Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra Kybernetiky

BAKALÁŘSKÁ PRÁCE

Automatické rozpoznávání typů otázek v různých jazycích

PLZEŇ, 2023 TEREZA KAŠÁKOVÁ

Prohlášení

Předkládám tímto k posouzení a obhajobě bakalářskou práci zpracovanou na závěr studia na Fakultě aplikovaných věd Západočeské univerzity v Plzni.

Prohlašuji, že jsem bakalářskou práci vypracoval(a) samostatně a výhradně s použitím odborné literatury a pramenů, jejichž úplný seznam je její součástí.

V Plzni dne 11. srpna 2023 .....

podpis

Poděkování

Ráda bych poděkovala Ing. Markétě Řezáčkové za odborné vedení, cenné rady a připomínky, které vedly k dokončení této práce. Dále bych ráda poděkovala MetaCentru za poskytnutí výpočetních zdrojů - Výpočetní zdroje poskytl projekt e-INFRA CZ (ID:90254), podpořený Ministerstvem školství, mládeže a tělovýchovy ČR.

Abstrakt

Tématem této práce je automatické rozpoznávání typů otázek v různých jazycích. Práce analyzuje typy vět a typy otázek, zaměřuje se na jejich intonaci a popisuje její pravidla v různých jazycích. Na problematiku se dívá z hlediska syntézy řeči, zvýrazňuje důležitost správného generování prozodie a konkrétně typického tónu různých typů otázek, aby umělá řeč byla co nejpřirozenější. Důležité je tudíž při předzpracování textu před jeho samotnou syntézou automaticky rozlišovat typy otázek. Předkládaná práce se o to snaží využitím pravidel, klasifikátorů a neuronové sítě T5. Tyto přístupy testuje pro data v několika jazycích a porovnává jejich přesnosti při klasifikaci otázek.

Klíčová slova

TTS, syntéza řeči, zpracování přirozeného jazyka, intonace, otázky, SVM, T5

Abstract

The subject of this thesis is the automatic recognition of question types in various languages. The thesis analyzes types of sentences and types of questions, focuses on their intonation and describes its rules in different languages. The issue is looked at from the perspective of speech synthesis, it emphasizes the importance of the correct generation of prosody and, in particular, the typical tone of different types of questions, so that artificial speech is as natural as possible. It is therefore important to automatically distinguish between the types of questions when pre-processing the text before the synthesis. The current work aims to do so using rules, classifiers and the T5 neural network. These approaches are tested using data in several languages and their accuracy values for question types classification are compared.

Keywords

TTS, speech synthesis, natural language processing, intonation, questions, SVM, T5

Obsah

Seznam obrázků

Seznam tabulek

Seznam ukázek

Kapitola 1

Úvod

Syntéza řeči byla stvořena jako prostředek pro zpříjemnění a zjednodušení komunikace mezi lidmi a stroji. Řeč je odjakživa nejpřirozenějším způsobem dorozumívání mezi lidmi, proto začalo být experimentováno i s její syntézou stroji, od mechanických strojů z měchů až po dnešní podobu řeči, v zásadě vždy generovanou pomocí počítačů.

Dnešní podoba syntetizované řeči je obecně na vysoké úrovni, co se týče srozumitelnosti, ale v případě přirozenosti je stále co zlepšovat. Jelikož přirozenost komunikace je hlavní cíl syntetizované řeči, je dnešní vývoj zaměřen právě na přirozenost.

Syntetická řeč se tvoří na základě fonetické a prozodické informace, přirozenost se odvíjí hlavně od té prozodické. Hlavními prozodickými charakteristikami jsou intonace, časování a intenzita. Intonace jako jediná napomáhá nejen přirozenosti, ale má jasná pravidla i podle významu promluvy, tudíž napomáhá i srozumitelnosti. To umocňuje nutnost generovat intonaci v syntetické řeči korektně.

Intonace mimo jiné udává hlavní rozdíl mezi zjišťovacími otázkami (otázky, které očekávají odpověď ano/ne) a doplňovacími otázkami (očekávají slovní odpověď). Tón na konci zjišťovacích otázek stoupá, aby se výrazněji odlišily od oznamovacích vět, které mají často stejný slovosled. Doplňovací otázky toto výrazné rozlišení nevyžadují, a proto jejich tón klesá.

Při syntéze řeči z textu je nutné, aby intonace byla generována pro každý typ věty dle daných pravidel konkrétního jazyka, jinak by působila velmi nepřirozeně - například pokud by byla řečena zjišťovací otázka s klesavou melodií, zněla by jako oznamovací věta, a nebylo by jasné, že je očekávána odpověď. Kvůli tomu je potřeba zjistit pro každou větu, o jaký typ věty se jedná. To, že se jedná o otázku, lze snadno zjistit podle otazníku na konci věty. O jaký konkrétní typ otázky se jedná, je však komplexnější problém. Přístupy k jeho řešení jsou tématem této bakalářské práce.

Kapitola 2

Syntéza řeči

Dle [14] je syntéza řeči definována jako proces umělého vytváření řeči. Řeč je nejpřirozenějším a historicky nejstarším způsobem komunikace mezi lidmi, proto začaly vznikat způsoby, jak by mohly s člověkem na stejné úrovni komunikovat i stroje. Cílem syntézy je generovat řeč, která bude nejen srozumitelná, ale i přirozená, a tím téměř nerozeznatelná od promluvy člověka. Aplikací syntézy řeči je mnoho, patří mezi ně pomůcky pro nevidomé nebo automatické předčítání textů. Syntetizovaná řeč může být využita i jako menší součást hlasových dialogových systémů nebo robotů.

2.1 Základní typy syntézy řeči

Syntézu [14] dělí podle toho, jakým způsobem je řeč modelována.

Artikulační syntéza se snaží modelovat proces vytváření řeči přesně tak, jak fyziologicky probíhá u člověka. Zahrnuje modely hlasivek i artikulátorů a jejich pohybů. Zároveň počítá se všemi omezeními dynamiky řečových orgánů. Kvůli této komplexnosti dochází k vyšší výpočetní náročnosti a analýza i syntéza je složitější. Přístup proto nebývá příliš využívaný.

Formantová syntéza využívá teorii zdroje a filtru. Používá model řeči, který má dvě složky - zdroj buzení, který modeluje dechové ústrojí a vytváří periodický signál pro znělé zvuky a náhodný šum pro neznělé, a akustický filtr, který modeluje hlasový trakt, a jak artikulační ústrojí mění proud vzduchu na zvuk, jeho parametry určují formanty hlasového traktu člověka. Po dlouhou dobu byl nejvyužívanějším přístupem, dokud nebyl nahrazen konkatenační syntézou. Nejčastěji je úkon realizován se sadou pravidel, která převede fonetickou informaci na vstupu (hlásky + prozodie) na parametry pro syntetizér. Kvůli malému množství používaných pravidel a parametrů je obtížné plně vystihnout variabilitu lidské řeči, proto formantová syntéza nepůsobí přirozeně.

Konkatenační syntéza byla do nedávné doby nejpoužívanější přístup. Je založena na reprezentaci řeči pomocí konečného počtu řečových jednotek. Ty jsou uloženy v inventáři řečových jednotek. Řečové jednotky mohou být různé, dbá se na to, aby vystihly co nejvíce koartikulačních jevů, daly se bez problému řetězit a nedocházelo k nespojitos-

tem. Přitom je výhodnější mít inventář co nejmenší, což vede k využívání menších jednotek na úrovni hlásek, nebo i menších. Fonémy jsou nejmenší zvukové jednotky, jsou souhrnem všech způsobů, jak hlásku vyslovit, kvůli její poloze mezi jinými hláskami. Jednotky lze i kombinovat, ale běžně lze užívat:

- Věty, fráze, slova
- Slabiky, demislabiky
- Fonémy
- Difony, trifony, polyfony
- Subfonémové jednotky

Proces syntézy pak spočívá v řetězení (konkatenaci) jednotek v plynulou řeč. Na rozdíl od předchozích způsobů nemodeluje proces vzniku reálné řeči. Konkatenační syntéza bez modifikací řetězí jednotky v přesném tvaru, jak jsou uloženy, syntéza s modifikacemi naopak upravuje spektrální a prozodické charakteristiky. Tato metoda používá segmenty reálné řeči, kopíruje tím hlas konkrétního řečníka, zároveň je přirozenější a kvalitnější.

Mimo tyto tradiční techniky syntézy řeči se v dnešní době nejčastěji využívají nejrůznější **neuronové sítě** s různými architekturami, které také nemodelují hlasový trakt člověka. Jsou trénovány z nahrávek lidské řeči a jejich fonetických přepisů. Při použití dat od jednoho řečníka síť kopíruje jeho hlas, což přidává přirozenosti. Často se však kvůli nutnosti velkých objemů dat využívají nahrávky více řečníků - buď jen k předtrénování, nebo pro kompletní trénování. [9]

2.2 Syntetizér řeči

Syntetizér řeči se obecně říká systému, který vytváří řeč. Na vstupu pro vytváření řeči je fonetická informace - udává, co má být vytvořeno pomocí posloupnosti fonémů/hlásek a prozodická informace - určuje, jak se má řeč vytvořit pomocí informace o melodii, časování a intenzitě. [14]

Výsledkem prvních historických pokusů byly různé mechanické syntetizéry. Christian Kratzenstein začal vytvářet samohlásky pomocí akustických rezonátorů již v roce 1779. V roce 1791 bylo navrženo Wolfgangem von Kempelenem mechanické zařízení napodobující činnost artikulačních orgánů člověka a je považováno za první syntetizér řeči.

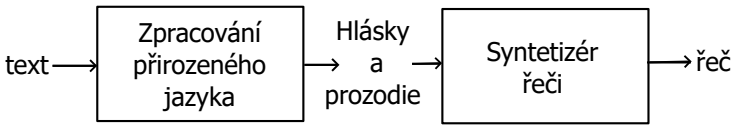
Od 20. let 20. století začaly vznikat elektronické syntetizéry. První zařízení (1922) používalo bzučák jako buzení dvou rezonančních obvodů modelujících první dva formanty hlasového traktu. Dokázalo tím produkovat zvuky jednotlivých souhlásek. Po přidání třetího formantu japonskými vědci v roce 1932 se zvýšila srozumitelnost. V roce 1939 vznikl první elektronický syntetizér schopný syntetizovat souvislou řeč, Voder. Jeho struktura je základem současných syntetizérů založených na teorii zdroje a filtru. V roce 1958 byl sestrojen i první artikulační syntetizér, DAVO.

S nástupem číslicových počítačů v polovině šedesátých let 20. století začaly být syntetizéry simulovány pomocí počítačů, vznikly digitální syntetizéry. První digitální syntetizér snažící se o konkatenační syntézu (popsanou v sekci 2.1) byl vytvořen v roce 1968 a v roce 1977 byl navržen syntetizér, který aktivně používal difony pro konkatenační syntézu, jak je známá dnes. [14]

Po výrazném vývoji počítačů v 90. letech přišel rozvoj neuronových sítí a jejich využívání pro nejrůznější úlohy. Do popředí se dostaly díky jejich schopnosti zpracovat obrovské objemy dat. Začaly se využívat i pro syntetizéry řeči, hlavně pro syntetizéry, které by tvořily řeč přímo z textu, aktuálně je nejznámější WaveNet[10], existují však i jiné, experimentující např. s GAN architekturou. [3]

2.3 Zpracování přirozeného jazyka pro TTS systémy

Syntetizér řeči vyžaduje kompletní fonetický přepis promluvy s informací o prozodii. To je však příliš složitý úkol pro běžného uživatele. Z tohoto důvodu vznikly TTS systémy (text-to-speech). [14] Jejich úkolem je vytvořit řeč z libovolného psaného textu bez jakékoliv další informace. Proto je nutné text před samotnou syntézou nejprve zpracovat a z textu odvodit fonetickou a prozodickou informaci automaticky. To je jedním z úkolů oboru zpracování přirozeného jazyka (angl. Natural language processing - NLP). Proces je znázorněn schématem na Obrázku 2.1.



