



FAKULTA APLIKOVANÝCH VĚD  
ZÁPADOČESKÉ UNIVERZITY  
V PLZNI

KATEDRA INFORMATIKY  
A VÝPOČETNÍ TECHNIKY

## Bakalářská práce

# Integrovaný systém pro automatizaci záznamu, úprav a publikaci přednášek

Martin Reich



FAKULTA APLIKOVANÝCH VĚD  
ZÁPADOČESKÉ UNIVERZITY  
V PLZNI

KATEDRA INFORMATIKY  
A VÝPOČETNÍ TECHNIKY

## **Bakalářská práce**

# **Integrovaný systém pro automatizaci záznamu, úprav a publikaci přednášek**

Martin Reich

### **Vedoucí práce**

Doc. Ing. Roman Mouček, Ph.D.

© Martin Reich, 2024.

Všechna práva vyhrazena. Žádná část tohoto dokumentu nesmí být reprodukována ani rozšiřována jakoukoli formou, elektronicky či mechanicky, fotokopírováním, nahráváním nebo jiným způsobem, nebo uložena v systému pro ukládání a vyhledávání informací bez písemného souhlasu držitelů autorských práv.

**Citace v seznamu literatury:**

REICH, Martin. *Integrovaný systém pro automatizaci záznamu, úprav a publikaci přednášek*. Plzeň, 2024. Bakalářská práce. Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Katedra informatiky a výpočetní techniky. Vedoucí práce Doc. Ing. Roman Mouček, Ph.D.

# Podklad pro zadání BAKALÁŘSKÉ práce studenta

Jméno a příjmení: **Martin REICH**  
Osobní číslo: **A22B0123P**  
Adresa: **Plovární 1458/21, Plzeň – Jižní Předměstí, 30100 Plzeň 1, Česká republika**  
Téma práce: **Integrovaný systém pro automatizaci záznamu, úprav a publikaci přednášek**  
Téma práce anglicky: **An integrated system for automating the recording, editing and publication of lectures**  
Jazyk práce: **Čeština**  
Související osoby: **Doc. Ing. Roman Mouček, Ph.D. (Vedoucí)**  
**Katedra informatiky a výpočetní techniky**

## Zásady pro vypracování:

1. Seznamte se současnými metodami a možnostmi převodu řeči do textu a zpět a metodami parafrázování
2. Seznamte se současnými systémy, nástroji a knihovnami pro transkripci řeči na text, převod textu na řeč a parafrázování
3. Navrhněte integrovaný systém, který dokáže automatizovat převod zvukového a obrazového záznamu přednášky v češtině či angličtině do textu, tento následně parafrázovat a zpětně převést do zvukové a obrazové podoby přednášky
4. Navržený systém dle bodu 3 implementujte.
5. Ověřte výsledné řešení na dostatečném počtu přednášek v českém i anglickém jazyce.

## Seznam doporučené literatury:

Dodá vedoucí práce

Podpis studenta:

Datum:

Podpis vedoucího práce:

Datum:



# Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného akademického titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Západočeská univerzita v Plzni má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona.

V Plzni dne 31. prosince 2024

.....

Martin Reich

## **Abstrakt**

TODO

## **Abstract**

TODO in english

## **Klíčová slova**

Převod řeči na text • AI • Převod textu na řeč • Parafrázování • Automatizace

## Poděkování

Rád bych tímto poděkoval vedoucímu bakalářské práce Doc. Ing. Romanu Moučkovi, Ph.D. za pomoc a odborné vedení při vypracování této práce.



# Obsah

<b>1</b>	<b>Analýza problému</b>	<b>2</b>
1.1	Studium současných metod převodu řeči do textu a zpět, včetně parafrázování . . . . .	2
1.2	Průzkum dostupných modelů, nástrojů a knihoven a jejich vyzkoušení	2
1.3	Návrh integrovaného systému a jeho implementace . . . . .	3
1.4	Testování systému . . . . .	3
<b>2</b>	<b>Převod řeči do textu</b>	<b>4</b>
2.1	Jak to funguje . . . . .	4
2.2	Metody . . . . .	5
2.2.1	Umělé neuronové sítě . . . . .	5
2.2.2	Hluboké učení . . . . .	6
2.2.3	Rekurentní neuronové sítě . . . . .	6
2.2.4	Transformer . . . . .	7
2.2.5	Konvoluční neuronové sítě . . . . .	8
2.3	Vybrané modely pro řešení daného problému . . . . .	8
2.3.1	Testování a výsledky modelů . . . . .	9
<b>3</b>	<b>Parafrázování</b>	<b>10</b>
3.1	Metody parafrázování . . . . .	10
3.2	Vybrané modely pro řešení daného problému . . . . .	10
3.2.1	Testování a výsledky modelů . . . . .	11
<b>4</b>	<b>Zkratky</b>	<b>12</b>
	<b>Bibliografie</b>	<b>13</b>

