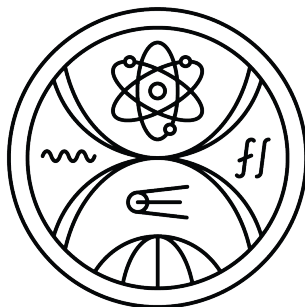


UNIVERZITA KOMENSKÉHO V BRATISLAVE  
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

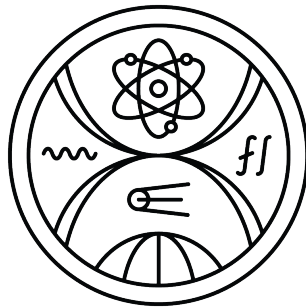


# DATASETY V SOFTVÉROVOM INŽINIERSTVE

Diplomová práca



UNIVERZITA KOMENSKÉHO V BRATISLAVE  
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



# DATASETY V SOFTVÉROVOM INŽINIERSTVE

Diplomová práca

Študijný program: Aplikovaná informatika  
Študijný odbor: Informatika  
Školiace pracovisko: Katedra aplikovanej informatiky  
Školiteľ: Ing. Lukáš Radoský





Univerzita Komenského v Bratislave  
Fakulta matematiky, fyziky a informatiky

## ZADANIE ZÁVEREČNEJ PRÁCE

**Meno a priezvisko študenta:** Bc. Jakub Murin  
**Študijný program:** aplikovaná informatika (Jednoodborové štúdium, magisterský II. st., denná forma)  
**Študijný odbor:** informatika  
**Typ záverečnej práce:** diplomová  
**Jazyk záverečnej práce:** slovenský  
**Sekundárny jazyk:** anglický

**Názov:** Datasets v softvérovom inžinierstve  
*Datasets in software engineering*

**Anotácia:** V softvérovom inžinierstve prebieha výskum efektívnych metód transformácie niektorých artefaktov vznikajúcich v rámci vývoja softvéru na iné artefakty vývoja softvéru. Napriek týmto snahám však vo výskumnej komunite často chýbajú vhodné datasety pre evaluáciu týchto metód. Metódy sú tak evaluované na netransparentných dátach v malých objemoch, ak vôbec sú evaluované. Rôzne metódy sú tak ťažko porovnateľné. Príkladom takejto úlohy je transformácia prípadov použitia na UML diagramy alebo zdrojový kód.

Zvoľte si úlohu softvérového inžinierstva a vhodným postupom vytvorte strojovo spracovateľný dataset, na ktorom bude možné rôzne metódy či prístupy riešiace danú úlohu evaluovať. Dataset adekvátne zdokumentujte a opíšte, vrátane štatistických ukazovateľov. Využite vytvorený dataset na evaluáciu niektorej z existujúcich metód na riešenie danej úlohy, prípadne využite vlastnú, hoci aj triviálnu metódu.

**Cieľ:** Vytvorenie datasetu alebo viacerých datasetov pre vybranú úlohu alebo úlohy softvérového inžinierstva

**Literatúra:** Frank Breiting, Alexandre Jotterand, Sharing datasets for digital forensic: A novel taxonomy and legal concerns, Forensic Science International: Digital Investigation, Volume 45, Supplement, 2023, 301562, ISSN 2666-2817.  
national Conference on Model Driven Engineering Languages and Systems: Companion Proceedings (MODELS '22). Association for Computing Machinery, New York, NY, USA, 396–403.

The Quest for Open Source Projects that Use UML: Mining GitHub; Hebig, R. & Ho-Quang, T. & Robles, G. & Fernandez, M.A. & Chaudron, M.R.V. (2016). In proceedings, ACM/IEEE 19th International Conference on Model Driven Engineering Languages and Systems, pages 173-183, Saint-Malo, France, October 2-7, 2016.

Alyami A, Pileggi SF, Sohaib O, Hawryszkiewicz I. 2023. Seamless transformation from use case to sequence diagrams. PeerJ Computer Science 9:e1444 <https://doi.org/10.7717/peerj-cs.1444>



Univerzita Komenského v Bratislave  
Fakulta matematiky, fyziky a informatiky

---

Song Yang and Houari Sahraoui. 2022. Towards automatically extracting UML class diagrams from natural language specifications. In Proceedings of the 25th Inter

**Kľúčové**

**slová:** Dataset, Budovanie datasetov, Softvérové inžinierstvo, Vývoj softvéru

**Vedúci:** Ing. Lukáš Radoský

**Katedra:** FMFI.KAI - Katedra aplikovanej informatiky

**Vedúci katedry:** doc. RNDr. Tatiana Jajcayová, PhD.

**Spôsob prístupnosti elektronickej verzie práce:**

bez obmedzenia

**Dátum zadania:** 08.10.2023

**Dátum schválenia:** 11.11.2023

prof. RNDr. Roman Ďurikovič, PhD.  
garant študijného programu

.....  
študent

.....  
vedúci práce

Čestne prehlasujem, že túto diplomovú prácu som vypracoval samostatne len s použitím uvedenej literatúry a za pomoci konzultácií u môjho školiteľa.