Obrázek 2.1: Schéma TTS systému [14]

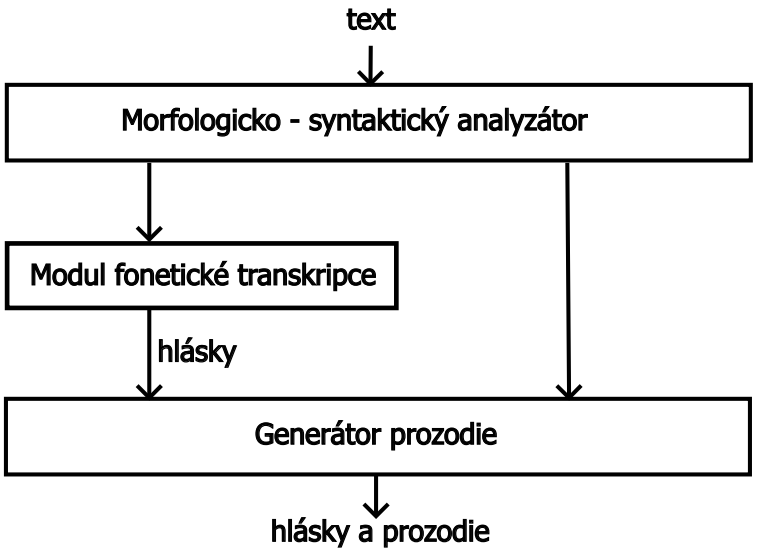
Obecné TTS systémy jsou výrazně odlišné od systémů, které využívají syntézu řeči z omezené oblasti. Ta se dle [8] využívá pro dialogové systémy, které jsou omezené na konečný počet vět nebo frází z dané oblasti. Syntéza pak často spočívá jen v dosazení konkrétních slov do předdefinovaných vět typu:

Vlak číslo ___ přijede v ___ hodin ___ minut na nástupiště číslo ___.

Taková syntéza působí velmi přirozeně, protože je složena z celých nahraných slov. Je pak ale nemožné vygenerovat jakoukoliv řeč mimo zadané věty. Proto je vhodné i syntézu řeči z limitované oblasti doplnit obecným TTS systémem a tím se zvyšuje nutnost jejich existence.

Syntetizér potřebuje vědět posloupnost fonémů a prozodické značky. Ty se většinou získávají ve dvou různých blocích - modul fonetické transkripce a generátor prozodie.

Kvůli možné nejednoznačnosti při určování těchto informací je často v modulu zpracování přirozeného jazyka obsažen i morfologicko-syntaktický analyzátor. Schematicky viz Obrázek 2.2.



Obrázek 2.2: Detail modulu zpracování přirozeného jazyka [14]

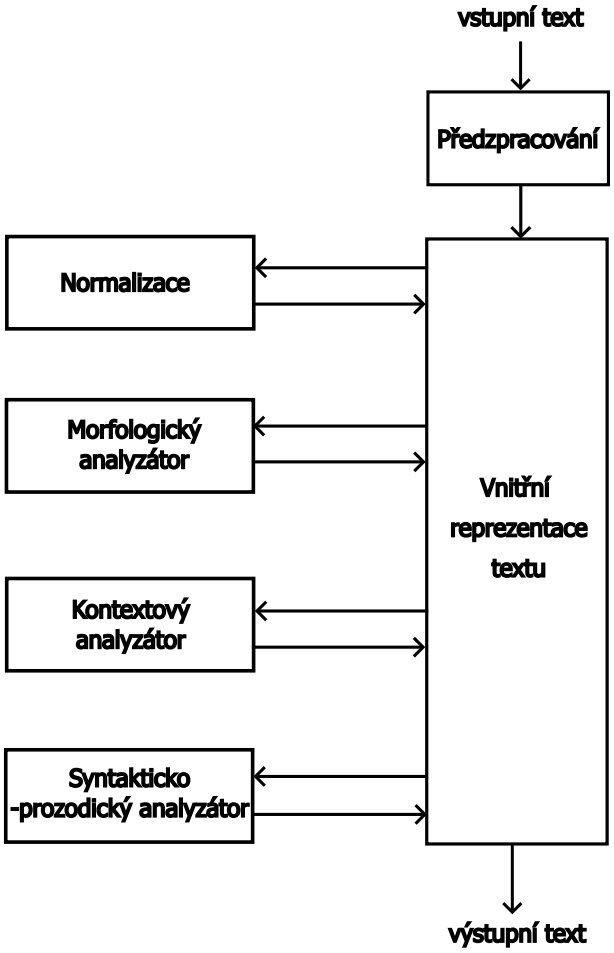
2.3.1 Morfologicko-syntaktická analýza

Morfologicko-syntaktická analýza odhaluje vnitřní struktury vět. [14] Fonetická transkripce i prozodie jsou značně závislé na syntaxi dané věty, je nutné znát vztahy mezi slovy. Jednotlivé komponenty analyzátoru bývají zřetězeny paralelně, aby mohly vzájemně využívat jimi produkované informace. Obrázek 2.3 znázorňuje tento proces.

Předzpracování slouží jako interface mezi vstupním textem a jeho vnitřní organizací v morfologicko-syntaktickém analyzátoru. Detekuje typ vstupního textu, filtruje přebytečné znaky textu a může dělit text na menší díly jako odstavce nebo věty. Detekce konců vět je důležitá především z hlediska intonace, detekce pouze pomocí interpunkce však může být nejednoznačná, jelikož například za některými zkratkami se mohou vyskytovat tečky.

Blok **normalizace** přepisuje text do plné slovní formy. Problematické mohou být:

- číslovky - obyčejné číslovky, časové údaje, obnosy peněz nebo telefonní čísla
- zkratky a akronymy - USA = Spojené státy americké x NATO = zůstává jako NATO
- symboly - % = procent.



Obrázek 2.3: Schéma morfologicko-syntaktického analyzátoru [14]

Při **morfologické analýze** se zjišťují všechny možné mluvnické kategorie, aniž by se bral v potaz kontext okolních slov. Dělí slova na ohebná a neohebná, u ohebných detekuje morfémy (předpony, přípony, koncovky a kořeny slov) za účelem zjištění, zda se jedná o stejná slova v různém tvaru.

Během **kontextové analýzy** se uvažují slova jako součást okolního kontextu. Tím se získají přesnější odhady mluvnických kategorií slov, ze kterých může vybírat. Pravděpodobnostní metody určují pravděpodobnosti mluvnických kategorií pro dvě slova (pomocí n-gramů nebo neuronových sítí). Nepravděpodobnostní metody využívají ano/ne pravidla, pomocí nichž přijmou nebo odmítnou danou kombinaci mluvnických kategorií, jsou založeny na klasifikačních a regresních stromech.

Syntakticko-prozodický rozbor probíhá tak, že hledá ve větách úseky, kde by mělo dojít k prozodickým modifikacím. Nejčastěji tak hledá ve větě fráze a hranice mezi nimi. Metody se dělí na tři skupiny - ručně odvozené heuristiky pomocí pozic interpunkce, metody používající rozšířené bezkontextové gramatiky a metody korpusově orientované, které využívají automaticky odvozená pravidla z rozsáhlých označkovaných textů.

2.3.2 Fonetická transkripce

Vstupní informací fonetické transkripce je text po morfologicko-syntaktické analýze. Úkolem fonetické transkripce je přesný a jednoznačný zápis mluvené řeči pomocí fonetické abecedy. Počet symbolů abecedy souvisí s přesností popisu. Přidáním významných alofonů (různé způsoby řečení fonému, hlavně kvůli vlivu okolních fonémů) se transkripce velmi zpřesní. Fonetická transkripce se může provádět ručně, ale z hlediska syntézy je vhodnější automatický postup. Dle [14] lze řešit pomocí dvou způsobů:

- fonetický slovník - uchovává vždy slovo v textové podobě a fonetické podobě; pro jazyky, kde existuje mnoho tvarů stejného slova, může být slovník příliš rozsáhlý, proto je výhodnější používat na jazyky, kde neexistuje skloňování
- fonetická transkripční pravidla - tento přístup využívá obecná pravidla, která jsou buď určena ručně expertem, nebo automaticky vyvozena z trénovacích dat; vhodnější pro jazyky se skloňováním

Často se využívá kombinace těchto metod, například pro češtinu je nejčastěji využit pravidlový přístup, ale výjimky a slova cizího původu jsou uchovávány ve slovníku.

Pro fonetický přepis lze využít i neuronové sítě, např. existují pokusy o G2P (grapheme-to-phoneme) konverzi založené na neuronové síti T5 - T5G2P. [17] Tento model v češtině dosahuje téměř stejných výsledků jako pravidlový přístup využívající zhruba 100 pravidel, který navíc obsahuje zhruba 170 000 výjimek ve slovníku.

2.3.3 Generování prozodie

Generování prozodie v syntetizované řeči napomáhá hlavně přirozenosti dané promluvy. [14] Hlavní prozodické charakteristiky jsou intonace, časování a intenzita. Automatické generování prozodie je složitá úloha, protože prozodie je částečně nezávislá na textu promluvy. Pro generování je využito dělení na fráze ze syntakticko-prozodického rozboru.

Intenzita se generuje na úrovni hlásek. Určuje se pomocí řečových korpusů, hledí se na pozici hlásky ve slově i v celé větě, rozlišuje se mezi přízvučností a nepřízvučností. V češtině se intenzita považuje za nejméně důležitou prozodickou charakteristiku, lze ji často zcela vynechat.

Generování časování zahrnuje generování trvání, pauz a i přízvuku. Trvání se generuje pro jednotlivé fonémy za pomoci pravidel nebo ze statistik z řečových korpusů. Pauzy jsou umisťovány na hranice frází, využívá se syntakticko-prozodická struktura věty. Pauzy mohou být kratší nebo delší podle toho, jak silně na sebe fráze navazují. Přízvuk se vytváří změnou tónu, intenzity a časování zároveň. Pro slovní přízvuk je nutné nalezení přízvučné slabiky. U češtiny je to například vždy první slabika, což úlohu zjednodušuje. Obecně pro všechny jazyky ale toto pravidlo neplatí.

Intonace je dána výškou hlasu. Její generování má dva kroky, jako první se k fonetické transkripci v některých aplikacích doplní symbolický popis prozodie (v jiných např. požadovaný průběh základní hlasivkové frekvence), a druhým krokem je generování intonace pomocí zvoleného intonačního modelu. Podrobněji se intonací zabývá kapitola 3.

Kapitola 3

Intonace

Intonace je v užším smyslu definována dle [19] jako melodická změna. Jde o přenesený význam melodie v hudbě, určuje sled tónů. Tón je přitom elementární periodický stabilní zvuk, jeho frekvencí je určena jeho výška. Řeč se skládá ze znělých a neznělých zvuků, pro znělé kmitají hlasivky a vytvářejí tóny, frekvenci těchto kmitů se říká základní hlasivková frekvence, značí se f0. U neznělých zvuků ke kmitání nedochází, tudíž po vykreslení základní hlasivkové frekvence do grafu by mělo na jejich místech docházet k nespojitostem.

Melodie může mít v promluvě podle [19] několik funkcí:

- **Lexikální** - V tónových jazycích může existovat více formálně stejných slov, která se můžou lišit pouze tónem, s jakým se vysloví určitá slabika. Tón tak ovlivňuje význam promluvy. Tónovým jazykem je například čínština [4], tóny jsou určeny průběhem jako:

  1. vysoce položený a rovný
  2. ostré stoupání ze střední polohy do vysoké
  3. hluboký tón následovaný uvolněním hlasivek stoupnutím
  4. klesání z vysoké polohy do hluboké

  Těmito tóny jsou pak rozlišitelná slova, která jsou jinak pomocí pinyinu (latinkového přepisu znaků) zapsaná až na daný tón stejně, ale mají různé významy:

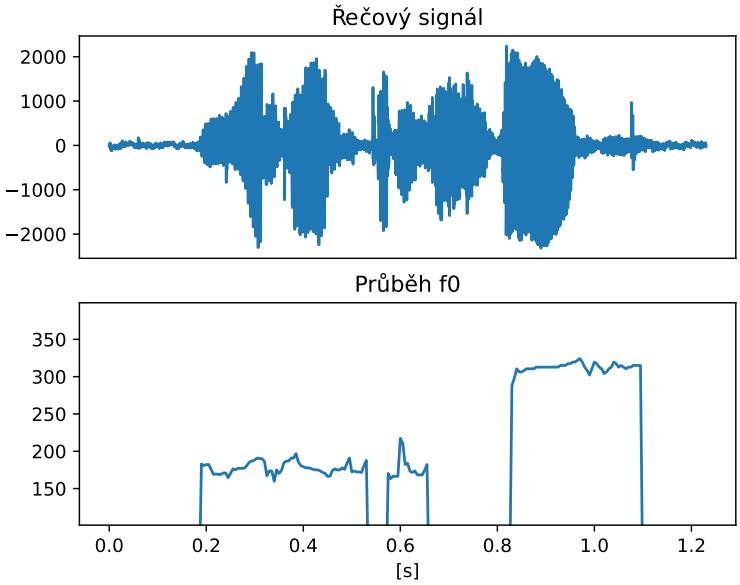
  1. mā - maminka
  2. má - konopí
  3. mǎ - kůň
  4. mà - nadávat

- **Větně významová** - Naznačuje typ věty, který není jednoznačný ze slovní struktury věty, tedy především rozdíl mezi oznamovací větou a zjišťovací otázkou o stejné struktuře, například:

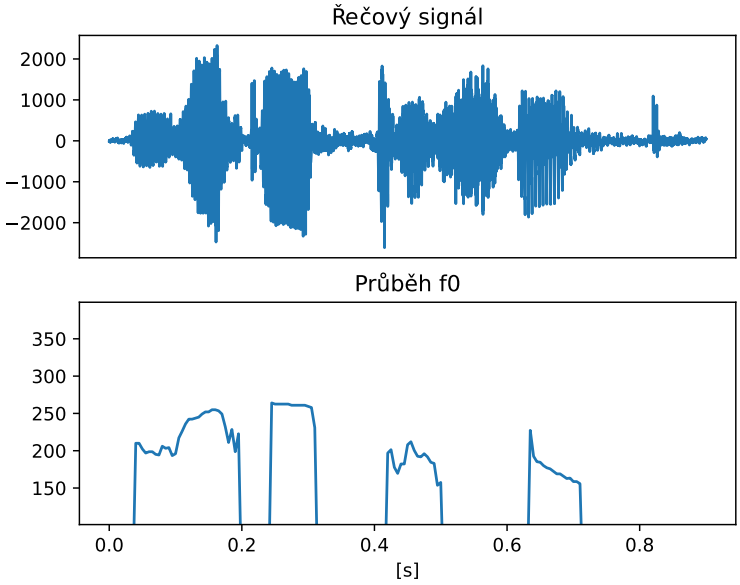
– Bude pršet? ↗

– Bude pršet. ↘

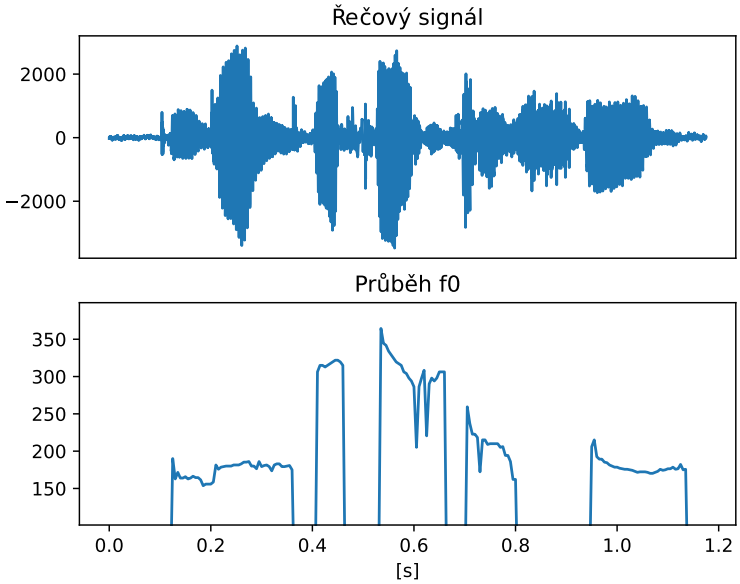
Detailně lze tento rozdíl v průběhu f0 vidět na obrázcích 3.1 a 3.2. Pro ještě bližší srovnání i s doplňovací otázkou je na obrázku 3.3 znázorněn průběh věty „Kdy bude pršet?“.



Obrázek 3.1: Průběh intonace zjišťovací otázky - „Bude pršet?“



Obrázek 3.2: Průběh intonace oznamovací věty - „Bude pršet.“



Obrázek 3.3: Průběh intonace doplňovací otázky - „Kdy bude pršet?“

- **Členící** - Pomáhá dělit věty na skupinky slov, které mají těsnější vzájemný vztah. Typické je klesnutí hlasem v oblasti mezi dvěma větami souvětí nebo využití nižšího intonačního rozpětí pro vložené věty, které nesou pouze doplňkovou informaci.
- **Diskurzní** - Melodie dává informaci, zda člověk plánuje pokračovat v řeči, nebo již očekává odpověď od druhého účastníka dialogu.
- **Afektivní** - Určuje, v jakém rozpoložení se mluvčí nachází z hlediska emocí, nálad, postojů nebo vlastností. Tyto skutečnosti ovlivňují kromě tónu i tempo, sílu zvuku a barvu hlasu.
- **Indexová** - Některé charakteristiky tónu a intonační kontury mohou příslušet právě jednomu řečníkovi, nebo určovat jeho příslušnost do sociální skupiny, jeho geografický původ, věk nebo status.

Dle [11] lze ale funkce tónu určit i zjednodušeně jako:

- Rozlišení ukončené a neukončené výpovědi
- Odlišení výpovědních typů
- Rozlišení neutrální od citově zabarvené výpovědi

3.1 Typy vět a jejich intonace v češtině

Podle záměru mluvčího dělí [20] věty na:

1. **Oznamovací** - vyjadřují sdělení nebo informaci, jsou ukončeny tečkou
   - *Otec přinesl zprávu.*
2. **Tázací** - vyjadřují otázku, jsou ukončeny otazníkem
   - Zjišťovací - otázka, která očekává odpověď ano/ne
     - *Přijdeš k nám?*
   - Doplňovací - otázka, která očekává slovní odpověď, využívá tázací zájmena, v angličtině se podle nich označují jako Wh-questions
     - *V kolik hodin přijedeš?*
   - Vylučovací - otázka, která dává na výběr z více možností pro odpověď
     - *Přijdeš k nám dnes, nebo zítra?*
   - Rozvažovací - vyjadřují osobní nejistotu
     - *Co dělat?*
   - Řečnické - nevyžadují odpověď, často se předpokládá záporná
     - *Může se nad tím někdo pozastavovat?*
3. **Žádací**
   - Rozkazovací - vyjadřují rozkaz nebo zákaz, ukončeny tečkou nebo vykřičníkem
     - *Zavři dveře.*
   - Přací - vyjadřují přání nebo citově zabarvenou žádost, jsou ukončeny nejčastěji vykřičníkem
     - *Kéž by se vrátil!*

Věty se mohou vyskytovat i ve formě zvolací, jsou dány velkým citovým zabarvením, vždy jsou ukončeny vykřičníkem. Tuto formu mohou mít věty, které by jinak byly označeny za oznamovací, tázací, i žádací.

Melodická část komunikace je poskládána z malých elementárních kontur, které mohou být v různých intonačních rozpětích. Rozlišují se tři melodémy: melodém ukončující klesavý, melodém ukončující stoupavý a melodém neukončující. [19] Přesněji pak může být melodém určen různými kadencemi.

Pro melodém ukončující klesavý:

- kadence klesavá
- kadence stoupavo-klesavá

Pro melodém ukončující stoupavý:

- kadence stoupavá
- kadence stoupavo-klesavá
- kadence rovno-klesavá

[11] rozlišuje kadence na příznakové a nepříznakové. Délka kadence bývá proměnlivá. Ukončující melodémy se realizují v posledním taktu výpovědi kvůli větnému přízvuku, který je umístěn na stejném taktu. Pokud je větný přízvuk jinde, realizuje se jinde i melodém. Při rozšíření intonačního rozpětí je kadence vnímána jako důrazová, používá se buď kvůli vyjádření emocí, pro kontrast, nebo určuje strukturu věty.

Oznamovací věty využívají melodém ukončující klesavý, realizují ho kadencí klesavou prostou. Doplňovací otázky využívají melodém ukončující klesavý, kadence přitom začíná na tázacím výrazu, které jsou typické právě pro doplňovací otázky (kdo, co, jaký,...). Zjišťovací otázky, které očekávají odpověď ano/ne, tyto výrazy nemají, proto je potřeba určit, že se jedná o otázku, pomocí tónu. Pro ně je využit melodém stoupavý.

Pravidla rozšířená v jednom jazyce nemusí být pravidlem v jiných, [2] se zabývá popisem intonace v různých jazycích, v čem jsou si podobné a v čem se liší.

3.2 Intonace v angličtině

Ačkoliv fonetika rozlišuje mezi americkou a britskou angličtinou a [2] popisuje obě zvlášť, rozdíly jsou v zásadě minimální. Americká angličtina je silně spojena s výraznější gestikulací a více změnami výrazu obličeje při vyšším tónu v řeči. Zároveň nemá slovní a segmentální tón. Hlavní funkce základní frekvence jsou tvorba přízvuku a kontur.

V angličtině nemá intonace vztah s gramatikou. Mnoho otázek končí stoupavým melodémem, často je to jediný příznak toho, že jde o otázku. Wh-otázky ve většině případů klesají. Zjišťovací otázky zhruba stejně často stoupají i klesají, existují však typy, které stoupají téměř vždy. Jsou to otázky, které opakují jinou otázku kvůli ujištění, že dotyčný správně slyšel. Otázky jsou pak celkově výše položené než jiné věty.

Věty zvolací jsou položené vysoko a na konci klesají. Oznamovací věty můžou přinášet nové téma, nebo odpovídat na otázku, oba typy využívají klesavý melodém, ale při odpovědi na otázku se spíše využije stoupavo-klesavá kadence. Rozkazy běžně vnáší nové téma, tudíž klasicky klesají, ale při opakovaném rozkazu se může kontura obrátit.

[13] se dále zabývá otázkami, které se dají vyložit dvěma způsoby. Konkrétně mezi vylučovacími a zjišťovacími, ve kterých je obsaženo slovo nebo. Následující otázky vypadají formálně stejně, liší se pouze tónem.

- Are you allergic to dairy or soy? ↗
- Are you allergic to dairy or soy? ↘

První případ se ptá na to, zda je dotyčný alergický, nezáleží v tom případě na co, očekává se odpověď ano/ne, tudíž se tónem stoupá. V druhém případě už víme, že dotyčný na něco alergický je, ale ptáme se, na kterou konkrétní věc, tudíž se více blíží doplňovací otázce a klesá. V češtině se rozdíl mezi spojkami nebo ve slučovacím a vylučovacím poměru rozlišuje čárkou před slovem nebo.

3.3 Intonace v němčině

Vědci argumentují, že běžná rytmická struktura angličtiny se s menšími úpravami dá aplikovat i na němčinu. [2] Tvrzení, že oznamovací věty používají klesavý tón, doplňovací otázky stoupavý a W-otázky (německá verze Wh-otázek) klesají, jsou považována za přílišné zjednodušení, používají se:

1. kadence klesavá - oznamovací věty a věty zvolací
2. kadence stoupavá - neukončená výpověď, otázky
3. kadence rovná - očekává se pokračování promluvy
4. kadence klesavo-stoupavá - nejistota
5. kadence stoupavo-klesavá - jistota, očividnost

Tento výčet však stále nepokrývá celou variabilitu tónu v hovorové němčině. Tabulka 3.1 znázorňuje používané melodémy podle typu věty a za jakým účelem je daná promluva řečena. Němci zároveň využívají mnohem užší frekvenční rozsah než jiné jazyky, tudíž promluva bývá celkově monotónnější.

Typ věty	Melodém	Účel
Oznamovací	klesavý	tvrzení
	stoupavý	nejisté tvrzení
Rozkazovací	klesavý	rozkaz
Přací	stoupavý	přání
Tázací (ano/ne)	klesavý	rázná otázka
	stoupavý	běžně
W-otázky	klesavý	běžně
	stoupavý	zopakovaná otázka

Tabulka 3.1: Účel využití melodémů v typech vět v němčině

Článek [5] pak doplňuje, že při dopředu známé odpovědi na ano/ne otázku se využívá klesavého melodému.

3.4 Intonace ve španělštině

Studií přibližujících prozodické charakteristiky španělského jazyka je méně a často si odporují. [2] Slova se dělí na přízvučná a nepřízvučná, přízvučná tvoří zhruba 63 % mluveného jazyka. Systém intonace se zakládá na jejich rozložení ve větě. Každá věta má tolik tónických skupin, kolik má slabik s přízvukem. Průběh je nejčastěji znázorňován úrovněmi danými transkripčním systémem INTSINT, od nejvyšší k nejnižší jsou to:

- T (Top), H (Higher), U (Upstepped), S (Same), M (Mid), D (Downstepped), L (Lower), B (Bottom)

Krátké oznamovací věty jsou tedy například dány sekvencí T,H a D nebo B - tj. jejich tón postupně klesá. Tato písmena vždy popisují úsek přízvučné slabiky. Značí tedy postupně klesavou intonaci.

Intonace má velkou roli v určování dvou typů tázacích vět. Ty, které naznačují gramatickou strukturou, že jsou otázkami (například existencí tázacích výrazů), na konci klesají a předchozí tóny jsou také na nízké úrovni, tím se podobají oznamovacím větám. Zjišťovací otázky mají naprosto stejný slovosled a strukturu jako oznamovací věty, proto se jejich tón musí lišit a narozdíl od oznamovacích vět stoupá. Některé otázky se mohou odklánět od těchto pravidel, pokud vyjadřují zdvořilost, opakování otázky nebo ujištění se o nějaké skutečnosti.

3.5 Intonace v ruštině

Prozodie ruštiny tvoří dva systémy, systém slovního přízvuku a systém větné intonace. [2] Prozodie zejména seskupuje zvuky a slabiky do vyšších celků. Občas může být prozodie jediné, co naznačuje účel výpovědi - otázka, neukončenost, kontrast. Přízvuk je v ruštině volný, může být na kterékoliv slabice, ale přesto existují pravidla na základě slovních druhů. Slovní prozodie je také velmi ovlivněna prozodií vět. Ruština seskupuje slova do větších celků pomocí přízvuku na prvním slově.

