazovky, avšak bez možnosti editácie používateľom (



Obr. 1).

Celé rozhranie sa výborne prispôsobuje používateľom nadstavenému rozlíšeniu a veľkosti jeho obrazovky. Základnou myšlienkou je vložiť textové pole pre vpisovanie do elementu, ktorého CSS atribút width (šírka) je zadaný pomerovo v percentách. Následne pomocou JavaScriptu je získaná dĺžka tohto elementu. Od tejto hodnoty je odčítaná určitá hodnota. Následne je takto mo-

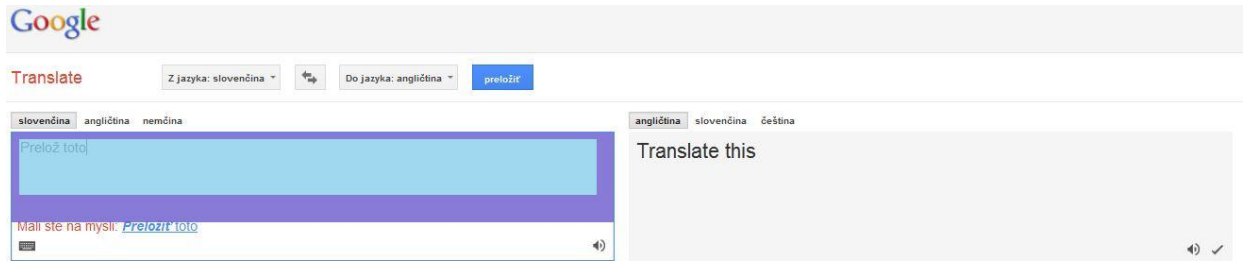
definovaná hodnota priradená CSS elementu width (šírka) textového poľa. Tým sa zabezpečí efekt rovnakého odstupu písaného textu od okraja elementu, bez ohľadu na parametre používateľovej obrazovky. Rovnako sa týmto predchádza aj odlišnostiam pri interpretácii vlastného box modelu jednotlivými webovými prehliadačmi. Tento postup je použitý aj pri nastavení okna pre zobrazovanie výsledkov prekladaného textu.



Obr. 1 – Rozhranie Google Translate. Bledomodrá časť ukazuje textové pole vložené do elementu s pomerovo nastavenou šírkou.

Yahoo! Babel Fish

Zanalyzujme si ešte webové rozhranie prekladača od spoločnosti Yahoo!. Pri prvotnom zobrazení stránky sa používateľovi zobrazuje iba textové pole, do ktorého vkladá text, ktorý chce prekladať. Následný výsledok z procesu prekladania sa zobrazuje v okne, ktoré sa zobrazí nad oknom textového poľa so zadávaným textom (Obr. 2)



Obr. 1.

Používateľ má možnosť nastavenia jazykov prekladu z natívneho HTML zoznamu možností, umiestneného priamo pod textovým poľom so zadávaným textom.

Na rozdiel od Google Translate nie je rozhranie prekladača od spoločnosti Yahoo! natoľko modifikovateľné. Textové pole pre vpisovanie prekladaného textu je vložené v kontajneri s pevnou šírkou. Optimalizácia pre rozličné webové prehliadače je samozrejmosť.



Obr. 2 – Rozhranie prekladača Yahoo! Babel Fish od spoločnosti Yahoo!

Záver

Prekladač od Yahoo! ponúka svojim používateľom klasické rozhranie s využitím dostupných HTML elementov v kombinácii s CSS. Google Translate pri budovaní rozhrania navyše využíva aj možnosti technológie na strane klienta a tým ponúka svojim užívateľom veľmi elegantné, moderné a na používanie intuitívne rozhranie. Rovnako tak aj celá aplikácia od Googlu pôsobí veľmi interaktívne a poskytuje používateľovi väčší komfort pri práci.

Výber implementačného prostredia

Prvou časťou nášho implementačného prostredia sa stal operačný systém. V tejto oblasti sme sa priklonili k operačnému systému Linux Fedora. Hlavným dôvodom bola jednoduchá práca, možnosť jednoduchého inštalovania ďalších programov a pravidelné aktualizácie systému.

V oblasti programovacieho jazyka sme zvolili programovací jazyk Java. Tento jazyk nám ponúka hlavne možnosti jednoduchého programovania a možnosti práce s webovými technológiami. Okrem toho budeme využívať aj programovací jazyk PHP pre tvorbu webového rozhrania. Pri práci s webovým rozhraním okrem toho využijeme aj JavaScript a framework JQuery.

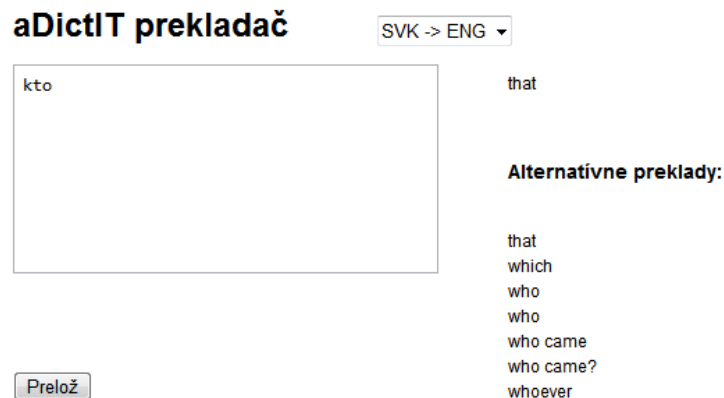
Na ukladanie dát využijeme databázový systém MySQL, ktorá nám zabezpečí rýchly a jednoduchý prístup k dátam. Oproti ostatným systémom pre nás ponúka najlepšie možnosti rozšírení.

3.1.2 Návrh

Návrh grafického rozhrania

Na základe vykonanej analýzy existujúcich rozhraní sme sa rozhodli navrhnúť elegantné a pre používateľa intuitívne rozhranie webovej aplikácie. Rozmery jednotlivých elementov by sa mali prispôbiť veľkosti používateľovej obrazovky ako aj nastavenému rozlíšeniu.

Zrealizované rozhranie je viditeľné na Obr. 3. Používateľ má možnosť si vybrať zadávaný a cieľový jazyk prekladu z ponukového menu nachádzajúceho sa v strede obrazovky. Prekladaný text používateľ zadáva do textového poľa v ľavej časti obrazovky. Celá akcia prekladu sa spustí stlačením tlačidla Prelož. Následný výstup z prekladu je zobrazený v elemente na pravej strane obrazovky. Pod týmto elementom sa zobrazí aj nadpis Alternatívne preklady a pod ním je vždy na novom riadku ponúknutá alternatíva prekladu. Elementy s prekladaným a preloženým textom sa automaticky prispôbia šírke používateľovej obrazovky pomocou CSS atribútu width (šírka) zadaného v percentách.



Obr. 3 - Ukážka webového rozhrania služby na štatistický preklad textu

Návrh webovej služby

Keďže náš prekladač má poskytovať služby nielen pre našu aplikáciu, ale má pomáhať aj ďalším programátorom chceme ho poskytnúť ako webovú službu. Táto služba by mala v prvej fáze poskytovať služby na preklad textu a získanie všetkých dostupných prekladov. V budúcnosti ju plánujeme rozšíriť o ďalšie možnosti, ako napríklad nastavenie prekladu a podobne.

3.1.3 Implementácia

Implementácia grafického rozhrania

Používateľ aplikácie na štatistický preklad textu zadáva požadované údaje pomocou grafického používateľského rozhrania (angl. GUI). Používateľ má možnosť zadať voľbu zdrojového a cieľového jazyka prekladu spolu s prekladaným textom. Tieto elementy, do ktorých používateľ

zadáva hodnoty, sú umiestnené v HTML formulári. Potvrdením HTML formulára tlačidlom Prelož, sú hodnoty odoslané na server v hodnotách globálnych PHP premenných `$_POST['type_here']` a `$_POST['language']`. Po ich spracovaní na strane servera je stránka s webovým rozhraním aplikácie znovu načítaná a používateľovi je zobrazený výsledok z celého procesu prekladu.

Implementácia webovej služby

Keďže sme si ako implementačné prostredie vybrali programovací jazyk Java, aj naša webová služba bude implementovaná v tomto jazyku. Tento fakt si vyžiadal nainštalovanie servera Tomcat na náš server.

Následne sme vytvorili webovú službu s dvoma metódami. Webová služba sídli na adrese:

<http://147.175.159.145:8080/aDictItTranslator/services/TranslatorService?wsdl>

V nej je možné volať metódy:

getFirstTranslate

je to metóda na získanie prvého a teda najpravdepodobnejšieho prekladu daného slova. Vracia teda objekt typu *string* v podobe preloženého slova. Táto metóda vyžaduje parametre:

word – typu *string* – je slovo, ktoré chceme preložiť

lang_from – typu *string* – trojpísmenové označenie jazyku v ktorom sa nachádza slovo

lang_to – typu *string* – trojpísmenové označenie jazyku v ktorom chceme mať preložené slovo

getTranslates

metóda získava všetky preklady daného slova. Vracia pole objektov *string*, v ktorom sa tieto preklady nachádzajú.

word – typu *string* – je slovo, ktoré chceme preložiť

lang_from – typu *string* – trojpísmenové označenie jazyku v ktorom sa nachádza slovo

lang_to – typu *string* – trojpísmenové označenie jazyku v ktorom chceme mať preložené slovo

3.1.4 Testovanie

Zobrazenie rozhrania

Pre testovanie viditeľnosti rozhrania sme použili nasledovný testovací scenár, podľa ktorého sme otestovali funkčnosť implementácie. Tento scenár sme opakovali v prehliadačoch IE6, IE7, IE8, IE9, Mozilla Firefox a Google Chrome všetky na platforme Windows.

Názov	Zobrazenie rozhrania	ID Testu	01-01-01
-------	----------------------	----------	----------

Rozhranie	Domovská stránka aplikácie	ID UC	01
Účel	Zobrazenie rozhrania webovej aplikácie		
Vstupné podmienky	Žiadne		
Výstupné podmienky	Používateľ má zobrazené rozhranie		
Krok	Akcia	Očakávaná akcia	Skutočná reakcia
1.	Zadanie adresy prekladača do internetového prehliadača	Zobrazenie všetkých elementov rozhrania na očakávaných pozíciách	Všetky elementy rozhrania sa zobrazili na očakávaných pozíciách

Schopnosť odpovede

Pre overenie odosielania odpovede služby sme použili nasledovný testovací scenár, podľa ktorého sme otestovali funkčnosť implementácie.

Názov	Odpoveď služby	ID Testu	01-01-02
Rozhranie	Domovská stránka aplikácie	ID UC	01
Účel	Overenie odpovede služby		
Vstupné podmienky	Zadaný výraz		
Výstupné podmienky	Používateľ vidí na obrazovke preložený výraz		
Krok	Akcia	Očakávaná akcia	Skutočná reakcia
1.	Získanie odpovede služby	Zobrazenie výsledku prekladu a alternatív prekladu	Výsledky prekladu vrátane alternatívnych výsledkov sa zobrazili v požadovaných HTML elementoch

3.2 Preklad slova

Ako používateľ chcem preložiť slovo.

3.2.1 Analýza existujúcich slovníkov

Na preklad slova z jedného jazyka do druhého je potrebné, mať slovník s dostatočne veľkým počtom hesiel. Veľká časť koncových používateľov používa webové slovníky. Na internete je ich dostatok a väčšinou poskytujú podobné rozhrania a funkcionality. Na účely nášho projektu sa však nemôžeme spoliehať na webové služby externých entít. Lepším riešením je disponovať kompletným slovníkom, čím získame flexibilitu manipulácie s ním a hlavne efektívnosť. Pre tento účel sme sa snažili nájsť dostatočne veľký slovník, teda taký, ktorý by pokryl čo najväčšiu časť jazyka. Zo začiatku sme sa rozhodli implementovať preklad z anglického do slovenského jazyka.

