

# Posudek oponenta bakalářské práce

Autor práce: **Jan Ulrych**

Název práce: **Systém pro tvorbu automatického generátoru zpráv ze strukturovaných dat**

## Obsah a cíle práce

Tématem práce Jana Ulrycha bylo navrhnout a implementovat systém, který přívětivým způsobem umožní generování novinových zpráv na základě definovaných šablon. Důraz je kladen zejména na burzovní zprávy, protože práce úzce souvisí s výstupy projektu TAČR-TL02000288. Motivací je urychlení a zefektivnění práce novinářů.

Typicky se ve zprávách a titulcích z burzovního zpravodajství objevují informace o růstu či poklesu akciového indexu, případně společnosti, které zaznamenaly největší růst či naopak pokles. Toto jsou informace, které se snadno zpracovávají a lze je jednoduše použít jako základ pro generování burzovního zpravodajství přes šablony. **Generování pomocí šablonovacího systému proto z mého pohledu dává smysl** a může být pro tuto oblast přínosné.

## Vyvinutý program a text práce

Celý systém je dle mého názoru dobře navržen a dekomponován, definice jednotlivých objektů a použitá NoSQL databáze MongoDB dává smysl. Architektura model-view-controller (MVC) a framework Spring umožňující Dependency Injection je rovněž volba, proti které nelze nic namítat. Frontend je řešen ve frameworku Angular. Celý systém běží v Dockeru, což oceňuji. Při spuštění jsem nenarazil na žádné problémy a navíc díky Dockeru jsem nic nemusel instalovat.

Text práce v první části popisuje problematiku generování zpráv, včetně poznatků od novinářů a požadavků na systém. Zbytek práce je důkladný popis návrhu a implementace systému, včetně instalace a testování. Text je doprovázen vhodně zvolenými diagramy a obrázky a této části téměř nelze nic vytknout.

V textu jsem nenalezl téměř žádné překlepy či chyby. Kvalita textu je na velice dobré úrovni. Rozsah práce je 108 stran, ale podstatnou část tvoří přílohy, popisující uživatelskou dokumentaci a popis vytvoření šablony pro burzovní zpravodajství. Reference obsahují 11 položek, což považuji vzhledem k povaze práce za adekvátní. U online referencí nicméně **chybí datum citace**, což by měl být povinný údaj.

## Testování

Jediný způsob (aspoň dle mé zkušenosti), jak do systému nahrát nová data je prostřednictvím skriptu v Pythonu. Tzn. nelze jednoduchým způsobem přes nějaké tlačítko nahrát nový CSV soubor, definovat šablonu a vygenerovat zprávu. Po spuštění služby je k dispozici i návod (tutorial), díky kterému lze snadno získat nápovědu.

System jsem testoval následujícím způsobem. Vytvořil jsem si v Pythonu skript, který vygeneruje volební model (tzn. politická strana a její preference v procentech). Cílem bylo automaticky vygenerovat zprávu následujícího znění:

„Nejlepšího výsledku dosáhla politická strana {{strana}} se ziskem {{procenta}}%. Do sněmovny by neprošly následující strany: {{strany\_pod\_pet\_procent}}“

nebo

„Nejlepšího výsledku dosáhla politická strana {{strana}} se ziskem {{procenta}}%. Před branami sněmovny by zůstaly: {{strany\_pod\_pet\_procent}}“

Dvě verze zpráv jsem úspěšně vygeneroval pomocí randomizéru. Přitom jsem objevil, že lze zadat pravděpodobnost větší než 1. Navíc se nekontroluje, zda všechny možnosti randomizéru sčítají do jedné. Poté jsem si prošel i předem připravená scénáře (tj. data a šablony). Vše fungovalo dle očekávání a bez problémů.

### **Hodnocení**

Práci považuji za velice povedenou, chybí mi pouze (např. v příloze) srovnání automaticky vygenerovaného textu s textem, který je napsán manuálně člověkem (novinářem). Tuto informaci jsem si ale poté dohledal sám přes výstupy burzovního robota na webu e15.cz.

Zadání práce bylo dle mého názoru splněno ve všech bodech. K práci mám doplňující dotazy (viz níže). Navzdory výše uvedeným připomínkám navrhuji hodnocení známkou **výborně** a práci doporučuji k obhajobě.

**V Plzni 19.5.2023**

**Ing. Jiří Martínek, Ph.D.**

### **Doplňující dotazy k práci**

- 1) Z jakého důvodu je nutné import dat řešit pouze skriptem v Pythonu? Proč není umožněno nahrání např. CSV souboru (pomocí tlačítka)?
- 2) Jak by bylo nutné systém rozšířit o možnost generovat i obrázky? Např. spojnicové grafy trendů? Případně pak generovat doprovodný text.
- 3) Existují kromě burzovního zpravodajství i další odvětví žurnalistiky (v rámci ČR), kde je poptávka po podobném řešení?