Samostatně stojící oznamovací věta je dána sérií stoupání a poklesů tónu, na konci věty na přízvučné slabice výrazně klesá. V souvislé řeči nebývá konečný tón úplný, v místě přízvuku může stoupat. Přízvuk také způsobuje nevýraznější melodii ostatních slov.

Otázky jsou pronášeny rychleji než oznamovací věty. Zjišťovací otázky nepoužívají tázací zájmena a slovosled je zcela volný, otázka je definována ostře rostoucím tónem na přízvučné slabice hlavního přízvučného slova a pokles na následujících slabikách. Speciálním případem jsou eliptické (neukončené) otázky začínající slovem a, např. „A ty?“, které jsou tvořeny stoupavým nebo klesavě-stoupavým tónem.

Tón doplňovacích otázek se dle některých názorů shoduje s tónem oznamovacích vět. Vyznačuje se však větším stoupáním před finálním klesnutím. Vzácnější typ doplňovacích otázek je charakterizován přízvukem na tázacím zájmenu a postupným ustálením.

Intonace rozkazovacích a zvolacích vět je příliš komplikovaná a její pravidla se nedají vyjádřit obecně.

3.6 Shrnutí

Intonace je nejdůležitější prozodická charakteristika, která je dána změnou melodie hlasu. Celková intonační kontura je poskládána z menších částí, tzv. melodémů, které se mohou dále realizovat různými kadencemi. Intonace je často provázána s přízvukem, hladina základního hlasivkového tónu se mění na přízvučných slabikách slov. Pravidla přízvuku se mohou v různých jazycích lišit velmi výrazně, čeština má přízvuk vždy na první slabice slova, ale například ruština má přízvuk volný a může být na kterékoliv slabice. Intonace umožňuje rozlišení mezi slovy v tónových jazycích, člení souvětí na věty klesnutím hlasem mezi předěly vět, ukončuje promluvu řečníka, určuje emocionální rozpoložení. Hlavně ale může rozlišovat mezi typy vět, a dále i mezi různými druhy otázek.

Otázky lze ve všech jazycích dělit na zjišťovací a doplňovací. Méně časté jsou otázky eliptické a vylučovací. Zjišťovací otázky očekávají odpověď ano/ne, v naprosto běžných podmínkách mají stoupavou intonaci, ve většině jazyků je jejich slovosled totožný s oznamovacími větami, a musí se proto odlišit mezi sebou výrazným rozdílem v intonaci. Doplňovací otázky očekávají slovní odpověď, obsahují tázací zájmena, která naznačují, jakou odpověď očekávají. Tázacími zájmeny jsou např. kdo, co, jaký, který. Tyto otázky se nezaměňují s oznamovacími větami, mají klesavou intonaci, která však stejně není totožná s intonací oznamovacích vět. Eliptické otázky se vyznačují tím, že nejsou dokončené, většinou neobsahují sloveso. Mohou ale obsahovat tázací zájmena, a tím připomínat doplňovací otázky. Na rozdíl od nich ale tónem stoupají. Vylučovací otázky jsou takové otázky, které dávají účastníkovi dialogu na výběr ze dvou možností. Vyznačují se spojkou „nebo“ ve vylučovacím poměru, tudíž nedávají prostor vybrat obě možnosti zároveň. Intonace na konci takového souvětí klesá.

Od pravidel intonace v otázkách se lidé odklánějí, když chtějí vyjádřit zdvořilost, pokud otázku opakují, nebo tvrzení vyjadřují s nejistotou a chtějí se ujistit o nějaké skutečnosti. Mluvčí pak každou promluvu může říct jinak i na základě různých emocí.

Kapitola 4

Cíl práce

Cílem bakalářské práce je automatické rozpoznání typů otázek pro využití při syntéze řeči z pohledu ukončujícího melodému dané otázky. Ten je důležité zjistit, aby bylo možné generovat prozodii při syntéze konkrétní otázky. Intonace a celkově prozodie pomáhají přirozenosti vytvořené řeči. Rozpoznání by mělo probíhat z obecného textu.

Ačkoliv existuje mnoho výjimek zdůvodněných emocemi, kontextem v plynulé řeči, či různými nářečími, při uvažování naprosto běžné neutrální výpovědi jsou přesně dány zjišťovací otázky stoupavým melodémem a doplňovací otázky klesavým melodémem (který se však ještě dále liší od klesavého melodému oznamovacích vět), a to ve všech jazycích, pro které je cílem práce tyto typy rozpoznávat, jak bylo uvedeno v kapitolách 3.1-3.6.

Pro práci budou využita data v češtině, angličtině, němčině, španělštině a ruštině. Pro zaručení co nejlepších výsledků je navíc nutné nejprve provést ruční korekci anotací tónů podle daných pravidel. Data obsahují převážně zjišťovací a doplňovací otázky, rozlišovat se bude hlavně mezi nimi. V datech je ale obsaženo i minimální množství vylučovacích a eliptických otázek.

Pro rozpoznávání budou využity techniky umělé inteligence a strojového učení. Pro učení s učitelem by měla být data rozdělena do dvou tříd, Td a Tu, určující klesavý a stoupavý tón. Uplatněn bude pravidlový přístup s ručně a automaticky odvozenými pravidly, SVM klasifikátor a neuronová síť Text-To-Text Transfer Transformer používaná ke zpracování přirozeného jazyka, které se zkráceně říká T5.

Kapitola 5

Rozpoznávání typů otázek

Tato kapitola se zabývá jednotlivými přístupy k rozpoznávání typů otázek a jejich dělení do dvou tříd. V sekci 5.1 je nejprve popsáno předzpracování dat k jejich následnému používání. Sekce 5.2 popisuje využití ručně i automaticky vytvořených pravidel. Sekce 5.3 a 5.4 pak k rozpoznávání typů otázek používají klasifikátory a neuronové sítě. Veškeré použité zdrojové kódy a konfigurační soubory jsou součástí přílohy na GitHubu.[1]

5.1 Příprava dat

Před tím, než byla data k čemukoliv využita, bylo nutné sjednotit jejich formát. Česká, anglická a německá data byla ve formátu:

Otázka \t anotace,

viz ukázka 5.1.

```
V čem myslíte, že je Škoda dobrá? Td
Jak přenést dědičnou informaci? Td
Čím se chlubí Praha? Td
```

Ukázka 5.1: Požadovaný formát dat

Anotace byly rozlišeny jako Tu (z angl. „tone up“, značící stoupavý melodém) a Td (z angl. „tone down“, značící klesavý melodém). Tento tabulátorově oddělený formát, známý jako *.tsv, je výhodný hlavně kvůli neuronové síti, která bude očekávat data právě v tomto formátu. Proto byla česká, anglická a německá data zachována v tomto formátu a španělská a ruská data byla převedena do totožného, aby s nimi bylo možné pracovat jednotně.

Španělská data (a i původní ruská data, která však kvůli velmi časté chybovosti anotací nebyla použita) měla formát takový, že každé slovo ve větě mělo za sebou znázorněnou

[1] `https://github.com/Kasakova/BP_kasakova`

fonetickou transkripci v systému IPA a daná anotace (⟨Tu⟩, ⟨Td⟩) byla umístěna na konkrétní pozici v textu/promluvě viz Ukázka 5.2. Fonetická transkripce není pro tuto úlohu důležitá, jelikož jde o zpracování textu, ne jeho fonetické podoby, lze tedy tuto informaci z dat zcela vypustit. Konkrétní pozice poklesu/stoupání tónu také není součástí úlohy, v TTS systému ji bude určovat jiný modul. Konkrétní anotace je však důležitá, proto je nutné ji ze závorek extrahovat a v datech ponechat. Zcela vymazáno bylo číslování řádků na začátku a ponecháním pouze řádků, které obsahují otazník, byly odfiltrovány oznamovací věty.

```
cnv00251 Qué[<T’>'ke] pasa[<Td><T’>'pasa<T?>]?
cnv00252 Qué[<T’>'ke] ha[<T’>'a] pasado[<Td><T’>pa'saðo<T?>]?
cnv00253 Qué[<T’>'ke] es[<Td><T’>'es<T?>]?
```

Ukázka 5.2: Původní formát španělských dat

Nová ruská data měla oproti španělským formát podobnější požadovanému, je zobrazen v Ukázce 5.3, místo *.tsv šlo o *.csv, což je formát velice podobný, jen oddělený středníkem místo tabulátoru. Bylo obráceno pořadí anotace a otázky oproti požadovanému formátu, muselo být proto upraveno na správný tvar.

```
Tu;В стоимость аренды включено топливо?
Tu;Все на ней ездят, а она молчит?
Tu;Может, всё ещё образуется?
```

Ukázka 5.3: Původní formát ruských dat

5.1.1 Korekce dat

Anotace vět nebyly vždy správné, autor této práce proto musel u jazyků, u kterých mu to jazykové schopnosti umožňují, ručně opravit některé anotace. Korekce byla provedena u dat českých, anglických, německých a španělských. V následujících příkladech jsou opravované věty uvedeny již se správnými anotacemi.

- Bylo nutné rozlišit eliptické otázky, které nejsou plnými větami (neobsahují podmět a přísudek), kde tón vždy stoupá.
  - *So what? Tu*
  - *Such as? Tu*
- V německých datech byla u specifických a méně běžných zájmen špatná anotace Tu pro všechny jejich výskyty.
  - *Wieso*
  - *Wieviele, Wievielmal, Wievielter*

  - *Wofür*
  - *Woraus*
  - *Worum*
  - *Wovor*
  - *Wozu*

- Dále bylo nutné rozlišit, kdy tázací zájmena uprostřed souvětí určují doplňovací otázku:

  - *Sám jste zařídil dva góly, jak jste viděl obě situace? Td*

  A kdy jsou jen běžnými spojkami:

  - *Je vydírání běžný způsob, jak zjistit pravdu? Tu*

  Toto rozlišení je zřejmé ve španělštině, kdy tázací zájmena mají přízvuk a spojky ne, například Cual x Cuál. V datech ale toto rozlišení není zachyceno správně, obě verze slov se vyskytují jak na pozici tázacích zájmen, tak i spojek. Před využitím tohoto rozdílu pro rozpoznávání by se musel v datech také opravit.

- Problematické bylo využití vylučovacích otázek, které mají zásadně klesavý tón, i když neobsahují tázací zájmena:

  - *Umoudří se lidé, či naopak dojde k nejhoršímu? Td*

- Některé chyby však byly nesporné a šlo o zcela jasné doplňovací otázky:

  - *Kam jede autobusová linka číslo dvacet? Td*

  Nebo zjišťovací otázky:

  - *Musel být ten spor řešen přes média? Tu*

Pokud by tyto případy byly špatně anotované již v trénovacích datech, klasifikátory by nemusely být schopné správně určit veškerá nově příchozí data, protože se učily z chybných a každá jejich další klasifikace by byla opět chybná. Proto byla oprava dat velmi důležitá a až takto upravená data se mohla využít pro automatické rozpoznávání typů otázek. Tabulka 5.1 znázorňuje počet opravených chyb a jejich poměrné zastoupení v datech.

5.1.2 Rozdělení dat a použitá metrika

Data byla rozdělena na trénovací a testovací v poměru 4:1 pomocí metody z knihovny Sci-kit Learn *sklearn.model_selection.train_test_split()*[12]. Pro průběžnou validaci neuronové sítě a výběr optimálních parametrů klasifikátorů pak byla oddělena od trénovacích dat stejným způsobem (ale v poměru 9:1) navíc validační data.

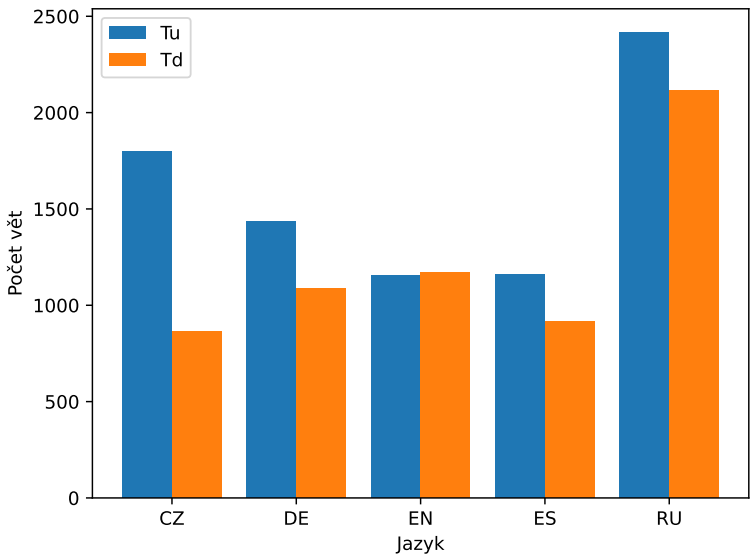
	Počet oprav	Poměr chybných [%]
CZ	21	0.79
DE	55	2.18
EN	47	2.02
ES	21	1.01

Tabulka 5.1: Chybovost anotací

Celkové množství dat a poměr otázek se stoupavým a klesavým tónem, které reprezentují třídy, do kterých se bude dělit, jsou znázorněna v tabulce 5.2. Rozdělení na trénovací, testovací a validační přibližně zachovává původní poměry Tu:Td.