Wiktionary - Wikislovník

Wiktionary je projekt, ktorý vznikol za účelom vytvorenia voľného viacjazykového slovníka. Princíp je podobný ako u Wikipédie. Heslá do slovníka pridávajú používatelia - dobrovoľníci. Pridávajú aj definície jeho významu, preklady do viacerých jazykov a slovné tvary vo všetkých pádoch. Slovenský slovník však v súčasnosti obsahuje len približne 1500 hesiel. O niečo lepšie

je na tom český slovník, ktorý má viac než 29000 hesiel, ale aj to pokrýva len veľmi malú časť jazyka na to, aby vytvoril plnohodnotný prekladač.

Freelang

Freelang je voľne dostupný softvér, ktorý slúži na precvičovanie cudzieho jazyka a na preklad slov. Poskytuje možnosť stiahnuť prekladový slovník z veľkého množstva dvojíc jazykov. Slovensko-anglický slovník obsahuje len o niečo viac než 3000 hesiel, teda len tie najčastejšie slová jazyka, čo pre prekladač textu, ktorý plánujeme vytvoriť rozhodne nemôže stačiť.

PC Translator

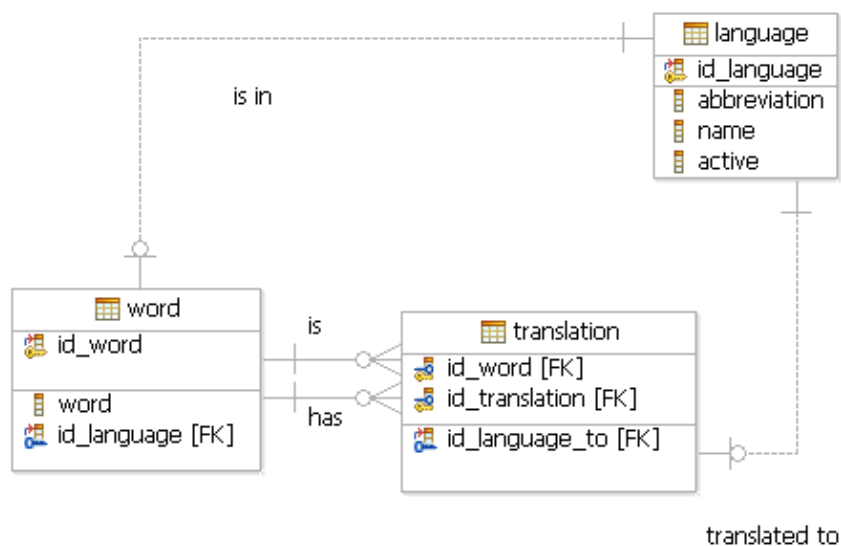
PC Translator je desktopová aplikácia, ktorá slúži ako prekladač. Je veľmi populárny na území Českej a Slovenskej republiky. Je dostupný v ôsmich jazykových verziách, pričom slovenský slovník obsahuje takmer 800 000 hesiel. Pri hlbšej analýze sme zistili, že každý prekladový slovník v aplikácii je uložený v súbore formátu DBF. Tento formát súboru sme schopní spracovať a prispôbiť pre naše potreby. V neskoršej fáze projektu by bolo možné pridať aj slovníky iných jazykov.

Zhodnotenie

Rozhodli sme sa použiť slovník z aplikácie PC Translator. Dôvodom je hlavne dobrá slovná zásoba a spracovateľnosť jeho dát. Ostatné dostupné slovníky neposkytujú dostatočné pokrytie jazyka a tým pádom nemôžu slúžiť k vytvoreniu plnohodnotného prekladača.

3.2.2 Návrh slovníka

Na obrázku Obr. 4 je zobrazený fyzický model slovníka. Pozostáva z tabuliek word, language a translation. Tabuľka language predstavuje jednotlivé jazyky ktorých slová sú v slovníku obsiahnuté. V tabuľke word je definované slovo a jazyk do ktorého patrí. Tabuľka translation je väzobnou tabuľkou ktorá určuje, že ktoré slovo je prekladom inému slovu v určenom jazyku.



Obr. 4 - Fyzický model slovníka

3.3 Detailný popis tabuliek

Tabuľka 1 - Entita fyzického modelu - language

Názov	language		
Kľúč	Názov	Typ	Opis
PK	id_language	int(11)	Primárny kľúč tabuľky
	abbreviation	varchar(5)	Skratka názvu jazyka
	name	varchar(100)	Názov jazyka
	Active	int(1)	Príznak aktívnosti jazyka, či je možné použiť pri preklade tento jazyk

Tabuľka 2 - Entita fyzického modelu - word

Názov	word		
Kľúč	Názov	Typ	Opis
PK	id_word	int(11)	Primárny kľúč tabuľky
FK	id_language	int(11)	Referencia na jazyk, do ktorého slovnej zásoby dané slovo patrí
	word	varchar(255)	Slovo v slovníku

Tabuľka 3 - Entita fyzického modelu - translation

Názov	translation		
Kľúč	Názov	Typ	Opis
PK	id_word	int(11)	Časť primárneho kľúča tabuľky, referencia na slovo, ku ktorému je definovaný preklad
PK	id_translation	int(11)	Časť primárneho kľúča tabuľky, referencia na slovo, ktoré je prekladom pôvodného slova v danom jazyku
FK	id_language_to	int(11)	Referencia na jazyk, do ktorého slovnej zásoby preklad patrí

3.3.1 Implementácia

Slovník prekladača je implementovaný ako databáza v MySQL. Pre získanie prekladu bola implementovaná metóda v jazyku Java. Na komunikáciu s databázou bolo použité rozhranie JDBC.

Princíp je veľmi jednoduchý. Vstupnými parametrami sú slovo, ktoré má byť preložené, jazyk, z ktorého slovo pochádza a výsledný jazyk. Výstupom metódy je zoznam slov, nakoľko jedno slovo môže mať viacero prekladov a synonym.

Zdrojový kód funkcie:

```

PreparedStatement selWord = cm.prepareStatement(
    "SELECT " +
        "W_TO.word " +
    "FROM " +
        "word W_FROM " +
        "JOIN translation T ON W_FROM.id_word = T.id_word " +
        "JOIN word W_TO ON W_TO.id_word = T.id_translation " +
    "WHERE " +

```

```

    "T.id_language_to = ? AND " +
    "W_FROM.word = ? AND " +
    "W_FROM.id_language = ?"
);

selWord.setInt(1, id_language_to);
selWord.setString(2, wordToTranslate);
selWord.setInt(3, id_language_from);

ResultSets = selWord.executeQuery();

List<String>translations = newArrayList<String>();

while(rs.next())
{
    translations.add(rs.getString("word"));
}

```

Testovanie

Pre overenie funkcionality slovníka, sme vytvorili nasledujúce testovacie scenáre ktoré realizujú preklad slov z angličtiny do slovenčiny a zo slovenčiny do angličtiny.

Tabuľka 4 - Akceptačný test prekladu do slovenčiny

Názov	Preklad slova eng->svk	ID testu	01-02-01
Rozhranie	GUI služby (možno premenovať)	ID UC	zzz
Účel	Preklad slova z angličtiny do slovenčiny		
Vstupné podmienky	Žiadne		
Výstupné podmienky	Žiadne		
Krok	Akcia	Očakávaná akcia	Skutočná reakcia
1.	Zo selectu na výber prekladu zvolíme možnosť eng->svk	V selecte sa zvolí možnosť eng->svk	V selecte sa zvolila možnosť eng->svk
2.	Do ľavého poľa napíšeme slovo car	V ľavom poli zostane napísané zadané slovo	V ľavom poli zostalo napísané zadané slovo
3.	Klikneme na tlačidlo prelož	Slovo (car) sa preloží a jeho preklad (auto) sa zobrazí v pravom poli	Slovo (car) sa preložilo a jeho preklad (auto) sa zobrazí v pravom poli

Tabuľka 5 - Akceptačný test prekladu do angličtiny

Názov	Preklad slova svk->eng	ID testu	01-02-02
Rozhranie	GUI služby (možno premenovať)	ID UC	zzz
Účel	Preklad slova zo slovenčiny do angličtiny		
Vstupné podmienky	Žiadne		
Výstupné podmienky	Žiadne		
Krok	Akcia	Očakávaná akcia	Skutočná reakcia
1.	Zo selectu na výber prekladu zvolíme možnosť svk->eng	V selecte sa zvolí možnosť svk->eng	V selecte sa zvolila možnosť svk->eng
2.	Do ľavého poľa napíšeme slovo voda	V ľavom poli zostane napísané zadané slovo	V ľavom poli zostalo napísané zadané slovo

3	Klikneme na tlačidlo prelož	Slovo (voda) sa preloží a jeho preklad (water) sa zobrazí v pravom poli	Slovo (voda) sa preložílo a jeho preklad (water) sa zobrazil v pravom poli
---	-----------------------------	---	--

4 Druhý šprint – pivo

Druhý šprint sme pracovne nazvali pivo, ako mierne silnú návykovú látku. V tomto šprinte sme sa zamerali na nasledujúce fakty:

1. Preklad vety zadanej v jazyku
2. Zapojenie článkov z databázy sme do procesu
3. Pridanie tvaroslovia

Výstupom druhého šprintu je prekladač, ktorý dokáže preložiť vetu a v prípade že ju nájde v našom korpuse sme článkov tak vetu dokáže aj správne vyskloňovať.

4.1 Preklad vety

Ako *používateľ* chcem preložiť *vetu*.

4.1.1 Analýza

Analýza existujúcich riešení na preklad textu

V súčasnosti je dostupných niekoľko možností ako rýchlo získať preklad textu. Používateľ má k dispozícii veľké množstvo webových služieb ale aj prekladačov, ktoré môžu byť distribuované i ako inštalácie do osobného počítača alebo notebooku. Väčšina z nich vytvára potenciálne preklady, ktoré následne ohodnotí percentom úspešnosti. Niektoré tiež poskytujú interakciu na overenie prekladu, či možnosť jeho korektného napísania.

V analýze sa budeme podrobnejšie zaoberať nasledujúcimi prekladačmi:

- **Google Translator**
 - najznámejší a najpoužívanejší
 - aktuálne zvláda preklad do 64 jazykov
 - nevýhodou je obmedzené používanie, keďže Google Translate API v2 bolo spoplatnené
 - medzi výhody patrí overenie prekladu používateľom
- **Bing**
 - vyvinutý z Live Search Translator a Windows Live Translator
 - druhá najpoužívanejšia webová služba slúži na preklad textu
 - podporuje 37 jazykov
- **World Lingo**
 - platená verzia rozpoznáva viac ako 141 jazykov
 - neplatená verzia rozpoznáva iba 32 jazykov a absentuje slovenčina
- **MOSES**
 - prekladač, vyvíjaný niekoľkými vysokoškolskými študentmi a pedagógmi niekoľkých krajín sveta
 - automatické učenie sa prekladov
 - nezávislosť od jazyka – stačí množina korpusov v dvoch rôznych jazykoch
 - absencia slovenčiny
- **PC Translator 2010**

- o produkt spoločnosti LangSoft s.r.o., ktorá sa jeho zdokonaľovaním zaoberá už od roku 1988
- o výhodou je modul, ktorý používateľa naučí cudzí jazyk
- o nevýhoda je menší počet jazykov

Preklad viet

Skúmanie výsledkov prekladu nemusí byť vždy objektívne. Posudzovať ich správnosť napríklad podľa korektnosti celej vety, nie je príliš vhodné. Preto je potrebné zadefinovať systém ohodnocovania viet, ktorý bude brať do úvahy aj iné aspekty. Pri jazykoch ako je slovenčina a čeština to môže byť napríklad skloňovanie a časovanie, pri angličtine zase vhodnosť použitia daného slova alebo výsledný slovosled vo vete. V nasledujúcej tabuľke je uvedené bodové skóre, podľa ktorého boli vety pri prekladoch vyhodnocované.

