

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

SPRACOVANIE DLHODOBÝCH MERANÍ
VYBRANÝCH CHARAKTERISTÍK INTERNETU
BAKALÁRSKA PRÁCA

2023

PETER TRENČANSKÝ

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

SPRACOVANIE DLHODOBÝCH MERANÍ
VYBRANÝCH CHARAKTERISTÍK INTERNETU
BAKALÁRSKA PRÁCA

Študijný program: Informatika
Študijný odbor: Informatika
Školiace pracovisko: Katedra informatiky
Školiteľ: Ing. Dušan Bernát, PhD.

Bratislava, 2023
Peter Trenčanský



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Peter Trenčanský
Študijný program: informatika (Jednoodborové štúdium, bakalársky I. st., denná forma)
Študijný odbor: informatika
Typ záverečnej práce: bakalárska
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Spracovanie dlhodobých meraní vybraných charakteristík Internetu
Processing and visualisation of long-term measurements of selected Internet metrics

Anotácia: Analyzujte štruktúru existujúcej databázy, ktorá obsahuje výsledky dlhodobých meraní odozvy rôznych uzlov v sieti Internet (výstupu ping, traceroute a pod.). Analyzujte dostupné nástroje pre spracovanie a vizualizáciu týchto dát. Navrhňte a implementujte systém s webovým rozhraním pre prácu s nameranými dátami, ktorý používateľovi umožní predovšetkým výber, triedenie a filtrovanie (podľa adries a času), ako aj výpočet štatistík a vizualizáciu ich hodnôt, časových priebehov, prípadne geografického rozloženia. Systém musí byť modulárny v tom zmysle, aby bolo možné konkrétne štatistiky a výstupy jednoducho dopĺňať. Riešenie overte na zobrazení časového vývoja priemernej, najmenej a najväčšej nameranej doby odpovede.

Vedúci: Ing. Dušan Bernát, PhD.
Katedra: FMFI.KI - Katedra informatiky
Vedúci katedry: prof. RNDr. Martin Škoviera, PhD.
Dátum zadania: 28.10.2022

Dátum schválenia: 31.10.2022

doc. RNDr. Dana Pardubská, CSc.
garant študijného programu

.....
študent

.....
vedúci práce

Pod'akovanie: Tu môžete poďakovať školiteľovi, prípadne ďalším osobám, ktoré vám s prácou nejako pomohli, poradili, poskytli dáta a podobne.

Abstrakt

Slovenský abstrakt v rozsahu 100-500 slov, jeden odstavec. Abstrakt stručne sumarizuje výsledky práce. Mal by byť pochopiteľný pre bežného informatika. Nemal by teda využívať skratky, termíny alebo označenie zavedené v práci, okrem tých, ktoré sú všeobecne známe.

Kľúčové slová: jedno, druhé, tretie (prípadne štvrté, piate)

Abstract

Abstract in the English language (translation of the abstract in the Slovak language).

Keywords:

Obsah

Úvod	1
1 Možnosti implementácie	3
1.1 Možnosti vývoja frontendového riešenia	3
1.2 Možnosti vývoja backendového riešenia	4
1.2.1 Možnosti implementácie vo vybranom jazyku	4
1.2.2 Databáza	7
2 Požiadavky a návrh riešenia	9
2.1 Analýza požiadaviek	9
2.1.1 Požiadavky na beh aplikácie	9
3 LaTeX	11
3.1 Obrázky	11
Záver	13
Príloha A	17
Príloha B	19

Zoznam obrázkov

3.1 Ukážka hry Červík	12
---------------------------------	----

Zoznam tabuliek

3.1	Doba výpočtu a operačná pamäť potrebná na spracovanie vstupu XYZ	12
-----	--	----

Úvod

Cieľom tejto práce je poskytnúť študentom posledného ročníka bakalárskeho štúdia informatiky kostru práce v systéme LaTeX a ukážku užitočných príkazov, ktoré pri písaní práce môžu potrebovať. Začneme stručnou charakteristikou úvodu práce podľa smernice o záverečných prácach [8], ktorú uvádzame ako doslovný citát.

Úvod je prvou komplexnou informáciou o práci, jej celi, obsahu a štruktúre. Úvod sa vzťahuje na spracovanú tému konkrétne, obsahuje stručný a výstižný opis problematiky, charakterizuje stav poznania alebo praxe v oblasti, ktorá je predmetom školského diela a oboznamuje s významom, cieľmi a zámermi školského diela. Autor v úvode zdôrazňuje, prečo je práca dôležitá a prečo sa rozhodol spracovať danú tému. Úvod ako názov kapitoly sa nečísluje a jeho rozsah je spravidla 1 až 2 strany.