# Analýza problému

# 1

## 1.1 Studium současných metod převodu řeči do textu a zpět, včetně parafrázování

V první fázi je nutné porozumět současným technologiím pro:

- Převod řeči na text: Mezi nejznámější metody patří systémy založené na hlubokých neuronových sítích, konkrétně na architekturách jako jsou konvoluční neuronové sítě (CNN) a rekurentní neuronové sítě (RNN), nebo i pokročilejší transformery.
- Převod textu na řeč: Nejnovější metody TTS využívají neuronové sítě, jako třeba TTS model od OpenAI, ElevenLabs modely, Google TTS model atd.. Tyto modely zajišťují přirozenější syntézu řeči a mohou pracovat s intonací a rytmem, což je klíčové pro parafrázované texty.
- Parafrázování: Parafrázování textu je možné provádět pomocí modelů NLP (natural language processing, zpracování přirozeného jazyka), které dokážou přepsat text jinými slovy, aniž by změnilý význam. Zde se využívají transformer modely, například T5 nebo GPT-4. Při parafrázování je výzvou zachovat srozumitelnost a přirozenost textu, zejména pokud se bude převádět zpět na řeč.

## 1.2 Průzkum dostupných modelů, nástrojů a knihoven a jejich vyzkoušení

Druhá fáze zahrnuje detailní průzkum stávajících nástrojů, modelů a knihoven, které podporují každý krok procesu a jejich vyzkoušení. **TODO dopsat stávající modely, knihovny atd.**

## 1.3 Návrh integrovaného systému a jeho implementace

Dále je potřeba si rozmyslet a navrhnout systém, který:

- Automatizuje převod mluveného slova na text: Tento krok zahrnuje přepis audia nebo video záznamů přednášky na text.
- Parafrázuje přepsaný text: Po přepisu je text předán k parafrázování, aby výsledný text byl přeformulovaný.
- Převádí parafrázovaný text zpět na řeč: Po parafrázování je třeba převést text na řeč, která by měla být pokud možno přirozená a příjemná k poslechu.
- Synchronizuje obraz a zvuk: Nakonec je potřeba synchronizovat nový zvuk s obrazem, který může být originální nebo nový ve formě animace či avatara.

Po navržnutí takového systému je potřeba ho implementovat (viz. implementace **Dodat odkaz po přidání implementace**)

## 1.4 Testování systému

Na závěr je potřeba zvolit vhodnou metodu pro testování navrženého a implementovaného systému. Tento krok také zahrnuje testování podle kritérií na reálných přednáškách v českém a anglickém jazyce. Mezi tyto kritéria může patřit:

- Kvalita: Testování přesnosti rozpoznávání řeči, přirozenosti parafrázování a srozumitelnosti výstupu.
- Uživatelské hodnocení: Testování s cílovými uživateli, kteří posoudí užitečnost a přirozenost celého procesu.
- Výkon: Hodnocení rychlosti a efektivity systému, aby byla zajištěna schopnost zpracovat delší přednášky bez výrazného zpoždění.

# Převod řeči do textu

## 2

Převod řeči do textu (Speech-to-Text, STT) je oblast, která se zabývá automatickým rozpoznáváním mluveného slova a jeho převodem na psaný text. Je to multidisciplinární obor zahrnující akustiku, jazykovědu, statistiku a strojové učení.

## 2.1 Jak to funguje

Převod řeči na text je software, který funguje tak, že naslouchá zvuku a poskytuje upravitelný, doslovný přepis na daném zařízení. Software to provádí prostřednictvím rozpoznávání hlasu. Počítačový program využívá jazykové algoritmy k třídění zvukových signálů z mluvených slov a převádí tyto signály na text pomocí znaků zvaných Unicode. Převod řeči na text funguje prostřednictvím složitého modelu strojového učení, který zahrnuje několik kroků.