Tabuľka 6 - Hodnotenie viet podľa úspešnosti prekladu.

Preklad	Bodov
<i>správny preklad</i>	5
<i>nesprávny preklad</i>	0
<i>časovanie</i>	3
<i>skloňovanie</i>	3
<i>slovosled</i>	4
<i>iné alebo zle preložené slovo</i>	1
<i>kombinácia chýb</i>	počet / 2

Majme konečnú množinu, ktorá obsahuje vety na preloženie. Postupne spravíme ohodnotenie každej z nich – hodnota i -tého ohodnotenia je $body_i \in \langle 0,5 \rangle$. Ich súčet tvorí čitateľa zlomku.

Menovateľ je zložený zo súčinu celkového počtu viet N a maximálneho počtu bodov – max_body (v našom prípade 5). Ich podiel vynásobený číslom sto tvorí percento úspešnosti daného prekladača.

Rovnica 1 - Výpočet úspešnosti prekladača.

$$úspešnosť = \left[\frac{\sum_{i=1}^N body_i}{N * max_body} \right] * 100;$$

Zhodnotenie prekladu viet

Z daného prieskumu (Príloha A) jednoznačne vyplýva, že najväčším problémom pre prekladače je jazyk obsahujúci veľa skloňovacích pravidiel. V prípade jednoduchších jazykov (ako napríklad angličtina) je výsledok zobrazený používateľovi buď celý správny alebo sa nedá vôbec pou-

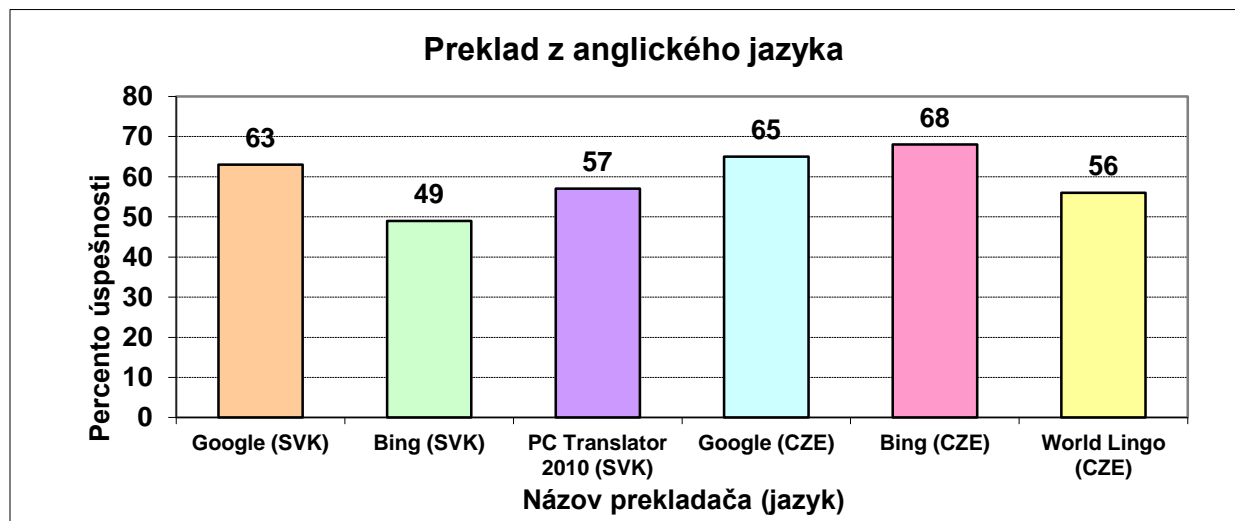
žiť. Jediným z problémom je aj časovanie slovies. Správny tvar sa však nedá zistiť, keďže vety sú zväčša vytrhnuté z kontextu.

Celkové výsledky:

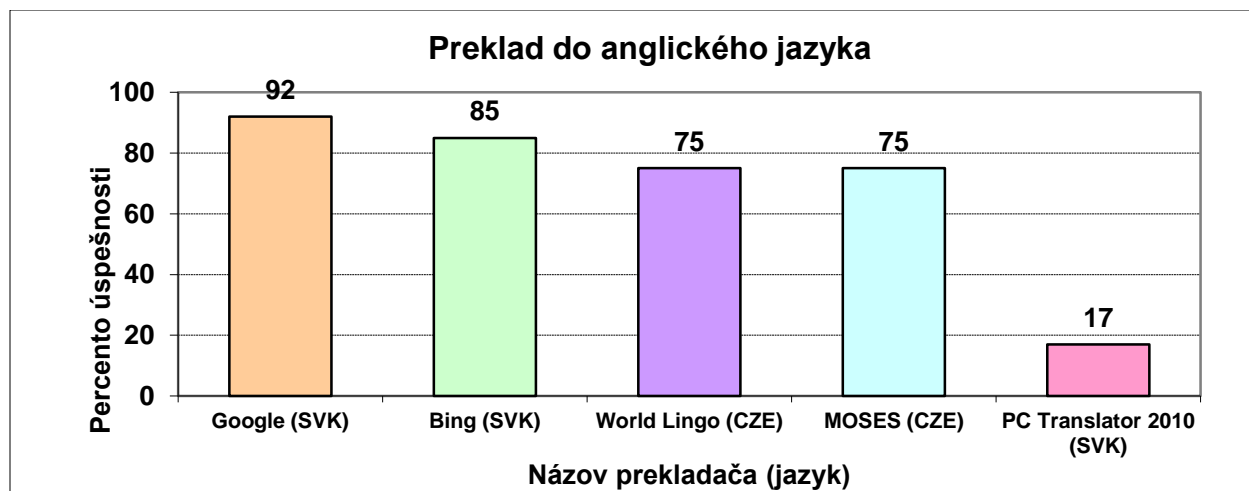
1. Google Translator – 74%
2. Bing – 67 %
3. World Lingo – 52 %
4. PC Translator – 37 %

Poznámka:

- MOSES dokáže prekladať iba jedným smerom a preto nebol zahrnutý do celkových výsledkov



Obr. 5 - Graf úspešnosti prekladu vety z anglického jazyka.



Obr. 6 - Graf úspešnosti prekladu vety do anglického jazyka.

Preklad častí textu - paragrafov

Predchádzajúca kapitola (+ Príloha A) rozoberala preklad jednej vety medzi dvoma jazykmi. Používateľ ale zväčša potrebuje získať ucelený pohľad na určitú časť kontextu, ktorej nerozumie. Preto je nevyhnuté, aby prekladače vedeli spracovávať zadané údaje nielen po slovách či vetách, ale aby aj odvodzovali vzťahy medzi jednotlivými časťami kontextu. Jednotlivo preložené vety by mohli byť z rôznych tém, vyskytovalo by sa miešanie časov, či mužských a ženských osôb. Preložená časť by pôsobila rušilo.

Na výpočet ohodnotenia viet použijeme Rovnica 1 z kapitoly „Preklad viet“ a bodovanie z Tabuľka 7.

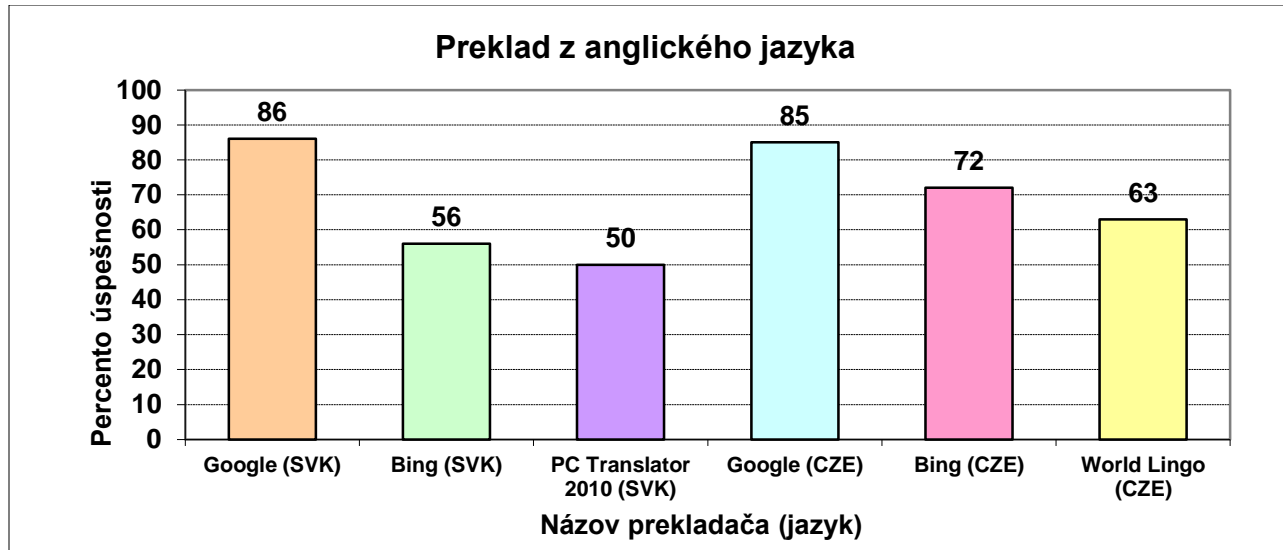
Tabuľka 7 - Hodnotenie viet podľa úspešnosti prekladu.

Preklad vety v kontexte	Bodov / slovo
<i>správny preklad</i>	5
<i>nesprávny preklad</i>	0
<i>časovanie</i>	3
<i>skloňovanie</i>	3
<i>slovosled</i>	4
<i>slovo navyše</i>	2
<i>iné alebo zle preložené slovo</i>	1

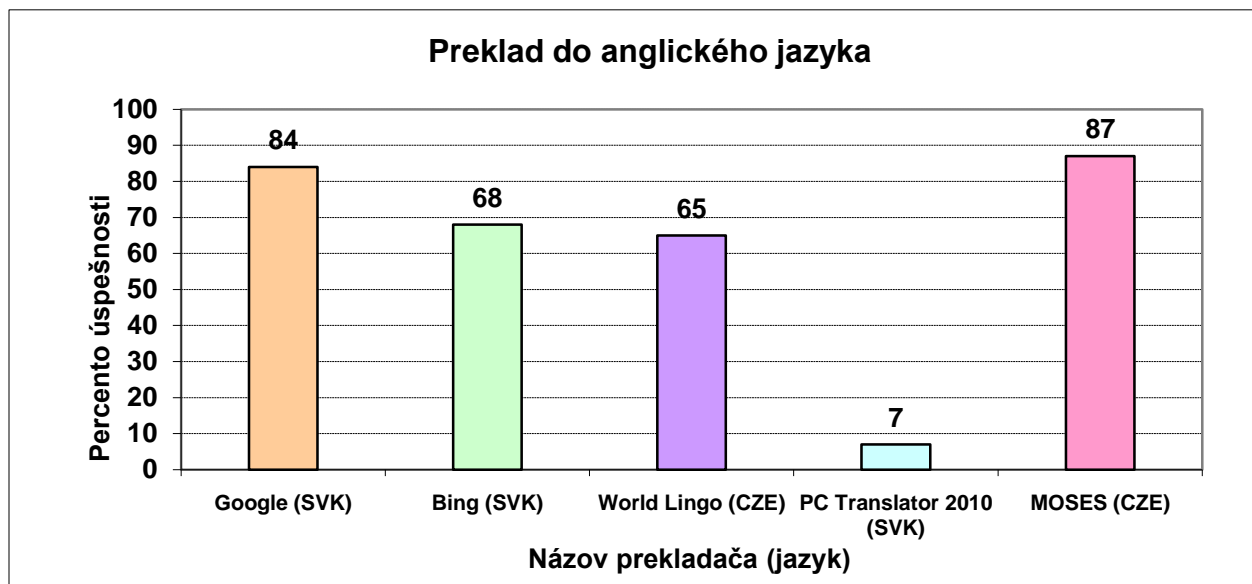
Zhodnotenie prekladu paragrafov

Prekladanie ucelených častí textu (Príloha B) je oveľa zložitejší proces ako preklad samostatnej vety. Existujúca dejová línia, kumulovanie prídavných mien, striedanie časov, či dialógy osôb. To všetko je potrebné zapracovať do algoritmu prekladu. Postupné prekladanie slov a ich vkládanie do vety môže mať úspech iba pri jazykoch, ktoré nepoznajú skloňovanie. V opačnom prípade však narážame na problém. Príkladom takéhoto prekladača je napríklad PC Translator 2010.

