

# TÍM č. 3, Plagiátor Wanted

**Členovia tímu (študenti): Bc. Jozef Lačný, Bc. Maroš Maršalek, Bc. Pavel Michalko, Bc. Ján Súkeník**

**Ved. tímu (pedagóg): doc. Mgr. Daniela Chudá, PhD.**

**Názov projektu: Plagiáty na webe**

## **O ČOM JE NÁŠ PROJEKT?**

Na webe sa v súčasnosti nachádza veľké množstvo informácií. Prístup k týmto informáciám je jednoduchý, rýchly a takmer nikdy nie je spoplatnený. Okrem mnohých pozitívnych následkov tejto dostupnosti sa však objavujú aj niektoré negatívne. Jedným z nich je rýchle rozširovanie problémov súvisiacich s plagiátorstvom. Jeho odhaľovanie dnes skoro vždy vykonávajú iba jednotlivci a robia to takmer výlučne manuálne.

Porovnávanie textov ľuďmi na prvý pohľad pôsobí dôveryhodne, trpí však mnohými nedostatkami – napríklad rýchlosť, subjektivnosť, či nepozornosť. Ďalším problémom je fakt, že ani odborník na konkrétnu oblasť nemôže detailne poznať všetku dostupnú literatúru a výskum, čím sa výrazne znižuje pravdepodobnosť odhalenia plagiátu. Tento problém sa stáva ešte výraznejším v prostredí webu, z dôvodu veľkého počtu dostupných textov. Je preto samozrejmé, že musíme celý proces v čo najväčšej miere automatizovať. Automatické odhaľovanie plagiátorstva však predstavuje komplexnú úlohu, ktorá pozostáva z rôznych častí – spracovanie dokumentov, vyhľadávanie relevantných zdrojov, analýza podobnosti textov, vizualizácia výsledkov, atď. Tie so sebou prinášajú množstvo ďalších problémov, ktoré je potrebné vyriešiť. Automatické metódy však napriek svojim stále existujúcim nedostatkom vedia rýchlo, objektívne a merateľne vyhodnotiť úroveň originality veľkého množstva textov. Ľudský úsudok stále nemôžeme z procesu vylúčiť úplne, ale automatické metódu môžu slúžiť ako pomocný nástroj a vďaka tomu výrazne uľahčiť a spresniť identifikáciu plagiátov. Našou odpoveďou na uvedený problém je aplikácia, ktorá umožňuje jednoduchým spôsobom overiť originalitu ľubovoľného textu vzhľadom na texty, resp. dokumenty dostupné na webe. Keďže nemáme prostriedky na to, aby sme si index webových dokumentov vytvorili podľa našich potrieb, rozhodli sme sa hľadať plagiáty s využitím už existujúcich webových vyhľadávačov. Na zvýšenie presnosti celého procesu sme navrhli viacero algoritmov (napr. na zostavovanie vyhľadávacích fráz), ktoré aktuálne testujeme v prototypu našej aplikácie. Dosahované výsledky spĺňajú naše očakávania a preto veríme, že si aplikácia nájde praktické využitie pri odhaľovaní plagiátov v akademickej ale aj komerčnej sfére.

## **ČO NÁM DÁVA PRÁCA NA TOMTO PROJEKTE?**

Práca na tomto projekte je pre nás jedinečnou príležitosťou oboznámiť sa s vývojom relatívne veľkého softvéru v tíme. Tímová práca je v súčasnosti neodmysliteľnou časťou väčšiny softvérových projektov. Takto si vyskúšame základné princípy komunikácie, rôzne podporné technológie alebo samotné procesy pri práci v distribuovaných tímoch. Považujeme to teda za výborný základ do budúceho zamestnania.

Napriek tomu, že tento projekt je riešený v akademickom prostredí, pristupujeme k nemu ako k reálnemu komerčnému projektu pre zákazníka. Dúfame, že tento prístup nám poskytne skúsenosti, ktoré budeme môcť využiť v budúcnosti. K prehĺbovaniu týchto skúseností prispieva aj používanie technológií a nástrojov, ktoré sa v súčasnosti v praxi používajú tiež. V dnešných dňoch sa do popredia dostávajú agilné prístupy k vývoju softvéru, preto sme si aj my zvolili tento prístup a prispôbili ho našim potrebám. Každý člen tímu má svoje

špecifické zodpovednosti, od podporných činností cez vývoj používateľského rozhrania až po implementáciu webových služieb. Úlohy v týchto oblastiach sú však rozdelené rovnomerne medzi všetkých členov tímu. Toto rozdelenie úloh je jedným z dôsledkov použitia agilných techník vývoja softvéru – jednotlivé úlohy plníme na základe ich priority a vyťažnosti tímu, špecializácia jednotlivých členov je až na druhom mieste. Tento prístup nám zabezpečil, že každý člen tímu sa stretol s vývojom všetkých častí aplikácie a tak nikto nie je úplne nenahraditeľným, čo by mohlo mať nepríjemné následky napríklad v prípade dlhšej práceneschopnosti. Toto rozdelenie navyše umožní každému členovi získať skúsenosti v rôznych oblastiach vývoja softvérových projektov.

Za veľký prínos projektu považujeme aj možnosť oboznámiť sa s rôznymi časťami riadenia v softvérových projektoch, čo je oblasť, s ktorou sa v praxi zo začiatku zrejme moc stretávať nebudeme. Či už ide o plánovanie a rozdeľovanie úloh, manažment rizík, dokumentácie, kvality, atď., všetky sú dôležité a dufame, že práca s nimi nám poskytne množstvo užitočných skúseností. Dostatočné pokrytie týchto oblastí totiž do veľkej miery ovplyvňuje kvalitu a výsledok softvérového projektu.

### **PREČO JE NÁŠ PROJEKT ZAUJÍMAVÝ?**

Vytváranie nových prác by malo v čo najväčšej miere vychádzať z poznatkov, skúseností a myšlienok ich autorov. Zámerné alebo neúmyselné odkopírovanie cudzej práce a jej následné prezentovanie ako vlastnej nie je dnes ničím výnimočným. Dôvody môžu byť rozličné, avšak výsledok je vždy rovnaký – znižovanie kvality vzdelávania a výskumu. Využívanie existujúcich znalostí je v procese objavovania nových poznatkov nevyhnutné. Prevzaté časti však musia byť korektne označené, aby mohol byť identifikovaný prínos autora. Preto je našim cieľom vytvoriť aplikáciu, ktorá bude odhaľovanie plagiátov uľahčovať a nútiť tým autorov korektne citovať všetky použité zdroje.

### **POUŽITÉ TECHNOLOGIE:**

Ruby on Rails, HTML 5, JavaScript

### **O ČOM TO VLASTNE JE?**

Vytvárame aplikáciu, pomocou ktorej sa bude dať jednoducho overiť originalita ľubovoľného textu vzhľadom na všetky dokumenty dostupné na webe. Využívanie práce iných autorov je podľa nás veľmi dôležité a nechceme mu nijako zabraňovať. Naším cieľom je ale motivovať autorov aby všetky použité zdroje korektne citovali a aj tým pomáhali procesu objavovania nových poznatkov.