1. Zvuk vychází z úst k vytvoření slova, také vytváří sérii vibrací. Technologie převodu řeči na text funguje tak, že zachycuje tyto vibrace a překládá je do digitálního jazyka prostřednictvím analogově-digitálního převodníku.
2. Analogově-digitální převodník přijme zvuk z audio souboru, podrobně zemře jeho vlny a provede nad nim filtraci, aby odlišil relevantní zvuky.
3. Zvuk je poté segmentován na setiny nebo tisíce sekundy a poté je přiřazen k fonémům. Foném je jednotka zvuku, která odlišuje jedno slovo od druhého v jakémkoli daném jazyce.
4. Fonémy jsou poté zpracovány sítí prostřednictvím matematického modelu, který je porovnává s dobře známými větami, slovy a frázemi.
5. Text je poté prezentován jako text nebo jako počítačový požadavek na základě nejpravděpodobnější verze audia.

## 2.2 Metody

Současné metody, které se používají k převodu řeči na text.

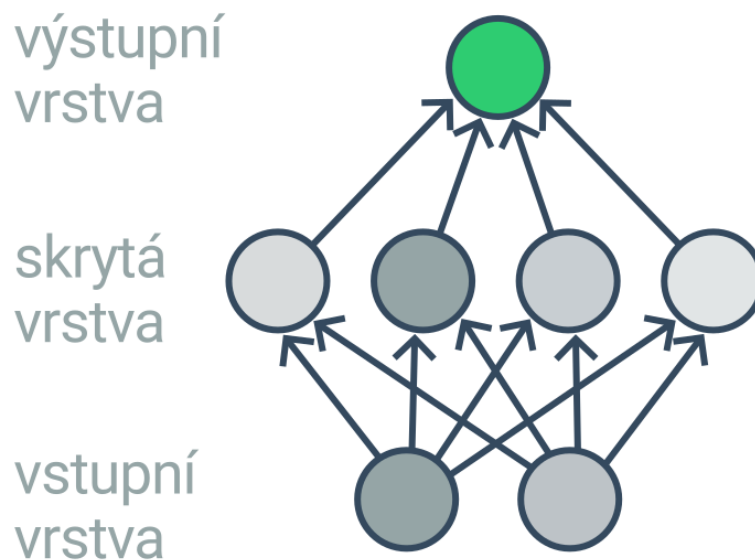
### 2.2.1 Umělé neuronové sítě

Neuronová síť je model strojového učení, který simuluje rozhodování lidského mozku. Používá procesy, jež napodobují způsob, jakým biologické neurony spolupracují, aby identifikovaly jevy, zhodnotily možnosti a dospěly k závěrům. Každá neuronová síť se skládá z vrstev uzlů, které fungují jako umělé neurony. Tyto vrstvy zahrnují vstupní vrstvu, jednu nebo více skrytých vrstev a výstupní vrstvu. Každý uzel je propojen s ostatními uzly a má svou vlastní váhu a prahovou hodnotu. Pokud výstup některého uzlu překročí stanovený práh, tento uzel se aktivuje a předá data do další vrstvy sítě. Jinak nedochází k žádnému předání dat. Neuronové sítě potřebují tréninková data, aby se učily a zvyšovaly svou přesnost. Jakmile jsou optimalizovány, stávají se silnými nástroji v oblasti informatiky a umělé inteligence, což nám umožňuje rychle klasifikovat a shlukovat data. Úkoly, jako je rozpoznávání řeči nebo obrazů, mohou trvat minuty namísto hodin. Výhody:

- Flexibilita a schopnost učení: Neuronové sítě se dokáží učit z tréninkových dat a zlepšovat svou přesnost, což je činí adaptabilními na různé úkoly a scénáře.
- Schopnost zpracovávat složité vzory: Neuronové sítě jsou schopné identifikovat složité vzory a vztahy v datech, které by tradiční algoritmy mohly přehlédnout.
- Všestrannost: Můžou být aplikovány v různých oblastech, včetně zpracování obrazů, přirozeného jazyka, doporučovacích systémů, medicíny a mnoha dalších.