Percentuálna úspešnosť bola znázornená pomocou dvoch stĺpcových grafov. Príjemným prekvapením je MOSES, ktorého schopnosti je možné porovnávať s takým gigantom ako je Google Translate. Škoda len, že sa špecializuje iba na preklad z češtiny do angličtiny.



Obr. 7 - Graf úspešnosti prekladu textu z anglického jazyka.



Obr. 8 - Graf úspešnosti prekladu textu do anglického jazyka.

Celkové zhodnotenie

Analýza problémovej oblasti bola zameraná na porovnanie prekladu medzi anglickým, českým a slovenským jazykom. Bola zložená z dvoch častí, ktoré jednoznačne potvrdili, že skloňovanie je najväčším „nepriateľom“ prekladača.

Počítanie percentuálnej úspešnosti bolo vykonané s pomocou presne zadaného modelu ohodnocovania. Výsledky však stále nie sú uspokojivé. Je potrebné ich neustále zdokonaľovať. To možno dosiahnuť napríklad aplikovaním novej funkcie, ktorá určí pravdepodobnosti výskytu za pomoci iných kauzálnych vzťahov.

V súčasnosti na trhu absentuje nástroj, ktorý by bol schopný konkurovať, či prípadne nahradiť najznámejšiu aplikáciu slúžiacu na preklad textov. Ním je Google Translate. Ten však „uštedril“

programátorom tvrdý úder pod pás – spoplatnil Translate API. Nastal teda čas pokúsiť sa vymyslieť niečo nové.

Analýza slovenského jazyka a spojených problémov

Slovenský jazyk

Pre preklad textu do slovenského jazyka je nevyhnutné analyzovať problémy spojené s jazykom. Slovenčina patrí medzi západoslovenské jazyky a je používaná viac ako šiestimi miliónmi ľudí. Zaraduje sa medzi tzv. flexné jazyky a je charakteristická bohatou morfológiou. Slová v slovenčine majú viacero tvarov a pravidlá pre tvorbu tvarov nie sú striktné definované. Tento fakt predstavuje problém pre tvorbu algoritmov, ktorých cieľom je úprava slova na konkrétny tvar. Na zápis slovenčiny sa používa slovenská abeceda pozostávajúca z modifikovaných písmen latinky a je rozšírená o diakritické znamienka. Písanie diakritických znamienok sa v neformálnom texte alebo komunikácií prostredníctvom internetu často nevyužíva [1, 2].

Získanie konkrétneho tvaru slova

Zmeniť slovo do požadovaného tvaru je dôležitý krok pri preklade. Bez neho by nebolo možné zo slov vyskladať spisovnú slovenskú vetu. V praxi sa najčastejšie využíva opačný proces, a to získanie základného tvaru slova – lematizácia. Keďže by stačilo tento proces otočiť, aby sme získali ľubovoľný tvar slova, analyzujeme možné prístupy lematizácie. Vo všeobecnosti sa delia do dvoch skupín a to [3]:

- slovníkové
- pravidlové

Slovníkové prístupy využívajú databázu slov jazyka a všetkých prislúchajúcich tvarov. Aktuálna databáza slovenských slov od Jazykovedného ústavu Ľudovíta Štúra pozostáva z 2 472 051 záznamov vo forme základný tvar a odvodený tvar. Jedinou nevýhodou tejto databázy je nepoužiteľnosť pre nové slova jazyka, ktoré v nej nie sú obsiahnuté. Úspešnosť získavania základného tvaru slovníkovým prístupom dosahuje veľmi dobré výsledky. Databáza slov od jazykovedného ústavu upravená na lematizátor dosiahla v experimente na bežných a odborných článkoch (3209 slov) úspešnosť lematizácie až 96,1% [2].

Pravidlové prístupy sa opierajú o pravidlá tvorby slovných tvarov a snažia sa ich aplikovať pre získavanie základného tvaru. Medzi nástroje využívajúce takýto prístup patrí nástroj *Tvaroslovník*¹. Autori vytvorili slová, ktoré predstavujú šablóny ohýbania pre iné slová a pomocou pravidiel podvojnovej výmeny je možné získavať ľubovoľné tvary slov. Výhodou v porovnaní so slovníkovým prístupom je možnosť získať základný tvar aj pre nové slová, ktoré sa v databáze slov nenachádzajú. Potrebné je iba nájsť vhodnú šablónu pre ohýbanie požadovaného slova. Nástroj *Tvaroslovník* v rovnakom experimente, na akom bola testovaná databáza slov od Jazykovedného ústavu Ľ. Štúra, dosiahol úspešnosť lematizácie 84,2% [2, 3].

¹ <http://nazou.fiit.stuba.sk/home/index.php>

Vidíme, že slovníkový prístup je pre úpravu slova na základný tvar efektívnejší, a preto predpokladáme, že bude efektívnejší aj pri získavaní odvodených slovných tvarov.

Analýza databázy SME článkov

Databázu SME článkov tvorí 99164 článkov, ktoré boli získané na fakulte v rámci výskumného projektu. Tie bolo zabalené v SQL súbore, ktorého spustením sa vytvorí a naplní tabuľka articles.

#	Stĺpce	Typ	Zotriedenie	Atribúty	Nulový	Predvolené	Extra	Akcia
1	id	int(11)			Nie	None	AUTO_INCREMENT	Zmeniť Odstrániť More ▼
2	title	varchar(255)	utf8_general_ci		Nie	None		Zmeniť Odstrániť More ▼
3	body	longtext	utf8_general_ci		Nie	None		Zmeniť Odstrániť More ▼
4	author	varchar(255)	utf8_general_ci		Áno	NULL		Zmeniť Odstrániť More ▼
5	published_at	datetime			Nie	None		Zmeniť Odstrániť More ▼
6	created_at	datetime			Nie	None		Zmeniť Odstrániť More ▼
7	updated_at	datetime			Nie	None		Zmeniť Odstrániť More ▼
8	logger_url	varchar(255)	utf8_general_ci		Áno	NULL		Zmeniť Odstrániť More ▼
9	perex	longtext	utf8_general_ci		Nie	None		Zmeniť Odstrániť More ▼
10	section	varchar(255)	utf8_general_ci		Nie	None		Zmeniť Odstrániť More ▼
11	category	varchar(255)	utf8_general_ci		Nie	None		Zmeniť Odstrániť More ▼
12	url	varchar(255)	utf8_general_ci		Nie	None		Zmeniť Odstrániť More ▼
13	ari	decimal(8,3)			Áno	NULL		Zmeniť Odstrániť More ▼
14	type	varchar(15)	utf8_unicode_ci		Nie	None		Zmeniť Odstrániť More ▼

↑ Označiť všetko / Odznačiť všetko Výber: Prechádzať Zmeniť Odstrániť Primárny Unikátny Index Celý text

Obr. 9 - Štruktúra tabuľky articles.

Primárnym kľúčom je „id“. V stĺpci body je uložený text celých článkov. Ostatné stĺpce aktuálne nie sú pre náš projekt podstatné.

Problémy:

1. text článkov obsahuje nepotrebné html značky
2. hľadanie slov alebo viet v celých článkoch je pomalé – potreba rozdelenia na menšie časti
3. vhodný návrh a prepojenie tabuliek

4.1.2 Návrh

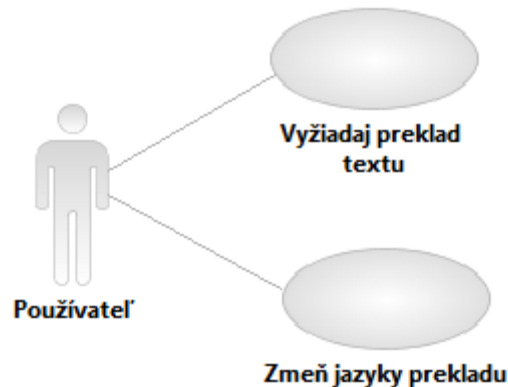
Metóda pre preklad textu

Preklad textu z jedného jazyka do druhého je komplikovaný proces, na ktorý samotné preloženie jednotlivých slov nie je postačujúce. Pre úspešný preklad je dôležité aby preložená veta dodržala sémantiku jazyka a vznikla tak zmysluplná veta. Analýza Slovenského jazyka ukázala, že vytvorenie algoritmu pre generovanie vhodných slov je veľmi náročný proces a kvôli nie striktno definovaným pravidlám tvorby slov nemôže dosahovať veľkú úspešnosť. Preto sme snahu o vytvorenie pravidiel tvorby viet zavrholi hneď na začiatku.

Jednou alternatívou riešenia je použitie prístupu spomínaného u nástroja Moses. Pre vytvorenie takéhoto riešenia je nevyhnutné získať paralelne preložené korpusy, ktoré budú slúžiť ako množina príkladov pre preklad textu. Avšak existujúce paralelné texty pre slovenčinu sú nevhodné pre preklad obyčajného textu a vytvorenie vlastných korpusev je časovo príliš náročný proces. Navrhujeme preto použiť prístup založený na štatistike, ktorý bude kombinovať preklad jednotlivých slov, získanie ich odvodených tvarov a generovanie potenciálnych prekladov. Ich správnosť sa overí pomocou štatistických metód na korpuse slovenských textov.

Prípady použitia

V našom systéme sme identifikovali nasledovné prípady použitia (Obr. 10):