	Počet otázek	Tu	Td
CZ	2661	1796	865
DE	2524	1435	1089
EN	2327	1157	1170
ES	2077	1159	918
RU	4531	2418	2113

Tabulka 5.2: Rozložení dat pro každý jazyk



Obrázek 5.1: Rozložení dat pro každý jazyk

Pro většinu jazyků je poměr dat pro obě třídy srovnatelný a blíží se poměru 1:1. Pro češtinu je už poměr více jak 2:1, ale stále obsahuje dostatečné množství vět i pro tón Td. Je jich ale nejméně ze všech jazyků, i když celkem je dat téměř nejvíce, hned po ruštině. Celkové množství dat pro konkrétní jazyk i počet dat pro jednotlivé typy otázek mohou

mít vliv na trénování klasifikátorů, ale zároveň je důležité, i jaké konkrétní otázky se v datech vyskytují.

Jelikož pro všechny jazyky je k dispozici dostatečné množství dat pro oba tóny, lze považovat za srovnatelnou ztrátu, která je způsobena chybnou klasifikací mezi oběma třídami. Proto lze pro určení přesnosti klasifikace využít metriku accuracy, která určuje poměr počtu správně zaklasifikovaných do obou tříd ku všem proběhlým klasifikacím, podle vzorce 5.1. Význam jednotlivých členů je blíže vysvětlen v obecné matici záměn pro třídy Tu a Td v tabulce 5.3.

- TP = true positive - počet správně zaklasifikovaných Td
- TN = true negative - počet správně zaklasifikovaných Tu
- FP = false positive - počet nesprávně zaklasifikovaných Td
- FN = false negative - počet nesprávně zaklasifikovaných Tu

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

		Klasifikátor	
		Td	Tu
Real.	Td	TP	FP
	Tu	FN	TN

Tabulka 5.3: Obecná matice záměn

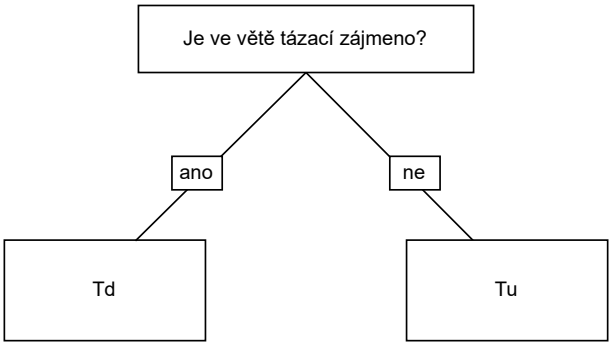
5.2 Znalostní systém

Pro rozpoznávání typu otázky byl nejprve využit pravidlový přístup, který je nejjednodušší, je dán zcela jednoduchými pravidly.

V tomto případě bylo typově využito jedno pravidlo:

Pokud věta obsahuje tázací zájmeno, tón je klesavý.

Tázací zájmena se nejčastěji vyskytují na začátku vět (obvykle na první nebo druhé pozici). Pokud jde ale o souvětí, můžou se vyskytovat uprostřed souvětí, na začátku věty druhé. Úkolem této kapitoly je zjistit, mezi kterými slovy by se mělo tázací zájmeno hledat. Použití tohoto pravidla je reprezentováno rozhodovacím stromem na obrázku 5.2.

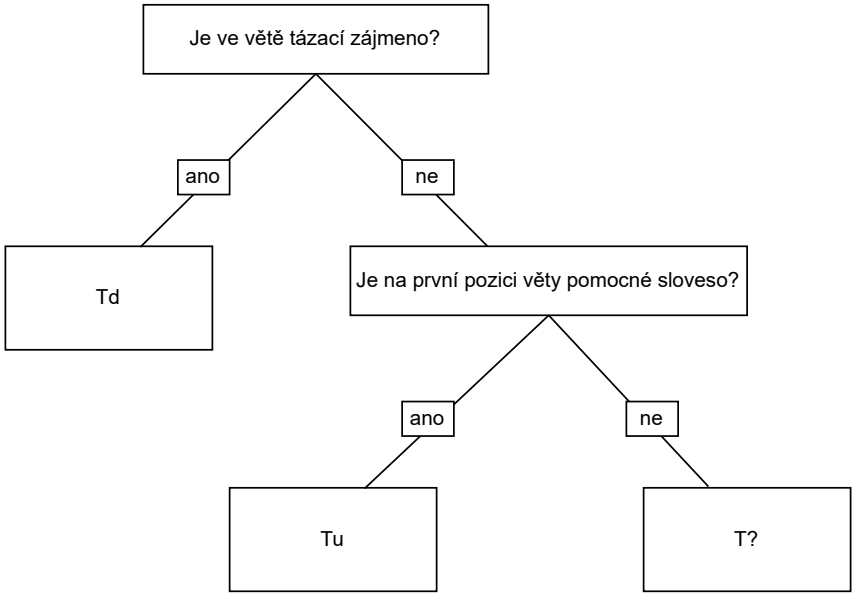


Obrázek 5.2: Rozhodovací strom s jedním pravidlem

Experimentováno bylo však i s druhým pravidlem:

Pokud je na první pozici ve větě pomocné sloveso, tón je stoupavý.,

které však nelze využít obecně pro všechny jazyky; pro angličtinu platí, ale pro češtinu nebo španělštinu nikoliv. Pro postup rozhodování s tímto pravidlem je rozhodovací strom z obrázku 5.2 upraven na tvar na obrázku 5.3.



Obrázek 5.3: Rozhodovací strom se dvěma pravidly

Za tímto účelem však musí být dodán seznam těchto tázacích zájmen, nebo v druhém případě pomocných sloves, ze kterého znalostní systém vybírá.[2] Toho bylo dosaženo dvěma způsoby, ručně u jazyků, u nichž mi to dovoluje jejich znalost, a poté automatickým vyvozením z trénovacích dat. Seznamy byly uchovávány ve formátu *.yaml.

[2]Kompletní konfigurační soubory jsou součástí přílohy na GitHubu.

5.2.1 Ručně odvozené znalosti

Pro znalostní systém byly ručně vytvořeny seznamy tázacích zájmen. Pro angličtinu byl nejkratší a část je zobrazena v ukázce 5.4, v souboru je seznam ještě doplněn o varianty s 's (např. where's) pro prvních 6 zájmen. Ty určují klesavý tón Td.

```
Td:
- where
- how
- who
- what
- why
- when
- which
- whose
- whom
```

Ukázka 5.4: Část konfiguračního souboru pro angličtinu - klesavý tón

```
Tu:
- do
- does
- did
- are
- is
- am
- was
- were
- will
- would
- can
- could
- shall
- may
- have
- has
- had
```

Ukázka 5.5: Část konfiguračního souboru pro angličtinu - stoupavý tón

Během práce bylo experimentováno i s částí konfiguračního souboru, která by určovala stoupavý tón. Hlavně v angličtině jsou na první pozici v zjišťovací otázce zásadně pomocná slovesa viz Ukázka 5.5. V reálném konfiguračním souboru je seznam ještě doplněn

o záporné varianty. Při rozpoznávání se pak můžou porovnávat slova vět s oběma seznamy, nebo jen zvlášť s každou částí. Následující část textu této kapitoly zkoumá, jaká část konfiguračního seznamu by se měla využít pro co nejlepší výsledky. Také zjišťuje, mezi kolika počátečními slovy má tázací zájmena a pomocná slovesa hledat, zohledňuje i variantu, kdy se hledá v celé větě. Pro vyhodnocení se využívají validační data.

Použití obou částí konfiguračního souboru znamená, že některé otázky dostanou klasifikaci Td, některé Tu, ale je nutné zavést kategorii „nezaklasifikováno“, která je reprezentována anotací T?, viz obrázek 5.3. Ta by měla být rozdělena pomocí dalšího pravidla nebo pravidel na Tu a Td.

Následující tabulky zobrazují matice záměn (angl. confusion matrix) pro využití obou částí konfiguračního souboru na určitý počet počátečních slov. Jelikož část konfiguračního souboru pro Tu počítá s pomocným slovesem na první pozici, bylo nejprve hledáno mezi prvními slovy věty viz Tabulka 5.4, kde přesnost byla 90,91 %.

		Klasifikátor		
		T?	Td	Tu
Real.	Td	6	86	3
	Tu	7	1	84

Tabulka 5.4: Matice záměn - EN - 1 slovo Tu, 1 slovo Td

Tázací zájmena pro Td však mohou být na první i druhé pozici viz Ukázka 5.6, takže při hledání Td ve dvou slovech a Tu v jednom slově se dojde k výsledkům v Tabulce 5.5 a přesnosti 93,58 %.

```
Where can I buy tickets? Td
And what happened then? Td
And, by the way, why should the afterlife be everlasting? Td
```

Ukázka 5.6: Poloha tázacích zájmen ve větách

		Klasifikátor		
		T?	Td	Tu
Real.	Td	1	91	3
	Tu	7	1	84

Tabulka 5.5: Matice záměn - EN - 1 slovo Tu, 2 slova Td

Překvapivě se však nejlepších výsledků dosáhne při hledání v prvních dvou slovech věty pro oba tóny (a obě části konfiguračního souboru), jak znázorňuje Tabulka 5.6, a to přesnost 94,12 %.

		Klasifikátor		
		T?	Td	Tu
Real.	Td	1	91	3
	Tu	6	1	85

Tabulka 5.6: Matice záměn - EN - 2 slova Tu, 2 slova Td

Z výsledků je však jasné, že výrazně větší část nezaklasifikovaných má mít tón Tu. To znamená, že pro optimální výsledky lze zanedbat Tu (pomocná slovesa) části konfiguračního souboru a využít jen Td (tázací zájmena) (viz obrázek 5.2), vše ostatní se pak zaklasifikuje jako Tu. Matice je v tabulce 5.7, přesnost stoupne na 97,33 %.

		Klasifikátor	
		Td	Tu
Real.	Td	91	4
	Tu	1	91

Tabulka 5.7: Matice záměn - EN - 2 slova Td

Pro ještě větší přesnost lze ještě započítat nepříliš časté případy, kdy jsou věty započaty nevýznamovým slovem a tázací zájmeno je odsunuto až na třetí pozici typu:

- *A s kým...*
- *Ale na který...*

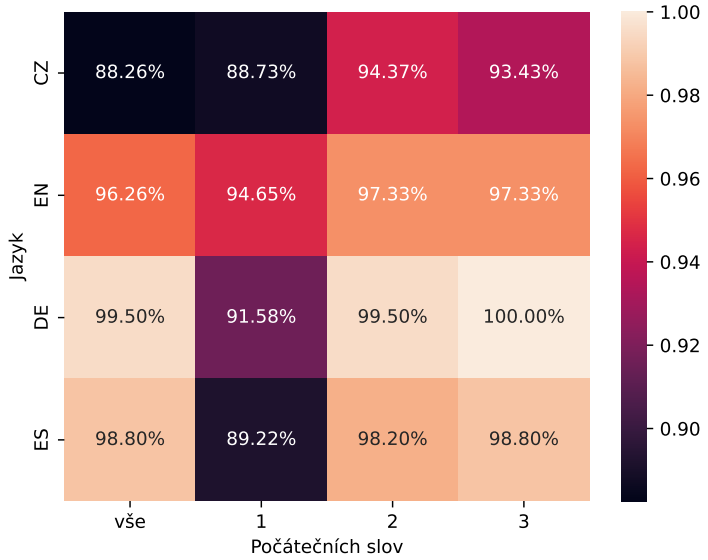
		Klasifikátor	
		Td	Tu
Real.	Td	91	4
	Tu	1	91

Tabulka 5.8: Matice záměn - EN - 3 slova Td

Z tabulek 5.7 a 5.8 je vidět, že pro angličtinu je výsledná přesnost 97,33 % stejná pro hledání zájmen v prvních 2 i 3 slovech. Obrázek 5.4 zobrazuje srovnání přesností klasifikace podle počtu počátečních slov dané věty, ve kterých hledá slova z Td seznamu pro všechny jazyky. Z něj by se dalo usoudit, že hledání ve třech slovech je opravdu výhodnější pro všechny jazyky kromě češtiny, ale na třetí pozici se již často vyskytují slova, která se píší stejně jako hledaná zájmena, ale ve větě mají funkci spojek, například:

- *Máte představu, jak situaci Českých drah zlepšit? Tu,*

proto bylo hledání omezeno jen na první dvě slova.