Nevýhody:

- Požadavky na tréninková data: Neuronové sítě potřebují velké množství kvalitních tréninkových dat, aby byly schopny efektivně se učit a generalizovat na nová data.
- Trénink neuronových sítí může být časově náročný a vyžaduje značné výpočetní zdroje, zejména při práci s velkými datovými sadami nebo složitými architekturami.
- Riziko přetrénování: Pokud jsou neuronové sítě trénovány na příliš malých nebo nevyvážených datových sadách, mohou se "naučit" specifické vzory v tréninkových datech a mít problémy s generalizací na nová, neznámá data.



Obrázek 2.1: Ukázka struktury neuronové sítě

### 2.2.2 Hluboké učení

Hluboké učení je podmnožina strojového učení, která využívá více vrstvé neuronové sítě, nazývané hluboké neuronové sítě, k simulaci složitějšího rozhodovacího procesu lidského mozku. Hlavní rozdíl mezi hlubokým učením a strojovým učením spočívá ve struktuře architektury základní neuronové sítě. Tradiční modely strojového učení, které nejsou hluboké, používají jednoduché neuronové sítě s jednou nebo dvěma výpočetními vrstvami. Naopak modely hlubokého učení mají tři a více vrstev, často stovky nebo dokonce tisíce vrstev. Zatímco modely s učením s učitelem vyžadují strukturovaná a označená vstupní data pro dosažení přesných výstupů, modely hlubokého učení mohou pracovat s učením bez učitele. Díky učení bez učitele dokážou modely hlubokého učení extrahovat charakteristiky, vlastnosti a vztahy, které potřebují k dosažení přesných výstupů z neupravených a neorganizovaných dat. Navíc tyto modely mohou vyhodnocovat a zdokonalovat své výstupy pro zvýšení přesnosti.

### 2.2.3 Rekurentní neuronové sítě

Rekurentní neuronové sítě (RNN) jsou modely hlubokého učení a jsou schopné zpracovávat sekvence dat tím, že mají vnitřní paměť, která uchovává informace o předchozích vstupech. Tato vlastnost jim umožňuje modelovat časové závislosti a

vztahy v datech. Na rozdíl od tradičních neuronových sítí, které zpracovávají vstupy nezávisle, RNN zohledňují kontext a historii dat. Výhody:

- Zpracování sekvenčních dat: RNN jsou ideální pro úkoly, kde je pořadí dat důležité.
- Flexibilita: Mohou zpracovávat vstupy a výstupy různé délky (např. různé délky vět).

Nevýhody:

- Vanishing gradient: Při trénování mohou gradienty zmizet, což brání učení dlouhodobých závislostí.
- Vysoké výpočetní nároky: RNN mohou být náročné na výpočetní výkon, zejména při práci s dlouhými sekvencemi.

**Dodat obrázek rekuretní neuronové sítě**

## 2.2.4 Transformer

Transformery jsou modely hlubokého učení a používají mechanismus vlastní pozornosti, který umožňuje modelu vážit různá slova v sekvenci na základě jejich relevance. To znamená, že model může posoudit, která slova mají vliv na ostatní slova, což je klíčové pro porozumění kontextu. Jelikož transformery nemají vnitřní strukturu pro zpracování sekvencí (na rozdíl od RNN 2.2.3), používají se k nim pozicové kódování, aby modely mohly rozpoznat pořadí slov v sekvenci. Tato kódování přidávají k vektorům slov informace o jejich pozici v sekvenci. Výhody:

- Paralelizace: Transformery umožňují paralelní zpracování vstupních dat, což urychluje trénink ve srovnání s RNN.
- Zachycení dlouhodobých závislostí: Díky mechanismu pozornosti jsou schopny efektivně sledovat vztahy mezi slovy na delší vzdálenosti.
- Flexibilita: Lze je snadno aplikovat na různé úkoly a adaptovat je pro konkrétní aplikace.

Nevýhody:

- Vysoké nároky na paměť: Vzhledem k tomu, že transformery zpracovávají celou sekvenci najednou, mohou mít vysoké požadavky na paměť, zejména při práci s dlouhými sekvencemi.
- Potřeba velkých dat: K dosažení dobrého výkonu vyžadují transformery velké množství tréninkových dat.

## 2.2.5 Konvoluční neuronové sítě

Konvoluční neuronové sítě (CNN) jsou specifickým typem neuronových sítí, které se ukázaly jako velmi efektivní při analýze vizuálních dat, jako jsou obrázky a videa. Tyto sítě se široce používají v oblastech, jako je počítačové vidění, rozpoznávání obrazů, analýza videí a dokonce i v zpracování přirozeného jazyka. CNN má tři hlavní typy vrstev.