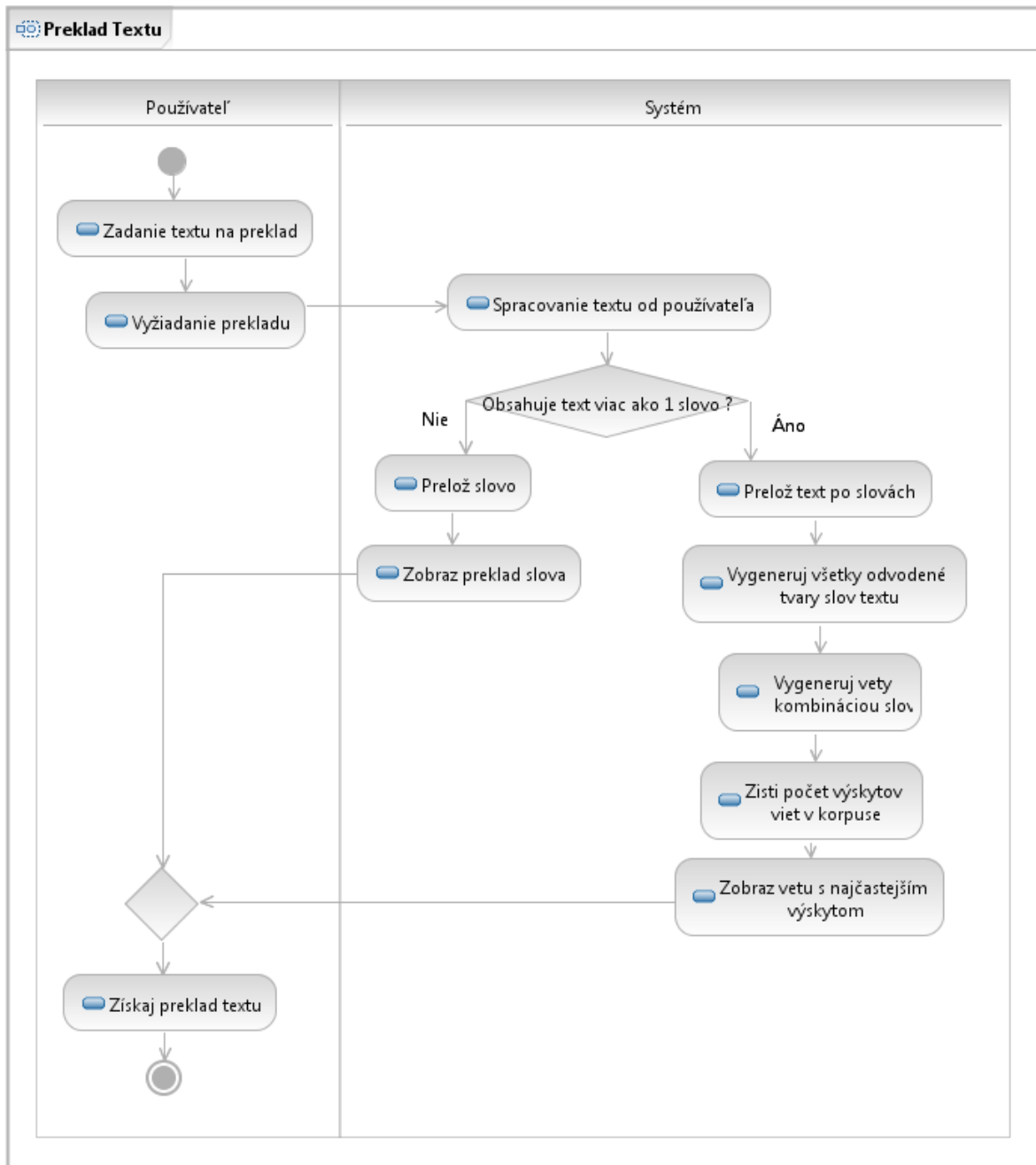
Obr. 10 - Diagram prípadov použitia.

Aby bolo možné získať preklad textu je nevyhnutné, aby súčasťou systému boli nasledujúce časti:

- slovník pre preklad jednotlivých slov
- generátor slovných tvarov pracujúci nad databázou slovenských slov od Jazykovedného ústavu Ľudovíta Štúra
- generátor viet zo všetkých kombinácií slov
- korpus slovenských textov
- vyhľadávač potenciálnych prekladov v korpuse
- webová stránka poskytujúca jednoduché rozhranie
- webová služba pre prepojenie používateľského rozhrania s jadrom aplikácie

Diagram aktivít

Ako je vidieť na Obr. 11, proces začína vložением textu na preklad do systému od používateľa. Prvým krokom je rozdelenie textu na slová, ktoré sú po jednom preložené pomocou slovníka. Tým dostaneme všetky významy preložených slov. Ďalším krokom je získanie všetkých odvodených tvarov preložených slov prostredníctvom databázy tvarov slovenských slov. To by malo

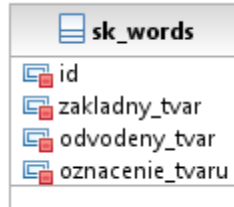


Obr. 11 - Diagram aktivít.

zabezpečiť, že sa vygenerujú správne vyskloňované slová. Nasleduje tvorenie viet kombinovaním vygenerovaných tvarov, čím vznikne množina potenciálnych prekladov. Posledným krokom je hľadanie týchto viet v korpuse. V prípade, že sa podarí nájsť presnú zhodu, hľadanie je zastavené a používateľovi sa zobrazí príslušná veta ako preklad. Zvyšné vety sú zobrazené len ako alternatívne preklady.

Fyzický model generátora slovných tvarov a viet

Na Obr. 12 je zobrazený fyzický model databázy slov z Jazykovedného ústavu Ľudovíta Štúra. Obsahuje len jednu tabuľku, v ktorej sú uložené všetky tvary slovenských slov, označenie týchto tvarov a prislúchajúce základné tvary.



Obr. 12 - Fyzický model generátora slovných tvarov.

V tabuľke sa bude často hľadať buď základný tvar slova alebo jeho odvodený tvar. Keďže obsahuje približne 2,4 milióna záznamov je dôležité, aby tieto atribúty boli indexované.

Návrh metódy využitia sme článkov

Problém č.1: Text článkov obsahuje nepotrebné html značky.

Každý z článkov obsahuje konkrétny text, ktorý je zapísaný vo forme vhodného výstupu na webovú stránku týždenníka SME. Obsahuje teda rôzne HTML značky, ktoré slúžia na oddelenie paragrafov, na zvýraznenie určitých častí, či na presun na inú stránku. Často krát sa v strede článku nachádzajú i reklamy. Pre štatistickú prácu a spracovanie dát je potrebné, aby tieto články neobsahovali HTML značky. Jedinou výnimkou je označenie paragrafu (odseku), ktoré neskôr posluží ako tzv. „okolie“ slova. To znamená, že za pomoci kontextu budeme vedieť určiť konkrétny význam slova. V slovenčine, ale i v iných jazykoch, existujú slová, ktoré majú niekoľko významov. Medzi takéto slovo patrí napríklad slovo hlava.

<h3>Čo je to RSS?</h3>

<p>Riešením je <acronym title="Really simple syndication">RSS </acronym>. Ide o veľmi primitívny spôsob, ako zo stránok sťahovať iba to, čo je na nich nové, rovnako ako z e-mailového servera sťahujete iba neprečítané e-maily. Autori stránok na internete uverejnia textový <acronym title="Extensible Markup Language">XML </acronym> výstup stránky, ktorý potom môže špeciálny program v pravidelných intervaloch kontrolovať a následne používateľa touto cestou informovať o novopridanom obsahu webstránky.</p>

Parse HTML tags
without paragraphs



Čo je to RSS?

<p>Riešením je RSS. Ide o veľmi primitívny spôsob, ako zo stránok sťahovať iba to, čo je na nich nové, rovnako ako z e-mailového servera sťahujete iba neprečítané e-maily. Autori stránok na internete uverejnia textový XML výstup stránky, ktorý potom môže špeciálny program v pravidelných intervaloch kontrolovať a následne používateľa touto cestou informovať o novopridanom obsahu webstránky.</p>

Obr. 13 - Odstránenie nepotrebných HTML značiek.

Problém č.2: Hľadanie slov alebo viet v celých článkoch je pomalé.

Vyhľadávanie kľúčových slov, prekladov, či viet nad veľkým korpusom prác je neefektívne. Text článku je preto potrebné rozdeliť na menšie časti – vety. Každá veta musí končiť bodkou,

otáznikom alebo výkričníkom. Tu však je potrebné zdôrazniť, že aj číslovania alebo iniciálky mien obsahujú bodku.

Algoritmus navrhnutého riešenia:

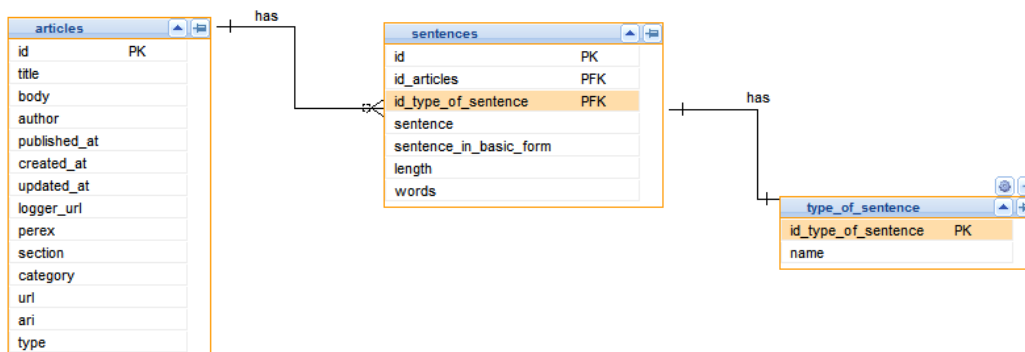
- Krok č.1.: načíta text článku z tabuľky articles
- Krok č.2.: odstráni značky paragrafov
- Krok č.3.: odstráni medzery za bodkou, dvojbodkou, otáznikom a výkričníkom
- Krok č.4.: odstráni v texte všetky dve a viac medzery za sebou
- Krok č.5.: nastaví začiatok vety na nulu
- Krok č.6.: hľadá symbol bodky, dvojbodky, otáznika alebo výkričníka a za ním veľké písmeno – nastaví koniec vety a nový začiatok druhej vety
- Krok č.7.: pomocou známeho začiatku a konca vytiahne danú vetu z kontextu a uloží do pola
- Krok č.8.: ak už nenájde symbol bodky, dvojbodky, otáznika alebo výkričníka a za ním veľké písmeno – ide o poslednú vetu, ktorú uloží do pola
- Krok č.9.: overí vytvorené pole
 - a.: zmazanie príliš krátkych viet (napr. číslovanie, skratka mena a pod.)
 - b.: zmazanie medzery, ak ňou veta začína alebo končí
- Krok č.10.: zistí typ vety pomocou posledného znaku
- Krok č.11.: zistí dĺžku a počet slov vo vete
- Krok č.12.: uloží vetu do tabuľky sentences

Problém č.3: Vhodný návrh a prepojenie tabuliek.

Tabuľka articles jednoznačne identifikuje daný článok pomocou svojho parametru id. Ten sa prenáša aj do tabuľky viet (sentences), ktoré z neho vznikli. Vystupuje tu ako cudzí kľúč s názvom id_articles. Dané vety môžu byť určitého typu (oznamovacie, opytovacie, rozkazovacie), čo je reprezentované pomocou ďalšieho cudzieho kľúča id_type_of_sentences. Ten pochádza z tabuľky type_of_sentence.

Platia nasledujúce vzťahy:

- článok má niekoľko viet
- veta je práve jedného typu



Obr. 14 - Návrh databázového modelu SME článkov.

4.1.3 Implementácia

Generátor slovných tvarov a viet

Generátor je implementovaný formou metódy v jazyku Java a komunikuje s MySQL databázou slovných tvarov. Pre komunikáciu s databázou používa rozhranie JDBC. Prostredníctvom generátora je možné získať:

- základný tvar slova,
- všetky odvodené tvary slova,
- všetky kombinácie všetkých slovných tvarov zo zadanej množiny slov.

Ukážka zdrojového kódu pre získanie základného tvaru slova:

```
public ArrayList<String> getLemma(final String word, final String tableName) throws SQLException {
    final String query = "SELECT * FROM " + tableName + " WHERE odvodeny_tvar='" + word + "'";
    final Statement statement = connection.createStatement();
    final ResultSet resultSet = statement.executeQuery(query);
    final ArrayList<String> output = new ArrayList<String>();
    int index = 0;
    try {
        while (resultSet.next()) {
            output.add(index, resultSet.getString("zakladny_tvar"));
            index++;
        }
    } finally {
        statement.close();
        resultSet.close();
    }
    return output;
}
```

Ukážka zdrojového kódu pre získanie odvodených tvarov slova:

```
public ArrayList<String> getWordForms(final String word, final String tableName) throws SQLException {
    final String query = "SELECT * FROM " + tableName + " WHERE zakladny_tvar = 'binary'" + word + "'";
    final PreparedStatement statement = connection.prepareStatement(query);
    final ResultSet resultSet = statement.executeQuery(query);
    final ArrayList<String> output = new ArrayList<String>();
    int index = 0;
    try {
        while (resultSet.next()) {
            output.add(index, resultSet.getString("odvodeny_tvar"));
            output.add(index+1, resultSet.getString("oznacenie_tvaru"));
            index += 2;
        }
    } finally {
        statement.close();
        resultSet.close();
    }
    return output;
}
```

Generátor kombinácií slovných tvarov, využíva metódu pre získanie slovných tvarov. Vďaka rekurzívnemu volaniu tak dokáže vytvoriť všetky možné kombinácie zo zadaných slov.