Obrázek 5.4: Porovnání přesnosti podle počtu slov hledání z Td seznamu

Jen pro porovnání je v tabulce 5.9 vidět využití jen Tu části konfiguračního souboru, který pro rozpoznávání není dostačující, dosahuje ze všech nejhorších výsledků - přesnost 73,26%.

		Klasifikátor	
		Td	Tu
Real.	Td	52	43
	Tu	7	85

Tabulka 5.9: Matice záměn - EN - 2 slova Tu

Pokud se tedy hledají jen tázací zájmena a pouze mezi prvními dvěma slovy vět (což se ukázalo jako nejlepší varianta), konečné výsledky jsou v tabulce 5.10. Pro ruská data nebyl ručně vytvořen seznam tázacích zájmen, není tudíž součástí těchto výsledků.

	Accuracy
CZ	94.17 %
DE	98.02 %
EN	96.78 %
ES	94.71 %

Tabulka 5.10: Přesnost ručních pravidel na testovacích datech

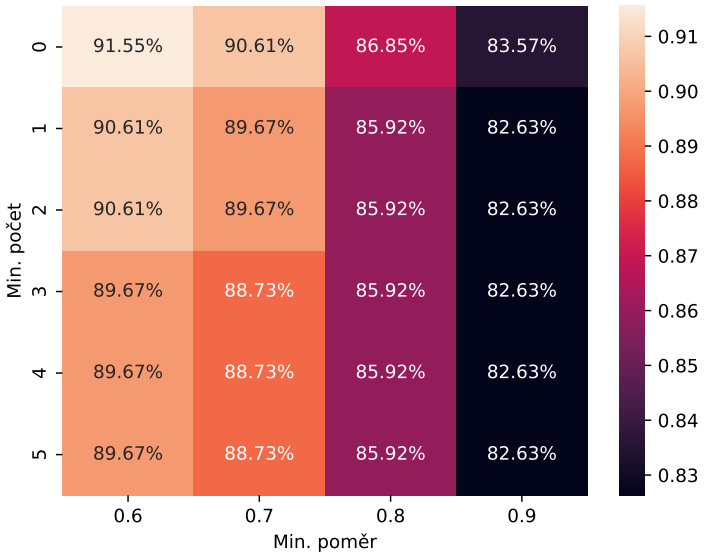
5.2.2 Automatické odvození znalostí - frekvence slov

Obecnějším způsobem, jak vytvořit seznamy zájmen, než je ručně vypisovat v každém tvaru, je vyvodit je přímo z trénovacích dat. Navíc není vyžadována žádná znalost daného jazyka. Pro automatické odvozování se budou na základě pokusů s ručně sepsanými pravidly uvažovat už jen seznamy slov k určení klesavého tónu Td.

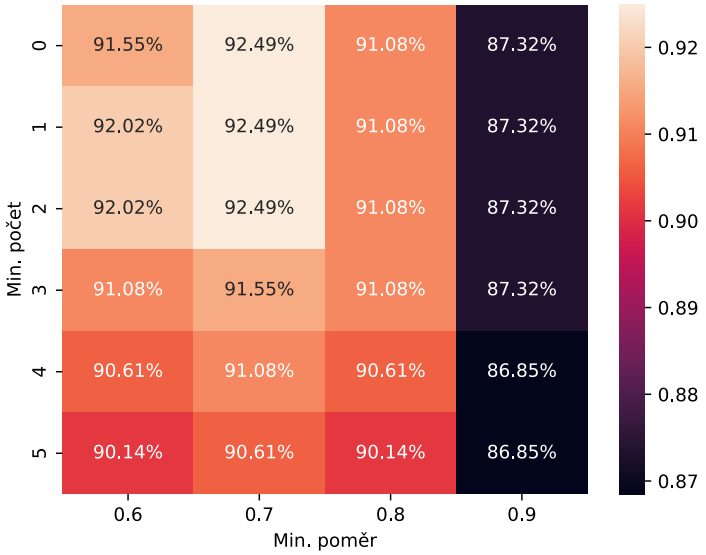
Pro odvození slov do konfiguračního souboru (obsahujícího seznam slov) je předpokladem, aby se dané slovo vyskytovalo častěji v Td otázkách než Tu. To se určí na základě zjištění četností pro všechna slova zvlášť v obou částech dat. U všech se tedy zkoumají dva parametry, pro které je vhodné odvodit prahy:

- Minimální absolutní počet výskytů slova - aby se odfiltrovala nevýznamová slova obsažená v dané třídě spíše náhodou
- Minimální poměr výskytů v dané třídě ze všech výskytů - aby se odfiltrovala slova, která nejsou dostatečně diskriminativní pro danou třídu

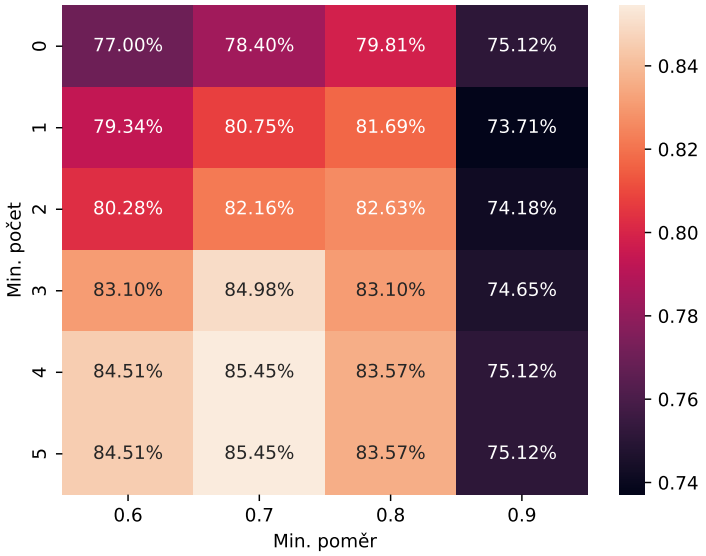
Pro výběr prahů byly pro každou kombinaci různých parametrů vytvořeny konfigurační soubory, ihned otestovány na validačních datech a určena jejich přesnost Accuracy mírou. Následující tabulky vyjadřují accuracy v závislosti na obou parametrech. Pro porovnání byl přístup otestován na prvních slovech vět na obrázku 5.5, na prvních dvou slovech na obrázku 5.6 a na kompletních větách na obrázku 5.7. Konfigurační soubor byl vždy otestován na tolika slovech věty, z kolika byl vytvořen.



Obrázek 5.5: Porovnání přesnosti pro různé parametry - CZ - první slovo



Obrázek 5.6: Porovnání přesnosti pro různé parametry - CZ - první dvě slova

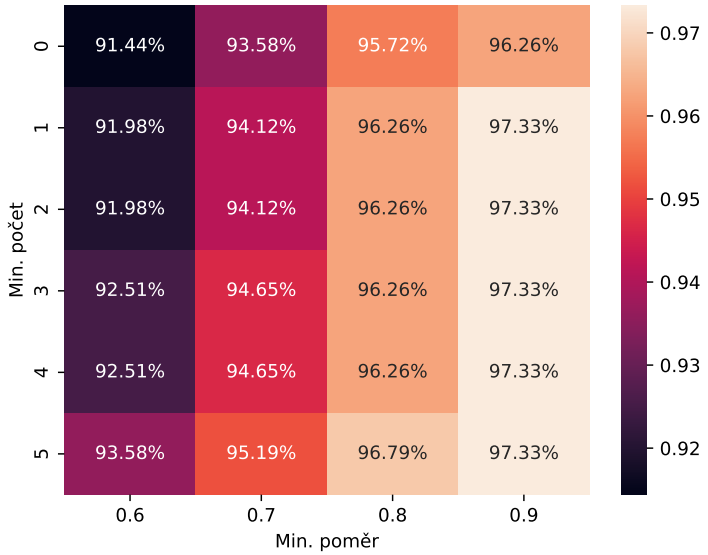


Obrázek 5.7: Porovnání přesnosti pro různé parametry - CZ - všechna slova

I z pokusů s ručně vytvořenými seznamy se vyvodilo, že optimální pro rozpoznávání jsou první dvě slova, z Obrázků 5.5, 5.6 a 5.7 lze vyvodit, že je nejlepší i pro automatické vytváření seznamu (na obrázku 5.6 je dosaženo nejvyšších přesností).

Pro vybrání konstantních parametrů. které by se daly využívat pro optimální využití automatizovaných seznamů, je nutné se podívat na rozložení přesnosti v rámci obou parametrů. V tabulkách jsou zvýrazněná jejich maxima nejsvětlejším políčkem. Pro češtinu by se dalo z tabulek vyvodit, že je výhodnější použít menší poměr v datech (viz obrázek

5.6, kde nejlepších výsledků je dosaženo pro poměr 0.7), ale test s angličtinou na obrázku 5.8 říká naprostý opak, nejlepších výsledků je dosaženo pro poměr 0.9.



Obrázek 5.8: Porovnání přesnosti pro různé parametry - EN - první dvě slova

Nutnost nižšího prahu pro lepší výsledky na českých datech vychází z vyššího poměru otázek typu:

- *Kdo? Tu*
- *Co? Tu*

v celých českých datech, tudíž se daná zájmena nezařadí na seznam. Tato tolerance, aby dané slovo mohlo být často i v druhém typu vět, je důležitá, ale zároveň nesmí být příliš velká. Při nízkém prahu totiž seznamy obsahují i slova, která nemají žádný význam pro konkretní tón, nebo dokonce slova, která v sekci 5.2.1 byla součástí té části konfiguračního souboru určující stoupavý tón, viz Ukázka 5.7.

```
Td:
- should
- does
- was
- the
- so
```

Ukázka 5.7: Část automatického konfiguračního souboru pro angličtinu pro nízký práh poměru (60% výskytů)

Příliš nízký nesmí být ani práh určující minimální četnost, při nulovém prahu se součástí seznamu stanou naprosto všechna slova, která se objeví v celých datech i jen jednou, pokud je to právě v Td otázce, viz Ukázka 5.8

```
Td:
- architect
- board
- bus
- computer
- factors
- water
```

Ukázka 5.8: Část automatického konfiguračního souboru pro angličtinu pro nízký práh četnosti (bez omezení)

Oba prahy by tudíž měly být kompromisem, pro výběr ideálních parametrů se využijí maxima pro všechny jazyky a parametry, pro které na validačních datech nastávají. Jsou znázorněny v Tabulce 5.11.

	Accuracy	Min. počet	Min. poměr
CZ	92.49 %	0-2	0.7
DE	98.51 %	1-2	0.8-0.9
EN	97.33 %	1-5	0.9
ES	98.20 %	4-5	0.7-0.8
RU	91.46 %	2	0.8

Tabulka 5.11: Maximální přesnosti pro pro hledání slov z automaticky vytvořeného konfiguračního souboru v prvních 2 slovech a jejich parametry, vyhodnoceno na validačních datech

Optimální budou některé z parametrů kolem průměrných, zvoleny byly minimálně 2 slova, které zabírají v Td otázkách alespoň 80 % ze všech výskytů. V tabulce 5.12 je vidět konečná přesnost na testovacích datech. Přesnosti klesly jednak testováním na testovacích datech místo validačních, ale i konkrétní volbou parametrů. Při dalším snížení dochází ve všech jazycích ale k problémům znázorněných v Ukázkách 5.7 a 5.8. Pro obecná data tudíž bude vyšší práh účinnější a parametry jsou tudíž optimální pro tvorbu souborů z dat pro nově příchozí jazyky.