V nasledujúcej kapitole nájdete ukážku členenia kapitoly na menšie časti a v kapitole 3 nájdete príkazy na prácu s tabuľkami, obrázkami a matematickými výrazmi. V kapitole ?? uvádzame klasický text Lorem Ipsum a na koniec sa budeme venovať záležitostiam záveru bakalárskej práce.

Kapitola 1

Možnosti implementácie

Každý softvér si musí určiť, pre aké platformu je vyvíjaný. Samozrejme, čím väčšiu škálu platforiem pokryje, tým je dostupnejší pre väčšie množstvo používateľov. Dnes drvivá väčšina používateľov používa istú podmnožinu zo systémov Android, Linux, Windows, iOS a MacOS X. Spoločnou črtou týchto platforiem je, že medzi nimi neexistuje taká dvojica, z ktorej by jedna platforma natívne podporovala programy napísané pre druhú platformu. Vyvíjať aplikáciu päťkrát, pre každú platformu v jej natívnom prostredí, však nie je veľmi efektívne.

Namiesto toho dnes existuje niekoľko riešení umožňujúcich zdieľať časti zdrojového kódu medzi riešeniami pre rôzne systémy. Ako príklad môžeme uviesť frameworky Flutter, React Native alebo MAUI pre .NET.

Ako univerzálne riešenie sa dnes používajú webové aplikácie. Takéto aplikácie majú niekoľko dôležitých výhod, ako napríklad rýchla dostupnosť, žiadna nutnosť inštalácie a jednoduché spustenie na všetkých platformách, na ktorých je dostupný webový prehliadač (napríklad Firefox, Chrome alebo Opera). Webové aplikácie zvyknú mávať dve oddelené časti - frontend, bežiaci vo webovom prehliadači klienta a backend, bežiaci na serveri. Tieto dve časti spolu komunikujú pomocou siete, najčastejšie pomocou protokolov HTTP(S) alebo WebSocket. Úlohou frontendu je poskytovať prívetivé rozhranie s ktorým môže používateľ pracovať, úlohou backendu je zase spracovávať dáta, komunikovať s databázou a poskytovať dáta frontendu. Práve pre takýto model som sa rozhodol.

1.1 Možnosti vývoja frontendového riešenia

Programovanie webových stránok sa dnes nezaobíde bez programovacieho jazyka JavaScript, ktorý používajú webové prehliadače. Alternatívne sa dajú použiť prekladače iných jazykov do WebAssembly (napríklad framework Blazor pre .NET) alebo jazyk TypeScript, ktorý je rozšírením JavaScriptu o typový systém a iné užitočné konštruk-

cie, zároveň je plne kompatibilný s JavaScriptom (je jeho nadmnožinou) a vo webových prehliadačoch sa spúšťa pomocou prekladu do JavaScriptu. Na vývoj webového frontendu je možnosť používať webové frameworky, ktoré umožňujú delenie stránok na moduly, ktoré sa dajú jednotlivo znovupoužívať a paraneťrizovať na rôznych podstránkach aplikácie, čím zjednodušujú spravovanie a čitateľnosť aplikácie. V súčasnej dobe sú najpopulárnejšie frameworky React, Vue a Angular. Kým React a Vue umožňujú prácu v JavaScripte ale voliteľne je možné vyvíjať aj v TypeScript, Angular prácu v TypeScript vyžaduje. Tieto frameworky sú podporované vo všetkých nových verziách moderných prehliadačov. Keďže všetky tieto frameworky fungujú na veľmi podobnom princípe, je veľmi ťažké povedať, ktorý by bola pre moju aplikáciu najvhodnejší. Vzhľadom na to som sa rozhodol pre Angular, v ktorom mám navyše praktických skúseností.

1.2 Možnosti vývoja backendového riešenia

Pre vývoj backendového riešenia je vhodné vybrať taký programovací jazyk, pre ktorý existuje framework umožňujúci vývoj API, ktoré bude podporovať protokol HTTP(s). Takýchto jazykov existuje veľmi veľa, dokonca pre niektoré jazyky existuje viacero rôznych knižníc umožňujúcich výstavbu takýchto API. Ako príklad môžeme uviesť jazyk Python a knižnice FastAPI alebo Flask a mnohé iné, jazyk Java a framework Spring Boot, jazyk C# a framework ASP.NET Core a mnohé iné. Všetky tieto možnosti majú dostatočné nástroje a je ťažké určiť, ktorý z nich je objektívne najvhodnejší, preto som sa rozhodol pre jazyk C# a ASP.NET Core, s ktorým mám najviac skúseností.