Vyhľadanie vety v korpuse

Ide o metódu implementovanú v jazyku Java, ktorá komunikuje s generátorom viet, a MySQL databázou slovenských textov. Cieľom je nájsť niektoré z potenciálnych viet v korpuse. Ak sa vetu podarí nájsť, je isté, že ide o takú, ktorá je spisovná, a tak môže byť zobrazená používateľovi. Pre komunikáciu s databázou používa rozhranie JDBC.

Ukážka zdrojového kódu pre hľadanie viet v korpuse:

```
public String findInDatabase(final ArrayList<String> vety, final String user, final String pass)
throws InstantiationException, IllegalAccessException, ClassNotFoundException, SQLException {
    final Iterator<String> itr = vety.iterator();
    PreparedStatement prepStatement = null;
    ResultSet resultSet = null;
    try {
        while (itr.hasNext()) {
            final String tmp = itr.next();
            final String query = "SELECT * FROM sentences WHERE sentence = '"+tmp+"'";
            prepStatement = connection.prepareStatement(query);
            resultSet = prepStatement.executeQuery();
            if (resultSet.next()) {
                return tmp;
            }
        }
    } finally {
        prepStatement.close();
        resultSet.close();
    }
    return null;
}
```

Implementácia metódy pre vyhľadávanie SME článkov

Problém č.1: Text článkov obsahuje nepotrebné html značky.

Pri odstraňovaní HTML značiek sú použité funkcie externej knižnice *JSoup* (v.1.6.1.) - *parse(String text)* a *text()*.

```
for(result = cmd.executeQuery("SELECT COUNT(id) FROM articles"); result.next();)
{
    pocet = result.getInt(1); // získaj pocet viet
}

for(i = 0; i <= pocet; i++) // prejdí všetky vety
{
    ps_select.setInt(1, i);
    for(result = ps_select.executeQuery(); result.next(); ps_update.executeUpdate()) // cyklus pre vyberanie konkrétneho článku
    {
        text = result.getString(2); // získaj text - parameter body
        text = text.replaceAll("<p>", "_start_parag_"); // nahrad <p> a </p> špeciálnym retazcom
        text = text.replaceAll("</p>", "_end_parag_"); // stringov
        text = Jsoup.parse(text).text(); // odstrán všetky HTML tagy
        text = text.replaceAll("_start_parag_", "<p>"); // nahrad všetky špeciálne retazce
        text = text.replaceAll("_end_parag_", "</p>");
        ps_update.setBytes(1, text.getBytes("utf-8")); // ulož článok späť do tabuľky articles
        ps_update.setString(2, result.getString(1));
    }
}
```

Obr. 15 - Odstránenie nepotrebných HTML značiek s použitím JSoup funkcií.

Problém č.2: Hľadanie slov alebo viet v celých článkoch je pomalé.

```
// ak na ziaciatku medzera
if (vety[j].startsWith(" ")==true)
    vety[j]=vety[j].substring(1,vety[j].length());

// ak na konci medzera
if (vety[j].endsWith(" ")==true)
    vety[j]=vety[j].substring(0,vety[j].length()-1);

// ak ziaciatku vety nie je pozadovny znak - nastal pripadny posun (korekcia)
if (vety[j].length()>2 && vety[j].charAt(1)==' ' && vety[j].charAt(2)>='A' && vety[j].charAt(2)<='Z')
    vety[j]=vety[j].substring(2,vety[j].length());

// odstrani skratku mena
if (vety[j].length()>2)
{
    // nastavi typ vety
    if (vety[j].endsWith(".")==true)
        typ_vety=1;
    if (vety[j].endsWith("?")==true)
        typ_vety=2;
    if (vety[j].endsWith("!")==true)
        typ_vety=3;

    // zisti pocet slov
    for (int q=0;q<vety[j].length()-1;q++)
        if (vety[j].charAt(q)==' ')
            pocet_slov++;

    ps_insert.setBytes(1, vety[j].getBytes("utf-8")); // nastav text vety
    ps_insert.setInt(2, id); // nastav id clanku
    ps_insert.setInt(3, typ_vety); // nastav typ vety
    ps_insert.setInt(4, vety[j].length()); // nastav dlzku vety
    ps_insert.setInt(5, pocet_slov); // nastav pocet slov
    ps_insert.executeUpdate(); // uloz do tabulky sentences
    pocet_viet++;
}
```

Obr. 16 - Implementácia algoritmu na vytváranie viet – krok 9 až 12.

Problém č.3: Vhodný návrh a prepojenie tabuliek.

Databázový model, ktorý sa nachádza na Obr. 16 bol zrealizovaný. V súčasnosti sú tabuľky naplnené dátami.

4.1.4 Testovanie

Pre testovanie funkčnosti prekladača sme navrhli nasledujúce testovacie scenáre, podľa ktorých bola overená funkčnosť prekladača.

Meno	Preloženie slova z angličtiny do slovenčiny	ID-Testu	02-01-01
Rozhranie	Stránka prekladača	ID-UC	01
Účel	Preloženie anglického textu		
Vstupné podmienky	Žiadne		
Výstupné podmienky	Žiadne		
Krok	Akcia	Očakávaná akcia	Skutočná reakcia
1	Zadanie textu pre preklad slo-	Zobrazenie zadaného textu	Zobrazenie zadaného

	va „pen“		textu
2	Zvolenie jazyka „ENG-SVK“ z vyberača jazykov	Zobrazenie správnej hodnoty vyberača „ENG-SVK“	Zobrazenie správnej hodnoty vyberača „ENG-SVK“
3	Stlačenie tlačidla „prelož“	Zobrazenie preloženého textu „pero“ a ďalších významov	Preloženie textu a jeho zobrazenie v zodpovedajúcich poliach

Meno	Preloženie vety z angličtiny do slovenčiny	ID-Testu	02-01-02
Rozhranie	Stránka prekladača	ID-UC	01
Účel	Preloženie anglického textu		
Vstupné podmienky	Žiadne		
Výstupné podmienky	Žiadne		
Krok	Akcia	Očakávaná akcia	Skutočná reakcia
1	Zadanie textu pre preklad vety „itwascoincidence“	Zobrazenie zadaného textu	Zobrazenie zadaného textu
2	Zvolenie jazyka „ENG-SVK“ z vyberača jazykov	Zobrazenie správnej hodnoty vyberača „ENG-SVK“	Zobrazenie správnej hodnoty vyberača „ENG-SVK“
3	Stlačenie tlačidla „prelož“	Zobrazenie preloženého textu „to bola náhoda“ a ďalších významov	Preloženie textu a jeho zobrazenie v zodpovedajúcich poliach

5 Použitá literatúra

- [1] SLOVAKE: *O jazyku*. 2011-03-13 [cit. 2011-05-01]. Dostupné na internete: <http://slovak.ee/sk/intro/language/general>
- [2] HORVÁTH, R.: Spracovanie textu v slovenčine, In: Podpora výkladu neznámeho pojmu pri prehl'adávaní v slovenčine, s. 3-6.
- [3] LACLAVÍK, M. et al.: Dostupné zdroje a výzvy pre počítačové spracovanie informačných zdrojov v slovenskom jazyku. In: 1st Workshop on Intelligent and Knowledge oriented Technologies, Institute of Informatics SAS, 2007, s. 92-98.

5.1 Príloha A – Preklad viet

A.1. Vysvetlenie označenia v prílohách

Pri tabuľkách, ktoré sa vyskytnú v nasledujúcich kapitolách, budú použité symboly. Ich význam je potrebný pre správnu orientáciu vo výsledkov.

Zoznam symbolov:

* - zadaný český výraz rovnakého významu

A.2. Z angličtiny do slovenčiny a češtiny

Prvým objektom záujmu je správnosť prekladu anglickej vety do slovenčiny či češtiny. Na vybraných dvanástich vzorkách bola vypracovaná ukážka ohodnocovania, ktorá zachováva farebnú konvenciu z tabuľky č.1.

č.	Anglická veta	Slovenský preklad			Český preklad		
		Google Translate	Bing Translate	PC Translator 2010	Google Translate	Bing Translate	World Lingo
1.	Where are you from?	Kde ste?	Odkiaľ si?	Odkiaľ ste?	Kde ste?	Odkud jsi?	Odkud jste ?
2.	I have a mother.	Mám matku.	Som matka.	Ja mám matku.	Mám matku.	Mám matku.	Mám matku .
3.	What are you doing in Slovakia?	Čo robíte v Slovensko?	Čo robíš na Slovensku?	Čo robíš na Slovensku?	Co děláte v Slovensko?	Co děláš na Slovensku?	Co děláš na Slovensku ?
4.	We will not block other countries.	Nebudeme blokovat iných krajín.	My neblokuje iných krajín.	My nebudeme blokovat iné krajiny.	Nebudeme blokovat jiných zemí.	Nám nebude blokovat další země.	Nebudeme blokovat ostatní země .
5.	There is no agreement.	Neexistuje žiadna dohoda.	Neexistuje žiadna dohoda.	Neexistuje žiadny dohoda.	Neexistuje žádná dohoda.	Neexistuje žádná dohoda.	Neexistuje žádná dohoda .
6.	Chinese Embassy rented the government hotel for celebration.	Čínskeho veľvyslanectva prenajať vládne hotela k oslave.	Čínskeho veľvyslanectva prenajíma vláda hotel pre oslavu.	Čínsky Embassy si prenajal vládny hotel pre oslavu.	Čínskeho veľvyslanectví pronajmout vládní hotelu k oslavě.	Čínské velvyslanectví pronajal hotel vláda pro oslavu.	Do Šanghaje pronajaté vládou hotel na oslavu .
7.	We are in a situation where we have to do things together.	Sme v situácii, kedy musíme robiť veci spoločne.	Sme v situácii, kde sme robiť veci dohromady.	My sme v situácii kde my musíme urobiť veci spolu.	Jsme v situaci, kdy musíme dělat věci společně.	Jsme v situaci, kdy musíme dělat věci dohromady.	Jsme v situaci, kdy musíme dělat věci dohromady .
8.	Our project is sound and workable.	Náš projekt je zdravý a funkčný.	Náš projekt je zvuk a funkčný.	Náš projekt má dobré jadro a výnosné.	Náš projekt je zdravý a funkční.	Náš projekt je zdravé a proveditelný.	Náš projekt je zvuk a proveditelný .
9.	Eurozone members are in the process of	Členovia eurozóny sú v procese ratifikácie návrhov.	Členov eurozóny sú v procese ratifikácie návrhy.	Eurozone členovia sú v procese ratifikujúcich	Členové eurozóny jsou v procesu ratifikace návrhů.	Členové eurozóny jsou v procesu ratifikace návrhy.	Zeměmi eurozóny jsou v procesu ratifikace

	ratifying proposals.			návrhov.			návrhy.
10.	There has been renewed optimism this week.	Tam bol obnovený optimizmus tento týždeň.	Došlo obnovená optimizmus tento týždeň.	Tam bol obnovený optimizmus tento týždeň.	Tam byl obnovený optimizmus v tomto týdnu.	Je tu nabytého optimizmu tento týden.	Bylo obnovenno optimizmus tento týden .
11.	These helped to boost investor sentiment with stock markets.	Tieto pomohol k posilneniu investičného prostredia s akciovými trhmi.	Tieto pomohol oživiť investora sentiment s akciové trhy.	Tieto pomohol povzbudiť investor cit s trhmi s cennými papiermy.	Tyto pomohl k posílení investičního prostředí s akciovými trhy.	To přispělo k posílení smýšlení investorů s akciové trhy.	Tyto přispělo k posílení investor sentiment s akciových trhů .
12.	The company has limited itself to making black and white e-readers.	Spoločnosť sa obmedzila na výrobu čiernebielejších e-čítateľia.	Spoločnosť sa obmedzila tvorby čierne a biele e-čítateľia.	Spoločnosť limitovaný samo o sebe tvorbe čiernebielejších e-čítateľia.	Společnost se omezila na výrobu černobílejších e-čtenářů.	Společnost se omezila výrobu černých a bílých e-čtenářů.	Společnost se omezila na výrobu černých a bílých E - čtenářů .
Výsledok prekladu		37,5 / 60 => 63 %	29,5 / 60 => 49 %	34 / 60 => 57 %	39 / 60 => 65 %	40,5 / 60 => 68 %	33,5 / 60 => 56 %

Tabuľka A.1: Preklad anglických viet.