	Accuracy	Min. počet	Min. poměr
CZ	92.11 %	2	0.8
DE	95.05 %	2	0.8
EN	95.71 %	2	0.8
ES	93.75 %	2	0.8
RU	90.51 %	2	0.8

Tabulka 5.12: Přesnost pro vybrané parametry automaticky vytvořeného konfiguračního souboru z prvních 2 slov, vyhodnoceno na testovacích datech

5.3 Klasifikátory - Sci-kit Learn

Pro rozpoznávání typů otázek lze využít i tradiční klasifikátory. Za jeden z nejpoužívanějších v oblasti klasifikace textu se považuje klasifikátor využívající SVM (Support Vector Machines), byl proto prioritizován. Pro jeho konkrétní implementaci byla zvolena Python knihovna Scikit-learn, která obsahuje jak zvolený klasifikátor SVC (Support Vector Classification), tak i mnoho dalších metod strojového učení.[12]

Principem SVM je najít nejlepší hyperplochu, která rozdělí dvě třídy dat o libovolné dimenzi. Optimální hyperplochu hledá pomocí support vektorů, které jsou určeny nejbližšími body dané třídy od dané hyperplochy. Klasifikátor se pak snaží nejmenší vzdálenosti hyperplochy od obou tříd maximalizovat. [18]

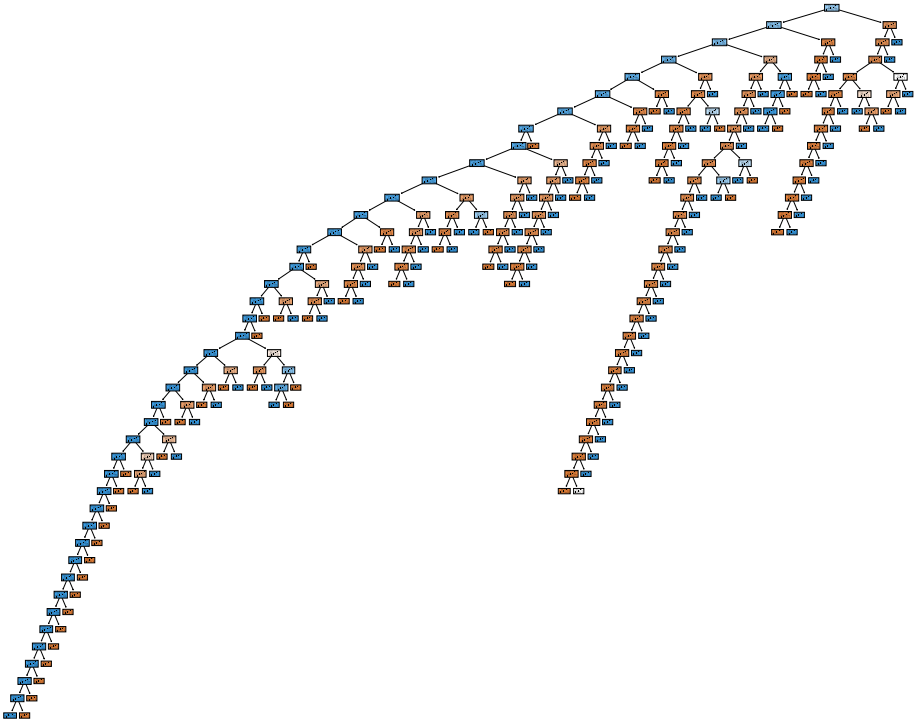
Klasifikátory v knihovně jsou uzpůsobeny klasifikaci dat ve formátu číselných vektorů, ne textu, tudíž je třeba textový formát na číselný převést. K tomu však existují také implementované metody. Bylo nutné vytvořit objekt *CountVectorizer()*, který z textu vytvoří vektory tím způsobem, že si uchovává uspořádaný seznam všech slov v textu. Následně z vět vytvoří vektory o délce celého seznamu slov, který má nenulové prvky pro slova, která se ve větě vyskytují, a odpovídá danému počtu. Pomocí instance *TfidfTransformer()* se navíc tento počet váží na základě četnosti daného slova v celých datech.

Takto vygenerované vektory, vytvořené vždy z celé věty, lze již využít pro klasifikaci, klasifikátor byl vždy natrénován na trénovacích datech a na testovacích otestována přesnost. Tabulka 5.13 porovnává SVM s klasifikátorem podle nejbližšího souseda a pěti nejbližších sousedů (kNN), které jsou znatelně méně přesné. Dále je pro porovnání využito rozhodovacího stromu (DecisionTree), který funguje na podobném principu jako znalostní systém v sekci 5.2.

Jelikož DecisionTree z dat vyvozuje znalosti zcela automaticky a nemá žádné povědomí o problematice, zkoumá všechna slova zvlášť, je struktura rozhodovacího stromu mnohem komplikovanější v porovnání se strukturou rozhodovacích stromů na obrázcích 5.2 a 5.3, viz obrázek 5.9. Tato struktura však umožňuje zamítnutí klesavého tónu například jen po nenalezení dvou různých tázacích zájmen. Nebo lze naopak vidět, že pro přijetí klesavého tónu je potřeba kombinace více slov, i když tázací zájmeno bývá ve větě pouze jedno.

	Klasifikátor			
	SVC	1nn	5nn	DecisionTree
CZ	88.91 %	70.49 %	66.92 %	89.85 %
DE	95.63 %	72.48 %	77.23 %	95.63 %
EN	97.63 %	73.61 %	75.97 %	96.56 %
ES	95.90 %	72.84 %	76.20 %	95.18 %
RU	89.62 %	64.24 %	71.30 %	85.98 %

Tabulka 5.13: Porovnání přesnosti klasifikátorů



Obrázek 5.9: Struktura automaticky vytvořeného rozhodovacího stromu

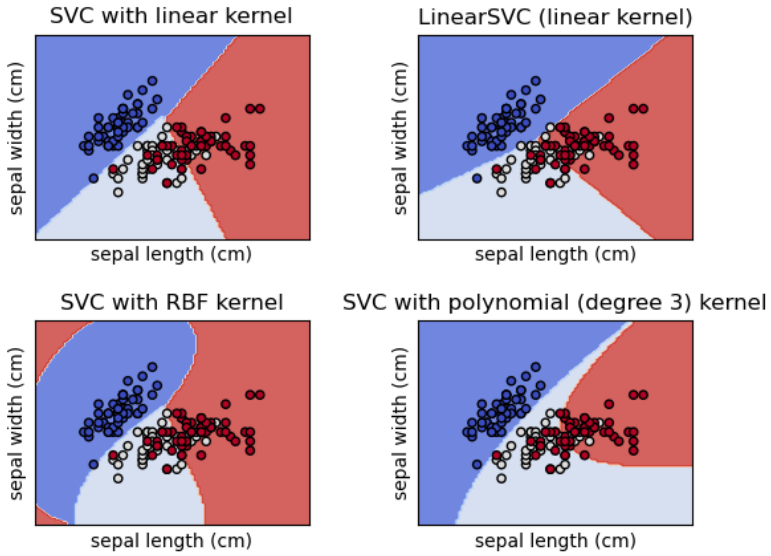
5.3.1 SVC

SVM dosahoval z testovaných klasifikátorů opravdu nejlepších výsledků, v Tabulce 5.13 byl porovnáván klasifikátor využívající SVM bez žádných dodatečných parametrů, takže byly využity ty výchozí - funkce radiální báze (rbf) jako kernel a regularizační koeficient roven 1. [12] V této kapitole je zjišťováno, zda změna parametrů neovlivní zcela zásadně přesnost SVM klasifikátoru.

Kernel je způsob transformace vstupních dat do požadované formy zpracování. Z několikadimenzionálního prostoru všech slov tudíž transformuje prostor do dvou dimenzí, aby mohly být lineárně separovatelné. Kromě funkce radiální báze se využívají sigmoid, polynomy různých stupňů, i pouze lineární. [6] Jejich funkce je vidět na obrázku 5.10.

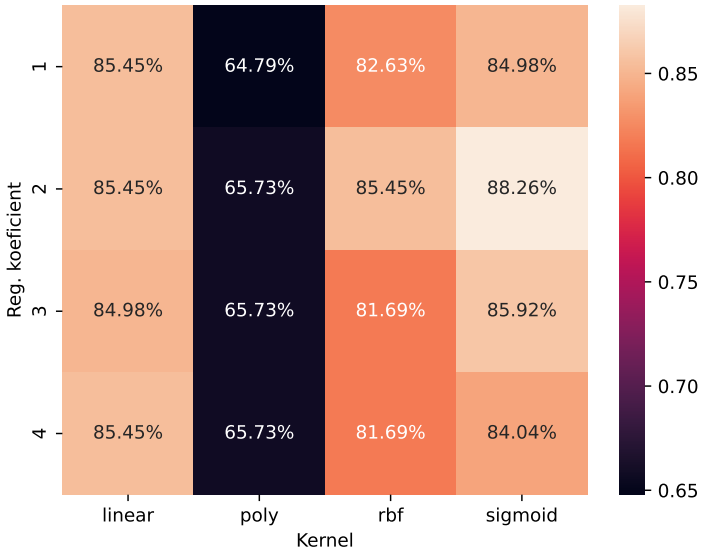
Regularizační koeficient určuje míru penalizace nesprávného zaklasifikování. Řeší kompromis mezi přetrénováním a nedotrénováním, více zašumělá data by měla mít větší

toleranci nesprávného zaklasifikování při trénování a tudíž menší parametr C. Výchozí hodnota je 1 a měla by být zásadně kladná. [21]



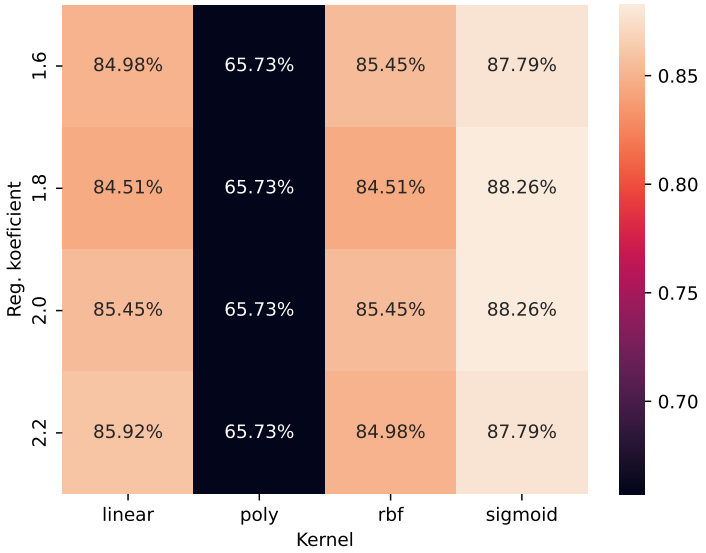
Obrázek 5.10: Rozdělení dat různými kernely [21]

Porovnání přesnosti klasifikace při použití různých kernelů a regularizačních koeficientů je znázorněno na obrázku 5.11. Pro vyhodnocování byla použita validační data.



Obrázek 5.11: Porovnání SVC pro různé parametry - CZ

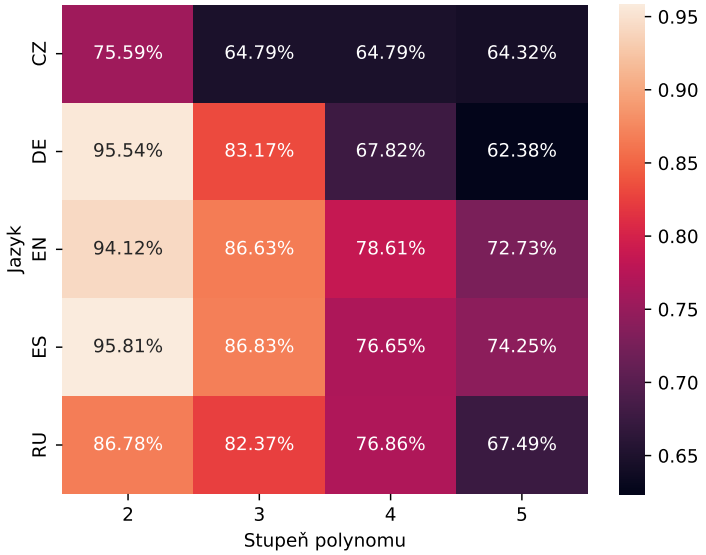
Maximum zřejmě nastává kolem regulačního koeficientu 2, porovnání s detailnějšími rozestupy je pak na obrázku 5.12.



Obrázek 5.12: Porovnání SVC pro různé parametry - CZ - detail

O něco lepších výsledků než výchozí rbf dosahuje lineární kernel, nejlepší výsledky má pro češtinu použití sigmoid kernelu. Výrazně nižší přesnost má kernel polynomu. Výchozí je polynom třetího stupně, stojí za to otestovat, jaký vliv na přesnost mají různé stupně polynomu.

Na obrázku 5.13 je však vidět, že zvýšení stupně polynomu z původních tří rozhodně nezvýší celkovou přesnost. Nejlepších výsledků dosahují polynomy druhého stupně - dostávají se přibližně na úroveň ostatních kernelů.



Obrázek 5.13: Porovnání SVC pro různé stupně polynomu

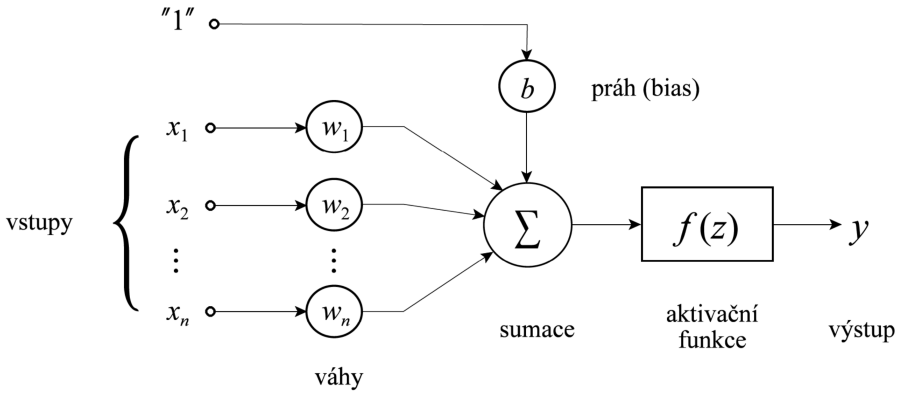
V tabulce 5.14 lze nalézt maxima pro všechny jazyky a parametry, pro něž bylo těchto maxim dosaženo. Na první pohled by se dal označit za nejlepší kernel sigmoid, ale obecně se sigmoid nedoporučuje používat v reálných aplikacích pro SVM, protože funkce nabývá pouze nezáporných hodnot (mezi 0 a 1), což může způsobit nežádoucí chování klasifikátoru. [18] Proto byly raději zvoleny výchozí parametry, jejichž přesnost na testovacích datech byla uvedena již v tabulce 5.13.