1.2.1 Možnosti implementácie vo vybranom jazyku

Napriek tomu, že už sme si vybrali jazyk, stále máme niekoľko rôznych možností, ako poňať vývoj aplikácie, najmä čo sa týka organizácie kódu alebo knižníc. Na komunikáciu s databázou sa dá použiť buď priamo knižnica ADO.NET, ktorá umožňuje manuálne vykonávanie SQL dotazov a ich ručné mapovanie do objektov, alebo knižnice typu ORM (Object Relational Mapping), ktoré pôsobia ako wrapper pre ADO.NET a sú schopné istú časť mapovania vykonať za nás. Najpoužívanejšie ORM v C# sú Entity Framework Core (ďalej len EFC) a Dapper. Kým Dapper mapuje len výsledky dopytov na objekty (tzv. micro-ORM), EFC je schopné mapovať kód v jazyku C# a knižnicu LINQ na SQL príkazy, čím v jednoduchších aplikáciách dokáže úplne odbúrať nutnosť písania SQL dopytov ručne.

Vzhľadom na možnú zložitosť dopytov som sa rozhodol nepoužiť EFC, no použijem Dapper ako skratku pre odbúranie opakujúceho sa kódu vznikajúceho pri používaní ADO.NET (otváranie a zatváranie spojenia, definície parametrov, mapovanie výsledkov na objekty).

Algoritmus 1.1: Ukážka kódu pre ten istý dopyt pomocou rôznych knižníc - ADO.NET

```
1     private NpgsqlConnection CreateConnection()
2     {
3         return new NpgsqlConnection(_connectionString);
4     }
5
6     public async Task<IEnumerable<IpAddressInfo>> GetIpAddresses2(
7         int offset = 0, int limit = int.MaxValue)
8     {
9         await using var connection = CreateConnection();
10        await connection.OpenAsync();
11        var command = connection.CreateCommand();
12        command.CommandText = "SELECT * FROM IpAddresses LIMIT
13            @limit OFFSET @offset";
14        command.Parameters.AddWithValue("@limit", limit);
15        command.Parameters.AddWithValue("@offset", offset);
16        var reader = await command.ExecuteReaderAsync();
17        var result = new List<IpAddressInfo>();
18        while (await reader.ReadAsync())
19        {
20            result.Add(new IpAddressInfo()
21            {
22                City = reader["City"] as string,
23                CountryCode = reader["CountryCode"] as string,
24                Id = Convert.ToInt32(reader["Id"]),
25                IpStringValue = reader["IpStringValue"] as string,
26            });
27        }
28        return result;
29    }
```

Algoritmus 1.2: Ukážka kódu pre ten istý dopyt pomocou rôznych knižníc - Dapper

```
1     private NpgsqlConnection CreateConnection()  
2     {  
3         return new NpgsqlConnection(_connectionString);  
4     }  
5  
6     public async Task<IEnumerable<IpAddressInfo>> GetIpAddresses(  
7         int offset = 0, int limit = int.MaxValue)  
8     {  
9         await using var connection = CreateConnection();  
10        return await connection.QueryAsync<IpAddressInfo>("SELECT *  
        FROM IpAddresses LIMIT @limit OFFSET @offset", new {  
            limit, offset });  
11    }
```

Algoritmus 1.3: Ukážka kódu pre ten istý dopyt pomocou rôznych knižníc - EFC

```
1     private ApplicationDbContext _dbContext;  
2  
3     public async Task<IEnumerable<IpAddressInfo>> GetIpAddresses(  
4         int offset = 0, int limit = int.MaxValue)  
5     {  
6         return await _dbContext.IpAddresses.Skip(offset).Take(limit  
7             ).ToListAsync();  
8     }
```

Pri štruktúre kódu je nutné sa rozhodnúť, či budeme endpointy deliť do controllerov alebo nie (tzv. minimal API). Kým výhodou controllerov je lepšia štrukturovateľnosť a väčšia prehľadnosť pri väčšom množstve endpointov, výhodou mimal API je mierne menej kódu a potenciálny kratší čas písania. Keďže aplikácia môže mať väčšie množstvo endpointov, rozhodol som sa použiť controllery.

