

# TÍM č. 4, aDictIT

## Členovia tímu (študenti):

Bc. Róbert Horváth,  
Bc. Peter Jurčík,  
Bc. Peter Macko,  
Bc. Vladimír Ruman,  
Bc. Peter Sládeček,  
Bc. Maroš Ubreži,  
Bc. Matúš Vacula

## Ved. tímu (pedagóg): Ing. Dušan Zeleník

Motto tímu: Audaces fortuna adiuvat

## Názov projektu: Štatistický preklad voľného textu

### O ČOM JE NÁŠ PROJEKT?

Preklad textu z jedného jazyka do iného je netriviálny problém, ktorý sa ľudia snažia automatizovať už niekoľko desaťročí. Prvé pokusy o automatizáciu procesu prekladu sa datujú už od obdobia druhej svetovej vojny. Aj napriek tomu, že sa táto problematika skúma už dlhší čas, dnešné prekladače nie sú ani zďaleka dokonalé. Ich preklady sú často nepresné a z dlhších viet často vytvoria nezmyselné skupiny slov.

Väčšina z týchto prekladačov pritom funguje na princípe paralelného korpusu, ktorého získanie je však príliš zložitá a preto väčšina menších prekladačov tohto typu nedosahuje pri prekladoch dostatočne dobré výsledky.

Veľké množstvo projektov na našej fakulte, ale aj vo svete, využíva možnosti automatických prekladačov ako Google Translate. Ten však v nedávnom čase spoplatnil svoje rozhranie pre vývojárov a tak neprijemne ovplyvnil veľa nádejných projektov.

Nami navrhnutý prekladač sa od ostatných odlišuje absenciou obojstranného korpusu. V našom riešení sa v prvej fáze všetky slová zadaného textu preložia po slovách. Následne sa z týchto prekladov vygeneruje množina prípustných viet. Táto množina je tvorená kombináciami možných prekladov slov z vety. Jednotlivé vety je nutné následne vyfiltrovať tak, aby sa z kombinácií odstránili vety, ktoré majú minimálnu šancu na to, aby sa stali požadovaným prekladom zadanej vety.

Ďalším krokom je vyhľadanie zadaných formulácií v korpuse viet cieľového jazyka. Podľa toho, na koľko sa jednotlivé výsledky približujú vetám v korpuse, sú výsledky vyhodnotené a zoradené. Používateľovi sa teda zobrazia iba tie výsledky, ktoré sú najpravdepodobnejším prekladom ním zadanej vety.

Výhodou nášho riešenia je schopnosť s minimálnym úsilím prekladať texty medzi ľubovoľnými jazykmi. Podporu ďalšieho jazyka je možné zabezpečiť pridaním slovníka a korpusu viet v danom jazyku.

### ČO NÁM DÁVA PRÁCA NA TOMTO PROJEKTE?

Keďže v prípade tohto projektu ide o tímové riešenie, ktorého rozsah sú dva semestre, prináša nám práca na ňom množstvo príležitostí. Hlavným prínosom je práca v tíme. Aj keď sme sa počas štúdia stretli s prácou v dvojiciach či trojiciach, v takom rozmere v akom je tímový projekt realizovaný, sme spoločné riešenie ešte nevytvárali. Tímová práca je pre náš

ďalší profesionálny život veľmi dôležitá, keďže sa s ňou IT špecialisti stretávajú azda v každej oblasti tohto odvetvia.

Ďalšou oblasťou riešenia, ktorá je priamo spojená s naším riešením, je realizácia projektu v prostredí webu. Naš prekladač pracuje ako samostatná webová služba, ktorá má dosť vysoké nároky na výkon servera. Práve preto je pre nás dobré pracovať s webovými technológiami založenými na jazyku Java a takisto s aplikačným serverom Apache Tomcat. Tieto a im podobné technológie dovoľujúce príchod pojmov, ako sú Web 2 či 3.0, sú veľmi perspektívne a je vysoko pravdepodobné, že prácu na nich budeme v našom profesionálnom živote vedieť zúročiť.

Veľkým prínosom pre nás je aj práca na nových technológiách ako je mladý projekt ElasticSearch. Táto technológia je nová a tak je práca s ňou samé prekvapenie. Prináša nám množstvo výhod, ale takisto prináša aj nevýhody, ktoré môžu nemilo prekvapiť.

Celkovo z tímového projektu a nášho riešenia získavame obrovské množstvo skúseností a pozitívnej energie. Tieto skúsenosti nám pri našom profesionálnom raste určite pomôžu a dovoľia nám preniknúť do IT priemyslu ešte hlbšie.

### **PREČO JE NAŠ PROJEKT ZAUJÍMAVÝ?**

Naš projekt je zaujímavý hlavne spôsobom prekladu, ktorý sa snažíme uplatniť v prekladači. Konvenčné prekladače využívajú obojstranný korpus, a teda majú rovnaký text v oboch prekladaných jazykoch. My na rozdiel od toho chceme prekladať na základe jednostranného korpusu a teda nie sme závislí na doslovne preložených textoch. Takýmto spôsobom teda získavame oveľa jednoduchšiu rozšíriteľnosť prekladača o nové jazyky. Táto výhoda spolu s faktom, že Google spoplatnil svoje prekladacie služby, by nám mala pomôcť pri spopularizovaní a následne nasadení nášho prekladača do reálnej prevádzky.

### **POUŽITÉ TECHNOLOGIE:**

Java, Apache Tomcat, MySQL, ElasticSearch, HTML, JavaScript

### **O ČOM TO VLASTNE JE?**

Naš projekt je zameraný na vývoj prekladača, ktorý bude na základe štatistiky vedieť prekladať texty pomocou jednostranného korpusu na strane cieľového jazyka.