- **Konvoluční vrstva:** Tato vrstva aplikuje konvoluční operaci na vstupní data pomocí malých filtrů (filtr pokrývá jen část vstupních dat). Hodnoty (váhy) ve filtru se vynásobí s hodnotami na vstupních datech, přes který je filtr "umístěn". Filtr se poté přesune, aby vytvořil reprezentaci nové části vstupních dat, a proces se opakuje, dokud nejsou pokryta celá vstupní data.
- **Pooling vrstva:** Podobně jako u konvoluční vrstvy, tato vrstva aplikuje filtry na vstupní data. Jediný rozdíl je, že filtr nemá žádné váhy a ze vstupních dat se vezme hodnota se největší nebo průměrnou hodnotou.
- **Plně propojená vrstva:** Na konci CNN se obvykle nachází jedna nebo více plně propojených vrstev, které kombinují výstupy z předchozích vrstev a předpovídají finální výstup.

Výhody:

- **Efektivita:** CNN dokáží zpracovávat velké množství dat s menšími výpočetními nároky díky sdílení vah a redukci rozměrů pomocí pooling vrstev.
- **Vysoká přesnost:** Tyto sítě dosahují vynikajících výsledků v úlohách klasifikace obrazů a detekce objektů.

Nevýhody:

- **Potřeba velkého množství dat:** Pro úspěšné trénování vyžadují CNN velké množství označených tréninkových dat.
- **Omezené schopnosti na neobvyklých datech:** Modely CNN mohou mít potíže s generalizací na atypická data, která se výrazně liší od tréninkových dat.

## 2.3 Vybrané modely pro řešení daného problému

TODO dopsat vybrané modely jejich výhody a nevýhody



## 2.3.1 Testování a výsledky modelů

**TODO** dodat graf/y s výsledky testování modelů podle rychlosti a přesnosti a napsat který model bude použit k řešení problému.

# Parafrázování

## 3

Parafrázování je proces přeformulování textu nebo myšlenek, při kterém zachováme původní význam, ale změníme slovní formulaci a strukturu. Cílem parafrázování je převést informace jinými slovy, často pro lepší pochopení, přizpůsobení určitému publiku nebo aby se text stal originálnější a nevznikl tak problém s plagiátorstvím.

## 3.1 Metody parafrázování

1. Synonymizace: Výměna některých slov za vhodné synonyma, která zachovají význam věty.
2. Použití jiného gramatického tvaru: Můžeme nahradit sloveso podstatným jménem nebo naopak, případně použít pasivní tvar místo aktivního (a opačně).
3. Změna větné struktury: Přeskládání slovosledu nebo struktury věty, aniž by došlo ke změně významu.
4. Zjednodušení nebo rozšíření textu: Pokud je text příliš komplikovaný, parafrázování může zahrnovat jeho zjednodušení. Naopak, pokud je text příliš stručný, může být rozšířen o doplňující detaily.
5. Změna slovní zásoby: Přeformulování pomocí odlišného stylu nebo slovníku.

Pro parafrázování se používají kombinace těchto metod.

## 3.2 Vybrané modely pro řešení daného problému

**TODO** dopsat vybrané modely jejich výhody a nevýhody

### 3.2.1 Testování a výsledky modelů

**TODO** dodat graf/y s výsledky testování modelů podle rychlosti a napsat který model bude použit a proč k řešení problému.

# Zkratky

## 4

TTS	text-to-speech
RNN	rekuretní neuronová síť
CNN	konvoluční neuronová síť
STT	speech-to-text

Tabulka 4.1: Tabulka zkratk

# Bibliografie

- [1] Jim Holdsworth and Mark Scapicchio (2024), *Deep learning*, IBM, <https://www.ibm.com/topics/deep-learning>.
- [2] *Neural networks*, IBM, <https://www.ibm.com/topics/neural-networks>.
- [3] *Transformer model*, IBM, <https://www.ibm.com/topics/transformer-model>.
- [4] *Recurrent neural network*, Wikipedia, [https://en.wikipedia.org/wiki/Recurrent\\_neural\\_network](https://en.wikipedia.org/wiki/Recurrent_neural_network).
- [5] *Techniques for paraphrasing*, An Intermediate Guide to Writing in English for University Study, <https://www.futurelearn.com/info/courses/english-for-study-intermediate/0/steps/35241>.