1.2.2 Databáza

S backendovým riešením úzko súvisí aj výber databázového providera pre ukladanie spracovaných dát. V dnešnej dobe sa používajú dva základné typy databáz - SQL a NoSQL. Keďže naše dáta budú mať jasnú štruktúru, bude vhodné použiť jedného z mnohých SQL providerov, napríklad PostgreSQL, MySQL, MS SQL SQLite atď. Keďže databáza s údajmi o meraniach je PostgreSQL, rozhodol som sa použiť tiež PostgreSQL pre zachovanie uniformity.

Kapitola 2

Požiadavky a návrh riešenia

2.1 Analýza požiadaviek

Požiadavky na aplikáciu vznikli po dohode s Ing. Dušanom Bernátom, PhD. Cieľom je vytvoriť webovú aplikáciu ktorá bude schopná vizualizovať výsledky meraní internetu v grafickej forme pomocou diagramov, grafov a máp.

2.1.1 Požiadavky na beh aplikácie

Požiadavky na beh serverovej časti:

- Systém musí byť schopný pracovať s PostgreSQL
- Systém nesmie zapisovať do databázy s meraniami
- Všetky časti systému musia bežať na Linuxovom prostredí (CentOS alebo Ubuntu)
- Systém musí byť schopný priebežne predspracovávať namerané dáta po týždňoch, aby boli vizualizácie zobraziteľné ihneď (bez dlhého čakania)
- Systém nesmie byť závislý od vonkajších servisov iných, ako databáza s meraniami (lokalizačné servery aj mapové servery musia bežať v rámci systému)

Požiadavky na beh klientskej časti:

- Aplikácia musí bez problémov bežať na všetkých moderných prehliadačoch
- Aplikácia musí vedieť vizualizácie zobrazovať bez prídlhého načítavania

Detaily implementácie boli ponechané na moje uváženie. Detaily rozhodovania sú popísané v kapitole 1.

Serverový systém bude implementovaný na .NET 6.0. Systém bude pozostávať z niekoľkých aplikácií, z ktorých jedna bude pôsobiť ako server pre klientskú časť, jedna

bude pôsobiť ako API pre vyťahovanie spracovaných dát a zvyšné budú background workery, ktoré budú mať za úlohu priebežne spracovávať namerané dáta. Predspracované dáta sa budú ukladať do PostgreSQL 15.0 databázy. Tieto workery budú spúšťané task schedulerom jedenkrát týždenne. Systém bude bežať v niekoľkých kontaineroch v platforme Docker. Ako samostatný kontajner bude bežať databáza, API na prístup k dátam, nginx server pre klientskú aplikáciu, každý background worker a Open Street Maps tile server. Ako údaje na lokalizáciu ip adries bola použitá voľná databáza db-ip lite [5].

Klientská časť bude implementovaná pomocou frameworku Angular 15.0. API bude volané pomocou vstavanej knižnice httpClient. Na zobrazenie presnej geografickej mapy sveta budú použité Open Street Maps pomocou knižnice Leaflet a jej wrapperu pre Angular ngx-leaflet. Na zobrazenie mapy sveta po krajinách je použitý mierne upravený verejne voľný svg template [7]. Na zobrazenie diagramov bude použitá knižnica ngx-charts.

Kapitola 3

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`.

3.1 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`.

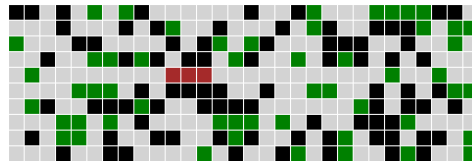
Na vkladanie obrázkov použite prostredie `figure`, ktoré obrázok umiestni na vhodné miesto, väčšinou na vrch alebo spodok stránky a tiež sa stará o automatické číslovanie obrázkov. Na každý obrázok sa treba v hlavnom texte odvolať. Napríklad ilustráciu hry Červík vidíme na obrázku 3.1. Pri odvolávaní 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 spomeňte aj v hlavnom texte. Napríklad v tabuľke 3.1 vidíme porovnanie časov niekoľkých fiktívnych programov.

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

$$\sum_{k=0}^n q^k = \frac{q^{n+1} - 1}{q - 1}. \quad (3.1)$$

Použitím prostredia `equation` bol tento výraz zarovnaný na stred na zvláštnom riadku



Obr. 3.1: 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 3.1: 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

a očíslovaný. Na toto číslo sa tiež môžeme odvolať príkazom `\ref`. Napríklad rovnica (3.1) 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 3.1.

Napokon, v texte nezabudnite citovať použitú literatúru pomocou príkazu `\cite`. Napríklad ďalšie detaily o systéme LaTeX nájdete v knihe od Tobiasa Oetikera a kolektívu [6]. Pre ukážku citujeme aj článok z vedeckého časopisu [3] a článok z konferencie [2], technickú správu [4], knihu [1] a materiál z internetu [8].