Bratislava, 2025

.....  
Bc. Jakub Murin





**Pod'akovanie:** Chcem sa pod'akovať ...

# Abstrakt

Bezpečie a ochrana sú neoddeliteľnou súčasťou našej existencie.

**Kľúčové slová:** Dataset, Budovanie datasetov, Softvérové inžinierstvo, Vývoj softvéru

# Abstract

Safety and security are an main part of our existence.

**Keywords:** Dataset, Dataset construction, Software engineering, Software development



# Obsah

<b>Úvod</b>	<b>1</b>
<b>1 Prípad použitia</b>	<b>3</b>
1.1 Dobrý prípad použitia . . . . .	3
1.1.1 Z ChatGPT . . . . .	3
<b>2 LaTeX</b>	<b>5</b>
2.1 Obrázky . . . . .	5
<b>Záver</b>	<b>7</b>



# Zoznam obrázkov

2.1 Ukážka hry Červík . . . . .	5
---------------------------------	---





# Zoznam ukážok kódu

2.1	Algoritmus na výpočet faktoriálu v jazyku C . . . . .	6
-----	---	---



# Úvod

Uvádzam túto prácu.



# Kapitola 1

## Prípad použitia

Dataset obsahuje aj prípady použitia. V tejto kapitole sa pozrieme na to, ako vyzerá dobrý prípad použitia a aké znaky môžeme sledovať.

### 1.1 Dobrý prípad použitia

Vo všeobecnosti platí, že chceme písať čo najmenej, čo najjasnejšie a ukázať spôsoby, akými systém reaguje na rôzne situácie. [3] *pg. 16*

**Správnosť > Presnosť**

Postup vytvárania:

1. Aktéri a ciele
2. Hlavné úspešné scenáre
3. Neúspešné prípady
4. Ošetrenie neúspešných prípadov

#### 1.1.1 Z ChatGPT

- **Orientácia na cieľ:** Use case by mal smerovať k dosiahnutiu konkrétnych cieľov alebo úloh pre používateľov systému. Mal by popisovať, čo chce používateľ dosiahnuť.
- **Narratívna forma:** Use case by mal byť napísaný v narratívnej forme, popisujúcej postupnosť interakcií medzi aktérom (používateľom) a systémom na dosiahnutie žiadanej úlohy.
- **Identifikácia aktérov:** Jednoznačne identifikujte aktérov zapojených v use case. Aktéri môžu byť používatelia, iné systémy alebo dokonca externé entity interagujúce so systémom.

- **Jasné a stručné:** Use case by mal byť jasný, stručný a zrozumiteľný pre všetkých zainteresovaných, vrátane používateľov, vývojárov a testerov.
- **Ne-technický jazyk:** Vyhnite sa technickému žargónu v use case, aby boli zrozumiteľné pre všetkých zainteresovaných, vrátane netechnických používateľov.
- **Obmedzenie rozsahu:** Use case by mal mať dobre definované hranice a zamerať sa na konkrétnu funkciu alebo vlastnosť systému.
- **Predpoklady a následky:** Jednoznačne stanovte akékoľvek predpoklady potrebné na spustenie use case a akékoľvek následky, ktoré musia platiť po dokončení use case.
- **Alternatívne cesty:** Popíšte alternatívne cesty alebo scenáre, ktoré používateľ môže v rámci use case použiť na dosiahnutie cieľa, vrátane výnimočných alebo chybových podmienok.
- **Vystopovateľnosť:** Use case by mal byť vystopovateľný do vyšších úrovňových požiadaviek a obchodných cieľov, poskytujúc jasný odkaz medzi cieľmi používateľov a funkčnosťou systému.