A.3. Zo slovenčiny a češtiny do angličtiny

V predchádzajúcej kapitole boli zadefinované určité anglické vzorky. Ich korektný manuálny preklad bol použitý ako základ pre spätné porovnanie výsledkov. Webové služby World Lingo a MOSES nepodporujú slovenský jazyk a preto bol použitý ich český ekvivalent.

		Anglický preklad				
č.	Slovenská veta	Google Translate	Bing Translate	World Lingo *	MOSES *	PC Translator 2010
1.	Odkiaľ pochádzate?	Where are you from?	From where are you from?	Where did you come from?	Where are you from?	From where come form?
2.	Mám matku.	I have a mother.	I have a mother.	I have a mother.	I have a mother.	Shall I mother.
3.	Čo robíte na Slovensku?	What are you doing in Slovakia?	What do you do in Slovakia?	What are you doing in Slovakia?	What do you do in Slovakia?	What are you about in Slovakia?
4.	Nebudeme blokovať ostatné krajiny.	We will not block other countries.	We do not block the other country.	We don't block the other countries.	We will not block the rest of the country.	Nebudeme block others landscapes.
5.	Neexistuje dohoda.	There is no agreement.	There is no agreement.	There is no agreement.	There is no agreement.	Don't exist agreement.
6.	Čínske veľvyslanectvo prenajalo k oslave vládny hotel.	The Chinese embassy rented the hotel to celebrate the government.	The Chinese Embassy prenájalo to celebrate a Government hotel.	China's embassy around to celebrate government the hotel.	The Chinese Embassy pronajalo to celebrate the government's hotel.	Chinese embassy rent to celebration government hotel.

7.	Sme v situácii, kedy musíme robiť veci spoločne.	We are in a situation where we do things together.	We are in a situation where we have to do things together.	We're in a situation, when we have to do things together.	We are in a situation where we have to do things together.	Be sitionary, when we have to do thing together.
8.	Náš projekt je zdravý a funkčný.	Our project is a healthy and functional.	Our project is healthy and functional.	Our project is a healthy and to the functional.	Our project is healthy and functional.	Our project be in good health and functional.
9.	Členovia eurozóny sú v procese ratifikácie návrhov.	Members of the Eurozone are in the process of ratification of the draft.	Members of the euro area are in the process of ratification of the proposals.	Euro area members are the process of ratification proposals.	The members of the euro zone are in the process of ratification of the proposals.	Member euro area are they in the process of ratification proposals.
10.	Tento týžden prišlo k obnoveniu optimizmu.	This week there has been renewed optimism.	This week came to renewed optimism.	This week there was to restore but.	This week there has been a renewed optimism.	This weekly come by renovation optimism.
11.	Tie pomohli k posilneniu investičného prostredia s akciovými trhmi.	They help to strengthen the investment climate in equity markets.	You have helped to strengthen the investment environment with stock markets.	You have helped to strengthen investment environment the stock markets.	You have helped to strengthen the investment environment in the stock market.	Those have helped to reinforcement invested settings with stock market.
12.	Spoločnosť sa obmedzila na výrobu čiernobielych e-čítačiek.	The company is limited to the production of black and white e-readers.	The company is limited to the production of black and white e-readers.	The company has limited itself to the production black and white e-readers.	The company has limited itself to the production of black and white e-čteček.	Society oneself restrict to production black and white e-scaler.
Výsledok prekladu		55 / 60 => 92 %	51 / 60 => 85 %	45 / 60 => 75 %	45 / 60 => 75 %	10 / 60 => 17 %

Tabuľka A.2: Preklad slovenských viet.

5.2 Príloha B - Preklad paragrafov

B.1. Vysvetlenie označenia v prílohách

Pri tabuľkách, ktoré sa vyskytnú v nasledujúcich kapitolách, budú použité symboly. Ich význam je potrebný pre správnu orientáciu vo výsledkov.

Zoznam symbolov:

* - zadaný český výraz rovnakého významu

B.2. Z angličtiny do slovenčiny a češtiny

Sú dané tri ukážky paragrafov v anglickom jazyku, ktoré je potrebné preložiť. Tabuľka bola rozdelená do dvoch častí – slovenský a český preklad. Jej posledný riadok obsahuje percento úspešnosti prekladu jednotlivého paragrafu a aritmetický priemer týchto prekladov.

č.	Anglický paragraf	Slovenský preklad		
		Google Translate	Bing Translate	PC Translator 2010
1.	If you are happy to be contacted by a BBC journalist please leave a telephone number that we can contact you on. In some cases a selection of your comments will be published, displaying your name as you provide it and location, unless you state otherwise. Your contact details will never be published. When sending us pictures, video or eyewitness accounts at no time should you endanger yourself or others, take any unnecessary risks or infringe any laws. Please ensure you have read the terms and conditions.	Ak ste radi, že kontaktoval novinára BBC, prosím zanechať telefónne číslo, aby sme vás mohli kontaktovať na . V niektorých prípadoch bude výber Vaše pripomienky uverejnené , zobrazí vaše meno, ktoré poskytujete, a miesto, pokiaľ neustanovuje inak. Vaše kontaktné údaje nebudú nikdy zverejnené . Pri odosielaní nás obrázkov, videa alebo očitých svedkov nikdy by ste ohrozili seba alebo iných, prijat akékoľvek zbytočné riziko alebo porušuje zákony . Prosím, aby ste si prečítali podmienky .	Ak máte radi dotknúť BBC novinár zanechajte telefónne číslo, ktoré sme vás kontaktovali na . V niektorých prípadoch sa uverejní výber vaše komentáre, ktoré zobrazuje vaše meno ako poskytnete a umiestnení, pokiaľ ste neurčia inak. Vaše kontaktné údaje nebudú nikdy uverejnené . Pri odosielaní nás obrázky, video alebo očitá svedok účtov v žiadnom okamihu by ste ohrozili sami seba alebo iné osoby, prijat akékoľvek nepotrebným rizikám alebo porušujú zákony . Uistite, ste si prečítali podmienky .	Ak vy ste šťastný , že by ste bol kontaktovaný BBC novinárom prosím zanechajte telefónne číslo ktoré my môžeme kontaktovať vy ďalej . Prípadne výber vašich komentárov bude zverejnený, zobrazovanie vaše meno ako vy poskytujete to a pamätové miesto, pokiaľ vy udáte cenu inak . Tvoj kontaktné detaily nikdy nebude publikované. Keď zaslanie nás obrázy, video alebo svedecké záznamy nikdy mal by si ohroziť seba alebo iný, zobrat nejaké nepotrebné riziká alebo porušovať nejaké práva . Prosím vás zabezpečte, že vy čítali podmienky .
2.	At the age of 17 he had got his parents' permission to travel to Yemen to study Arabic. He briefly returned to California but couldn't settle so he headed back to Yemen from where he wrote to ask his father if he could go to Pakistan to continue his studies. Frank Lindh replied: "I trust your judgment and hope you have a wonderful adven-	V 17 rokoch dostal súhlas rodičov k ceste do Jemenu študovať arabčinu . On stručne sa vrátil do Kalifornie, ale nebola schopná vysporiadať , takže zamieril späť do Jemenu, odkiaľ písal sa spýtať otca , či by mo-	Vo veku 17 mal dostal jeho rodičia dovolenia cestovať do Jemenu študovať arabčina . Nakrátko vrátil do Kalifornie ale nemohol vyriešiť tak viedol späť na Jemen od kde napísal požiadať svojho otca, ak mu mohol ísť	Vo veku 17 on mal jeho rodičovské oprávnenie cestovať do Jemenu študovať arabčinu. Vrátil sa do krátko Kalifornia ale nemôcť účtovať tak on vrátil sa späť do Jemenu odkiaľ on písal žiada jeho otca či on môže skočiť na

	ture."	hol ísť do Pakistanu, aby pokračoval vo svojom štúdiu. Frank Lindh odpovedal: "Dúfam, že vaše rozhodnutie a dúfame, že budete mať nádherné dobrodružstvo."	do Pakistanu pokračovať v jeho štúdie. Frank Lindh odpovedal: "Som veriť vášho rozsudku a dúfam, že ste krásnym dobrodružstvom."	Pakistan pokračovať jeho štúdie. Bez porta Lindh odpovedal: "Ja dôverujem vášmu úsudok a nádej, ktorú máte zázračné dobrodružstvo."
3.	China is investing billions of dollars in its space programme. It has a strong space science effort under way, with two orbiting satellites having already been launched to the Moon. A third mission is expected to put a rover on the lunar surface. The Asian country is also deploying its own satellite-navigation system known as BeiDou, or Compass.	Čína investuje miliardy dolárov do svojho vesmírneho programu. Má silné kozmickéj vedy úsilie na ceste, s dvoma satelitov na obežnej dráhe, ktoré už začali na Mesiace. Tretí misie sa očakáva, aby rover na mesačnom povrchu. Ázijské krajiny sa tiež nasadením vlastný satelitný navigačný systém, známy ako Beidou, alebo Compass.	Čína je investujú miliardy dolárov do svojho priestoru programu. Má silné vesmírna veda úsilia prebiehajú, s dvoma obiehajúceho satelitov s už začala na mesiac. Očakáva sa, že tretia misia dať rover na povrchu. Ázijské krajiny je tiež nasadenie svoj vlastný satelitný navigačný systém, známy ako BeiDou alebo kompas.	Čína investuje miliardy dolárov v ich priestorový program. To má silnú priestorovú vedeckú snahu v prúde, s dvoma obiehajúci satelity už bývajú vypustený Moon. Tretina rozkaz má opčný predaj auto - rover na slabom povrchu. Ázijská krajina je taktiež nasadenie svoj vlastný satelitný- navigačný systém známy ako BeiDou, alebo Compass.
	Priemer výsledku prekladu [jednotlivé výsledky]	86 % [84 %, 84 %, 91 %]	56 % [54 %, 54 %, 59 %]	50 % [50 %, 71 %, 28 %]

Tabuľka B.1: Preklad anglických paragrafov do slovenského jazyka.