Algoritmus 3.1: Algoritmus na výpočet faktoriálu v jazyku C

```

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

Záver

Na záver už len odporúčania k samotnej kapitole Záver v bakalárskej práci podľa smernice [8]: „V závere je potrebné v stručnosti zhrnúť dosiahnuté výsledky vo vzťahu k stanoveným cieľom. Rozsah záveru je minimálne dve strany. Záver ako kapitola sa nečísluje.“

Všimnite si správne písanie slovenských úvodzoviek okolo predchádzajúceho citátu, ktoré sme dosiahli príkazom \uv.

V informatických prácach niekedy býva záver kratší ako dve strany, ale stále by to mal byť rozumne dlhý text, v rozsahu aspoň jednej strany. Okrem dosiahnutých cieľov sa zvyknú rozoberať aj otvorené problémy a námety na ďalšiu prácu v oblasti.

Abstrakt, úvod a záver práce obsahujú podobné informácie. Abstrakt je kratší text, ktorý má pomôcť čitateľovi sa rozhodnúť, či vôbec prácu chce čítať. Úvod má umožniť zorientovať sa v práci skôr než ju začne čítať a záver sumarizuje najdôležitejšie veci po tom, ako prácu prečítal, môže sa teda viac zamerať na detaily a využívať pojmy zavedené v práci.

Literatúra

- [1] X. Autor1 and Y. Autor2. *Názov knihy*. Vydavateľstvo, 1900.
- [2] X. Autor1 and Y. Autor2. Názov článku (väčšinou z konferencie). In *Názov zborníka (väčšinou názov konferencie spolu s ročníkom)*, pages 1–100. Vydavateľstvo, 1900.
- [3] X. Autor1 and Y. Autor2. Názov článku z časopisu. *Názov časopisu, ktorý článok uverejnil*, 4(3):1–100, 1900.
- [4] X. Autor1 and Y. Autor2. Názov technickej správy. Technical Report TR123/1999, Inštitút vydávajúci správu, June 1999.
- [5] DB-IP, 2022. Dostupné z <https://db-ip.com/db/download/ip-to-city-lite>.
- [6] Tobias Oetiker, Hubert Partl, Irene Hyna, and Elisabeth Schlegl. *Nie príliš stručný úvod do systému LaTeX2e*. 2002. Preklad Ján Buša ml. a st.
- [7] simplemaps.com, 2017. Dostupné z <https://simplemaps.com/resources/svg-world>.
- [8] Univerzita Komenského v Bratislave. Vnútorňý predpis č. 7/2018, Úplné znenie vnútorného predpisu č. 12/2013 Smernice rektora Univerzity Komenského v Bratislave o základných náležitostiach záverečných prác, rigorózných prác a habilitačných prác, kontrole ich originality, uchovávaní a sprístupňovaní na Univerzite Komenského v Bratislave v znení dodatku č. 1 a dodatku č. 2 smernica rektora Univerzity Komenského v Bratislave o základných náležitostiach záverečných prác, rigorózných prác a habilitačných prác, kontrole ich originality, uchovávaní a sprístupňovaní na Univerzite Komenského v Bratislave, 2013. [Citované 2020-10-19] Dostupné z https://uniba.sk/fileadmin/ruk/legislativa/2018/Vp_2018_07.pdf.

Príloha A: obsah elektronickej prílohy

V elektronickej prílohe priloženej k práci sa nachádza zdrojový kód programu a súbory s výsledkami experimentov. Zdrojový kód je zverejnený aj na stránke <http://mojadresa.com/>.

Ak uznáte za vhodné, môžete tu aj podrobnejšie rozpísať obsah tejto prílohy, prípadne poskytnúť návod na inštaláciu programu. Alternatívou je tieto informácie zahrnúť do samotnej prílohy, alebo ich uviesť na oboch miestach.

Príloha B: Používateľská príručka

V tejto prílohe uvádzame používateľskú príručku k nášmu softvéru. Tu by ďalej pokračoval text príručky. V práci nie je potrebné uvádzať používateľskú príručku, pokiaľ je používanie softvéru intuitívne alebo ak výsledkom práce nie je ucelený softvér určený pre používateľov.

V prílohách môžete uviesť aj ďalšie materiály, ktoré by mohli pôsobiť rušivo v hlavnom texte, ako napríklad rozsiahle tabuľky a podobne. Materiály, ktoré sú príliš dlhé na ich tlač, odovzdajte len v electronickej prílohe.