## Kapitola 2

# Ukážky užitočných príkazov v systéme LaTeX

V tejto kapitole si ukážeme príklady niektorých užitočných príkazov, ako napríklad správne používanie tabuliek a obrázkov, číslovanie matematických výrazov a podobne. Konkrétne príkazy použité v tejto kapitole nájdete v zdrojovom súbore `latex.tex`. Všimnite si, že pre potreby obsahu a hlavičky stránky je v zdrojovom súbore uvedený aj skrátený názov tejto kapitoly. Ďalšie užitočné príkazy nájdete aj v kapitole ??, na ktorú sme sa na tomto mieste odvolali príkazom `\ref`.

### 2.1 Obrázky

Vašu prácu ilustrujte vhodnými obrázkami. Pri použití programu `pdflatex` je potrebné pripraviť obrázky vo formáte `pdf`, `jpg` alebo `png`. Vektorové obrázky (napr. `eps`, `svg`) je najvhodnejšie skonvertovať do formátu `pdf`, napríklad programom `Inkscape`.

Na vkladanie obrázkov použite prostredie `figure`, ktoré obrázok umiestni na vhodné miesto, väčšinou na vrch alebo spodok stránky a tiež sa stará o automatické číslovanie obrázkov. Na každý obrázok sa treba v hlavnom texte odvolať. Napríklad ilustráciu hry Červík vidíme na obrázku 2.1. Pri odvolávaní sa na číslo obrázku používame príkaz `\ref`. Pri vložení alebo zmazaní obrázku tak nemusíme ručne všetky ostatné obrázky prečíslovať.

Podobne tabuľky vkladajte pomocou prostredia `table`, pričom samotnú tabuľku vytvoríte príkazom `tabular`. Každú tabuľku potom spomeníte aj v hlavnom texte. Napríklad v tabuľke 2.1 vidíme porovnanie časov niekoľkých fiktívnych programov.

Obr. 2.1: Ukážka hry Červík. Červík je znázornený červenou farbou, voľné políčka sivou, jedlo zelenou a steny čiernou. Hoci tento popis obrázku je dlhší, v zdrojovom texte je aj kratšia verzia, ktorá sa zobrazí v zozname obrázkov.

Tabuľka 2.1: Doba výpočtu a operačná pamäť potrebná na spracovanie vstupu XYZ. V tomto popise môžeme vysvetliť detaily potrebné pre pochopenie údajov v tabuľke.

Meno programu	Čas (s)	Pamäť (MB)
Môj super program	25.6	120
Speedy 3.1	32.1	100
VeryOld	244.1	200

Ukážka kódu 2.1: Algoritmus na výpočet faktoriálu v jazyku C

---

```

1 int factorial = 1;
2 for(int i = 1; i <= n; i++) {
3     factorial *= i;
4 }
```

---

V texte môžete tiež potrebovať dlhšie matematické výrazy, ako napríklad tento

$$\sum_{k=0}^n q^k = \frac{q^{n+1} - 1}{q - 1}. \quad (2.1)$$

Použitím prostredia `equation` bol tento výraz zarovnaný na stred na zvláštnom riadku a očíslovaný. Na toto číslo sa tiež môžeme odvolať príkazom `\ref`. Napríklad rovnica (2.1) predstavuje súčet geometrickej postupnosti.

V práci tiež možno budete uvádzať úryvky kódu v niektorom programovacom jazyku. Môže vám pomôcť prostredie `lstlisting` z balíčka `listings`, v ktorom môžete nastaviť aj jazyk a kód bude krajšie sformátovaný. Ukážku nájdete ako Algoritmus 2.1.

Napokon, v texte nezabudnite citovať použitú literatúru pomocou príkazu `\cite`. Napríklad ďalšie detaily o systéme LaTeX nájdete v knihe od Tobiasa Oetikera a kolektívu [?]. Pre ukážku citujeme aj článok z vedeckého časopisu [2] a článok z konferencie [?], technickú správu [1], knihu [4] a materiál z internetu [?].



# Záver

Cieľom práce bolo ...



# Literatúra

- [1] A Alyami, SF Pileggi, O Sohaib, and I Hawryszkiewicz. Seamless transformation from use case to sequence diagrams. *PeerJ Comput Sci*, 9:e1444, Jun 2023.
- [2] Frank Breiting and Alexandre Jotterand. Sharing datasets for digital forensic: A novel taxonomy and legal concerns. *Forensic Science International: Digital Investigation*, 45:301562, 07 2023.
- [3] Alistair Cockburn. *Writing Effective Use Cases*. Addison-Wesley, 2001.
- [4] Gregorio Robles, Truong Ho-Quang, Regina Hebig, Michel R. V. Chaudron, and Miguel Angel Fernández. An extensive dataset of uml models in github. *2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR)*, pages 519–522, 2017.