č.	Anglický paragraf	Český preklad		
		Google Translate	Bing Translate	World Lingo
1.	If you are happy to be contacted by a BBC journalist please leave a telephone number that we can contact you on. In some cases a selection of your comments will be published, displaying your name as you provide it and location, unless you state otherwise. Your contact details will never be published. When sending us pictures, video or eyewitness accounts at no time should you endanger yourself or others, take any unnecessary risks or infringe any laws. Please ensure you have read the terms and conditions.	Pokud ste rádi, že kontaktoval novináře BBC, prosím zanechat telefonní číslo, abychom vás mohli kontaktovat- na . V některých případech bude výběr Vaše připomínky zveřejněny, zobrazí vaše jméno, které poskytnete, a místo, pokud nestanoví jinak. Vaše kontaktní údaje nebudou nikdy zveřejněny . Při odesílání nás obrázků, videa nebo očitých svědků nikdy byste ohrozit sebe nebo ostatní, přijmout jakákoli zbytečná rizika	Pokud máte rádi kontaktovat novináře BBC prosím zanechte telefonního čísla, abychom vás mohli kontaktovat. V některých případech výběr vaše připomínky bude zveřejněno, zobrazení vaše jméno jako poskytnout místo a to pokud výslovně jinak. Vaše kontaktní údaje nebudou nikdy zveřejňují. Při odesílání nás obrázky, video nebo svědka účty v čas by ohrozit sebe nebo ostatní, podstupovat zbytečná rizika nebo porušují žádné zákony.	Pokud ste spokojeni se být kontaktován BBC novinář nechte prosím telefonní číslo, které můžeme kontaktovat. V některých případech výběr Vaše připomínky budou zveřejněny, zobrazení své jméno v poskytovat a umístění, pokud se stát jinak. Vaše kontaktní údaje nikdy nebudou zveřejněno . Když nás posílá obrázky, video nebo očitých svědků v žádném okamžiku by ohrožuješ sebe nebo ostatní, učinit zbytečným rizikům nebo

		nebo porušuje zákony. Prosím, aby jste si přečetli podmínky.	Ujistěte se, že jste si přečetli a podmínky.	porušení právních předpisů. Zkontrolujte si přečteš podmínky.
2.	At the age of 17 he had got his parents' permission to travel to Yemen to study Arabic. He briefly returned to California but couldn't settle so he headed back to Yemen from where he wrote to ask his father if he could go to Pakistan to continue his studies. Frank Lindh replied: "I trust your judgment and hope you have a wonderful adventure."	V 17 letech dostal svolení rodičů k cestě do Jemenu studovat arabštinu. On stručně se vrátil do Kalifornie, ale nebyla schopna vypořádat, takže zamířil zpět do Jemenu, odkud psal se zeptat otce, zda by mohl jít do Pákistánu, aby pokračoval ve svém studiu. Frank Lindh odpověděl: "Doufám, že vaše rozhodnutí a doufáme, že budete mít nádherné dobrodružství."	Ve věku 17 let dostal jeho rodiče povolení k vycestování do Jemenu studovat arabsky. Krátce se vrátil do Kalifornie ale nemohla vyrovnat tak zamířil zpět do Jemenu z napsal, kde se zeptat svého otce, kdyby mohl jít do Pákistánu pokračovat ve studiu. Frank Lindh odpověděl: "Váš úsudek a doufám, že máte úžasné dobrodružství."	Ve věku 17 měl jeho rodiče povolení k vycestování do Jemenu studovat arabsky. Krátce vrátil do Kalifornie ale nemohl usadit tak zamířil zpět k Jemenu odkud napsal zeptat na jeho otce, kdybych mohl jít do Pákistánu pokračovat ve studiu. Frank Lindh odpověděl: "Věřím tvému úsudku a doufat, že máte úžasné dobrodružství."
3.	China is investing billions of dollars in its space programme. It has a strong space science effort under way, with two orbiting satellites having already been launched to the Moon. A third mission is expected to put a rover on the lunar surface. The Asian country is also deploying its own satellite-navigation system known as BeiDou, or Compass.	Čína investuje miliardy dolarů do svého vesmírného programu. Má silné kosmické vědy úsilí na cestě, se dvěma satelitů na oběžné dráze, které již byly zahájeny na Měsíc. Třetí mise se očekává, aby rover na měsíčním povrchu. Asijské země se také nasazením vlastní satelitní navigační systém, známý jako Beidou, nebo Compass.	Čína investuje miliardy dolarů v jeho kosmického programu. Má silnou space science úsilí na cestě, s dvěma orbitální satelity, již byla zahájena na měsíci. Třetí mise se očekává rover na měsíční povrch. Asijské země je také nasazení své vlastní satelitní navigační systém, známý jako BeiDou nebo kompas.	Čína je investovat miliardy dolarů v jeho kosmického programu. Má silné vesmírné vědě úsilí na cestě, se dvěma obíhající družice s již byla zahájena na měsíc. Třetí mise se očekává, že dal Rover na měsíčním povrchem. Asijské země rovněž umísťovat vlastní satelitní navigační systém zvaný beidou nebo kompas.
Prieemer výsledku prekladu [jednotlivé výsledky]		85 % [81 %, 85 %, 89 %]	72 % [63 %, 80 %, 73 %]	63 % [44 %, 89 %, 56 %]

Tabuľka B.2: Preklad anglických paragrafov do českého jazyka.

B.3. Zo slovenčiny a češtiny do angličtiny

Spôsob použitý pri vetách sa teraz aplikuje na časti článku. Vytvorený korektný preklad je použitý ako počiatočná báza. Na preložený odsek sa aplikujú pravidlá a dosiahnuté výsledky sa zaznačia do tabuľky.

č.	Slovenský paragraf	Anglický preklad				
		Google Translate	Bing Translate	World Lingo *	PC Translator 2010	MOSES Translate *
1.	Ak by ste radi kontaktovali	If you would like to contact the BBC	If you would like to contact	If you would like contacted	IF you'd radial road contac-	If you would like to have

	<p>BBC novinára, prosím zanechajte telefónne číslo, aby Vás mohol kontaktovať. V niektorých prípadoch môžu byť Vaše komentáre zverejnené, zobrazí sa meno, ktoré ste poskytli, a miesto, pokiaľ neuviedete inak. Vaše kontaktné údaje nebudú nikdy zverejnené. Poslaním obrázkov, videa alebo účtov očitých svedkov môžete ohroziť seba alebo ostatných a podstupujete nežiaduce riziko ani neporušujete zákon. Prosím, aby ste si prečítali podmienky.</p>	<p>journalist, please leave a phone number so that you can contact. In some cases your comments may be published, the name you provided, and place, unless you indicate otherwise. Your contact details will never be published. The mission of images, video or eyewitness accounts, you could endanger yourself or others, and run the risk of adverse or break the law. I ask that you read the conditions.</p>	<p>a BBC journalist, please leave a phone number that You could contact you. In some cases may be your comments published, you will see the name that you provided, unless you specify otherwise, and location. Your contact details will never be disclosed. The Mission of the images, video, or you can endanger themselves or other witness accounts of the očitých and podstupujete adverse risk nor the neporušujete law. Please read the terms and conditions.</p>	<p>the BBC journalist, please leave your phone number, to you to contact. In some cases, they may be your comments published, displays the name, which you provided, and the place, where do otherwise.Your contact data shall not be never published. The mission pictures, video or accounts eyewitness accounts you can endanger himself or others, and you're taking undesirable risk or smart law. Please, you to read conditions.</p>	<p>ted BBC journalist, please resignation telephone number, that you has been allowed contact. In some cases be able to Your comments made public, will portray oneself the name of, who are offer, and place, as far as nondisclosure otherwise. Your contact poop will not never made public. Sent images, videos or accounts ocular witness may endanger myself or of others and substage objectional risk nor' intact law. Please, so that you yourself read - outs conditions.</p>	<p>contacted the BBC journalist, please leave a telephone number, so that you can contact. In some cases, your comments may be published, shows the name that you have provided, and the place where the options otherwise. Your contact details will never be made public. The mission of the pictures, videos or accounts eyewitnesses can endanger yourself or others, and you run the risk not breaking the law or undesirable. Please, you read our terms.</p>
2.	<p>V 17 rokoch dostal súhlas rodičov k ceste do Jemenu študovať arabčinu. Nakrátko sa vrátil do Kalifornie, ale nebol schopný usídlit' sa, takže zamieril späť do Jemenu, odkiaľ</p>	<p>In 17 years, received parental permission to travel to Yemen to study Arabic. Briefly returned to California, but was unable to settle, so headed back to Yemen, where he asked his father if he could go to</p>	<p>In the 17 years he received the consent of the parents on the road to Yemen to study Arabic. Briefly he returned to California, but was not ca-</p>	<p>In 17 years he got consent of the parents to the road to Yemen to study Arabic. For a short while he went back to California, but he was not able</p>	<p>Volts 17 ages got parental consent to ways to Jemenu study Arabic. Briefly oneself handed back to in Kalifornie, but he was unable settle, accor-</p>	<p>In the 17 years he got parental consent to the trip to Yemen to study Arabic. Briefly returned to California, but was unable to settle elsew-</p>

	<p>sa spýtal otca, či by mohol ísť do Pakistanu, aby pokračoval vo svojom štúdiu. Frank Lindh odpovedal: "Verím v tvoje rozhodnutie a dúfam, že zazi- ješ nejaké dob- rodružstvo."</p>	<p>Pakistan to conti- nue their studies. Frank Lindh said: "I believe in your decision and hope that you'll expe- rience an adventu- re."</p>	<p>pable of ta- king up resi- dency, headed back to Ye- men, from where he asked his fat- her if he could go to Pakistan to continue their study. Frank Lindh replied: "I believe in your decision and I hope that zaziješ some adven- ture."</p>	<p>to settle, so he went back to Yemen, where he asked his fat- her, whether he could go to Pakistan, to continue the studio. Frank 8, and said: "I believe in your decision, and I hope I find some adventure."</p>	<p>dingly headed back to Jeme- nu, from whe- re oneself questioningly father, or might be go- ing to Pakista- nu, that he kept on at of his studio. Franconian Lindh replied: "Believe in your decision and hope that the experien- ce some ad- venture."</p>	<p>here, so he headed back to Yemen, where he said his father, whether he could go into Pakistan to continued his studies. Frank Lindh said: "I believe in your decision and I hope that desire Being an ad- venture."</p>
<p>3.</p>	<p>Čína investuje miliardy dolárov do svojho ves- mírneho prog- ramu. Silné úsilie kozmickej vedy je v plnom prúde, dva sate- lity na obežnej dráhe už vyštar- tovali na Me- siac. Od tretej misie sa očaká- va auto na me- sačnom po- vrchu. Ázijské krajiny tiež nasadzujú vlast- ný satelitný navigačný sys- tém, známy ako Beidou, alebo Compass.</p>	<p>China is investing billions of dollars into its space prog- ram. Strong efforts space science is in full swing, two satellites already in orbit to the moon vyštartovali. From the third car of the mission is expected on the lunar surface. Asian countries also deploy its own satellite navigation system known as Beidou, or Compass.</p>	<p>People's Re- public of Chi- na is investing billions of dollars into its space prog- ramme. The strong efforts of the interna- tional space science is in full swing, two satellites in orbit already vyštartovali for a month. From the third mission to the lunar surface is expected to be a car. Asian countries also are rolled out its own satelli- te navigation system, known as Beidou, or Compass.</p>	<p>China invests billions of dollars into the space program. Strong efforts space science is in full swing, two satellites in orbit for I to the moon. From the third mission, the expected the car in the lunar surface. Asian coun- tries also dep- loys its own satellite navi- gation system, known as beidou, or compass.</p>	<p>China invests zillion dollars in your cosmic program. Strong effort cosmic scien- ce be under- way, two sate- llites on circulating road for be off on moon. Away from the the third station one- self expects car on monthly surface. Asiatic lan- dscapes too risk own satel- litic nautical the system, known as Beidou, or Compass.</p>	<p>China is inves- ting billions of dollars into its space prog- ram. Strong efforts of spa- ce science is in full swing, two satellites in orbit longer took off for the moon. The third mission is expected from a car on the surface of the moon. Asian coun- tries also dep- loys its own satellite navi- gation system, known as Beidou, or Compass.</p>

Priemer výsledku prekladu [jednotlivé výsledky]	84 % [87 % , 93 % , 72 %]	68 % [64 % , 86 % , 55 %]	65 % [42 % , 86 % , 67 %]	7 % [0 % , 21 % , 0 %]	87 % [81 % , 88 % , 93 %]
--	----------------------------------	----------------------------------	----------------------------------	-------------------------------	----------------------------------

Tabuľka B.3: Preklad slovenský paragrafov do anglického jazyka.