	Accuracy	Kernel	Reg. koeficient
CZ	88.26 %	Sigmoid	2
DE	98.51 %	Sigmoid	2-5
EN	98.40%	Sigmoid	1
ES	100 %	Linear	2-3
RU	90.36 %	Sigmoid	1

Tabulka 5.14: Maximální přesnosti pro SVC a jejich parametry

5.4 Neuronová síť T5

Neuronová síť je dnes již velice rozšířený způsob rozpoznávání, je založen na napodobení funkce neuronů v mozku. Původní nejjednodušší model se jmenuje perceptron. Skládá se z jednoho neuronu, vážených vstupů, prahu a výstupu. Jeho strukturu ukazuje obrázek 5.14. Z nich se vyvinuly vícevrstvé sítě (kde výstup jednoho neuronu je vstupem dalšího), které se skládají z více vrstev i více neuronů v jedné vrstvě. Informace se tak šíří jedním směrem ze vstupní vrstvy až po výstupní. Některé typy neuronových sítí obsahují i zpětnou vazbu, kde výstup poslední vrstvy je přiveden zpět na vstup. Architektura sítě je pak s její aktivační funkcí hlavní charakteristikou neuronové sítě. [7] [15]



Obrázek 5.14: Schéma perceptronu [15]

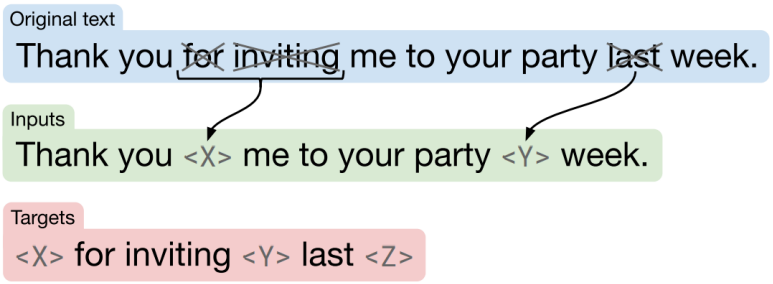
Neuronová síť T5 (Text-To-Text Transfer Transformer) [16] je založena na transfer learningu. Při její tvorbě bylo experimentováno s několika architekturami, jako nejvhodnější

se ukázala enkodér-dekodér architektura. Blok enkodéru zakóduje vstup na vektor příznaků, dekodér příznaky dekóduje na výstup. Text-to-text framework umožňuje použití stejného modelu (se stejnou ztrátovou funkcí a hyperparametry) pro zcela různé úlohy zpracování přirozeného jazyka. Proto ale vyžaduje vstupní data v totožném formátu, jaký je potřeba výstup. Jedině tak je model schopný pochopit, co je po něm vyžadováno. Původní anglický T5 model byl nejdříve bez učitele předtrénován na datech Common Crawl Corpus[3], ze kterých byly vybrány pouze kompletní anglické věty. Výsledný korpus obsahuje zhruba 750 GB dat. [1] Modely pro ostatní jazyky byly natrénovány v průběhu roku 2022 na KKY ZČU v Plzni na základě částí Common Crawl korpusu pro konkrétní jazyk (věty, na kterých byla T5 pro dané jazyky předtrénována, byly staženy v říjnu 2021). Objem dat k předtrénování modelů pro všechny jazyky je v tabulce 5.15.

Jazyk modelu	Velikost trénovacích dat [GB]
CZ	122
DE	649
EN	750
ES	683
RU	2568

Tabulka 5.15: Velikost datasetů pro předtrénování modelů

Proces předtrénování T5 pro daný jazyk probíhá na obyčejných větách tak, že se vynechávají určitá slova z textu a síť se snaží je doplnit, viz Obrázek 5.15. Pro trénování bylo porušeno 15 % z textu. Vynechání 10-25 % slov mělo obdobné výsledky, horších výsledků trénování dosahoval model až při vynechání 50 % slov, tento vyšší poměr by také zpomaloval učení. Takto se ladí všech zhruba 220 milionů základního T5 modelu. [16]



Obrázek 5.15: Předtrénování T5 modelu [16]

S neuronovou sítí T5 bylo pracováno za pomoci Python knihovny t5s. [22] Použit

[3]https://commoncrawl.org/

byl skript *fine-tuning.py*, který slouží k dotrénování neuronové sítě pro konkrétní úlohu.[4] Dotrénování probíhá podle údajů v konfiguračním souboru, jehož název je nutné předat jako argument při spuštění.

V konfiguračním souboru by měly být názvy souborů s daty, T5 modelu a tokenizeru. Dále obsahuje specifikace k načtení a trénování dat - např. počet epoch nebo konstantu učení. Data by měla být rozdělena na trénovací, testovací a validační a ve formátu tsv, tudíž přesně, jak byla upravena v kapitole 5.1.

Modely byly trénovány na výpočetních zdrojích organizace MetaCentrum. Trénování probíhalo v 10 epochách po 1000 krocích, konstanta učení bylo rovna 0.01. Po každé epoše byl model validován otestováním na validačních datech. Nakonec byl finální model otestován jak na validačních, tak testovacích datech. Výsledné přesnosti ukazuje tabulka 5.16.

	valid	test
CZ	94.84 %	93.42 %
DE	98.51 %	96.04 %
EN	96.26 %	96.57 %
ES	98.80 %	95.19 %
RU	89.81 %	91.61 %

Tabulka 5.16: Přesnost T5 na validačních a testovacích datech

Během trénování se model snaží zvyšovat přesnost na validačních datech, proto je často vyšší než přesnost na testovacích. U španělských a německých dat je tento rozdíl největší, ale po nahlédnutí do výsledků lze říci, že větší rozdíl je pravděpodobně způsoben jen konkrétním náhodným rozdělením validačních a testovacích dat, a nedošlo k přetrénování. Natrénované T5 modely jsou součástí přílohy[5].

[4]Přetrénované T5 modely byly autoru této BP dodány. Součástí práce bylo jen dotrénování modelu na zadanou úlohu klasifikace typů otázek, proto původní modely nejsou součástí přílohy.

[5]Z důvodu velikosti nejsou součástí přílohy na GitHubu, ale jsou uloženy zvlášť na Google Disku: `https://drive.google.com/drive/folders/19MWfAiByBFOWcqqCpOnDlaOXfoDMjDxm?usp=sharing`

Kapitola 6

Shrnutí výsledků a analýza chyb

Práce porovnává metody rozpoznávání typů otázek, které dělila podle stoupavé a klesavé intonace.

Jako první se pro klasifikaci vycházelo z informace, že doplňovací otázky s klesavým tónem v sobě mají obsaženy tázací zájmena. Na tomto základě bylo vytvořeno pravidlo, které vyhledává podle předem daného seznamu zájmen, zda se ve větě nějaké nachází. Pokud ano, určí tón jako klesavý.

Tento seznam byl vytvářen dvěma způsoby. Jeho sepsání ručně zaručuje, že všechna slova jsou opravdu tázacími zájmeny a jsou vypsána ve všech tvarech, ve kterých se mohou objevit. To však nelze provést pro ty jazyky, pro které není k dispozici jazykový expert, který by takový seznam mohl vytvořit.

Proto byl zpracován druhý přístup, který nevyžaduje žádnou znalost jazyka a vychází pouze z dostupných dat. Podle četnosti slov v typech otázek vybere taková, která jsou vždy dostatečně diskriminativní pro danou třídu. Z principu by to měla být právě tázací zájmena, ale tato metoda nezaručuje, že všechna vybraná slova jsou skutečně tázací zájmena. Také nemá jak zajistit, že budou zapsána ve všech jejich možných tvarech, využije jen slova, která jsou reálně obsažena v trénovacích datech.

Pro ruční i automatická pravidla bylo po testování rozhodnuto, že k hledání tázacích zájmen docházelo pouze mezi prvními dvěma slovy věty. Automatická pravidla byla odvozována také pouze z prvních dvou slov.

Následně byly zvažovány klasické klasifikátory, které jsou uzpůsobeny číselným vektorům. Tf-idf transformace sice zvýhodnila často se objevující slova ve větách a rozpoznávání tedy probíhalo hlavně podle nich, ale jinak je zcela odtrženo od jakéhokoliv reálného kontextu a významu. Po vyzkoušení různých parametrů nedosahovala žádná kombinace výrazně vyšších výsledků, nelze proto žádné obecně doporučit, a tak byly využity výchozí parametry.

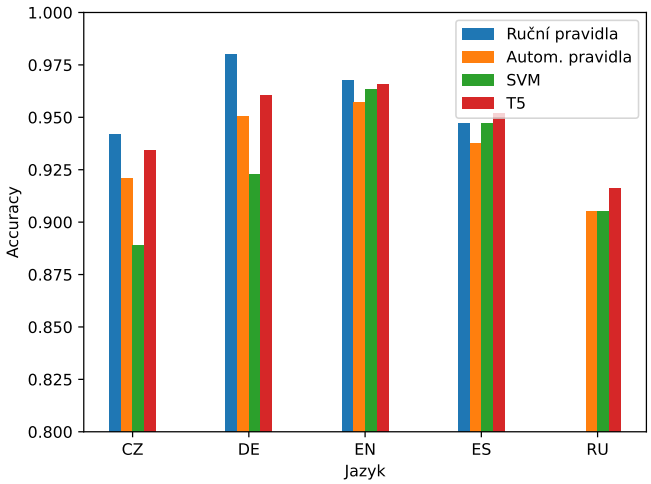
T5 je velký předtrénovaný model neuronové sítě používané ke zpracování přirozeného jazyka v různých oblastech. Probíhalo její dotrénování na konkrétní úlohu v konkrétním jazyce. Předtrénování zajistí, že model „rozumí“ danému jazyku. Dotrénování proběhlo pomocí t5s knihovny za pomoci pouhého konfiguračního souboru, což výrazně usnadnilo

implementaci.

Porovnání přesnosti všech použitých metod otestovaných na testovacích datech ukazuje tabulka 6.1[1]. Zvýrazněna je vždy nejvyšší přesnost pro daný jazyk, na obrázku 6.1 lze navíc vidět přesnost graficky.

	pravidla	aut. pravidla	SVM	T5
CZ	**94.17 %**	92.11 %	88.91 %	93.42 %
DE	**98.02 %**	95.05 %	92.28 %	96.04 %
EN	**96.78 %**	95.71 %	96.35 %	96.57 %
ES	94.71 %	93.75 %	94.71 %	**95.19 %**
RU	—	90.51 %	90.51 %	**91.61 %**

Tabulka 6.1: Nejlepší výsledky pro všechny jazyky každým způsobem klasifikace



Obrázek 6.1: Porovnání výsledků různých klasifikátorů

Všechny metody dosahují vysoké přesnosti, kromě využití SVM pro češtinu vždy nad 90 %. Z tabulky lze usoudit, že nejlepších výsledků je dosaženo využitím ručně vytvořených pravidel a T5 neuronové sítě. Obě metody mají tu vlastnost, že zakládají na reálné znalosti problému - pravidla pomocí člověka a neuronová síť díky předtrénování na obrovském množství dat. Ruční pravidla mají tu výhodu, že nevyžadují žádné trénování. Jejich velkou nevýhodou je, že vyžadují existenci jazykového experta, který by seznam tázacích zájmen vytvořil. Výhoda T5 je, že je celý proces plně automatizovaný, nejsou potřeba žádná ručně vytvářená pravidla, T5 se naučí vše pouze z dat. Nevýhoda T5 je velká výpočetní složitost i velikost modelu, oproti pravidlům navíc musí spoléhat pouze na to, co je obsaženo v trénovacích datech.

[1]Pole s pravidly pro ruský jazyk je prázdné, protože nebyl k dispozici jazykový expert, který by ručně vytvořil kompletní seznam tázacích zájmen.

Automatická pravidla jsou principem velmi podobná těm ručním a mohou být jejich náhradou, pokud neexistuje pro daný jazyk expert. Dosahují však horších výsledků, protože tento přístup nezaručuje výběr správných slov, podle kterých má klasifikovat, silně záleží na obsahu trénovacích dat, ze kterých se pravidla vyvozují. SVM pak může být nejjednodušší pro použití, ale svých nejlepších výsledků může dosáhnout pouze konkrétním výběrem parametrů specifickým pro konkrétní jazyk, což použití ztěžuje. I tak ale nedosahuje výsledků ostatních metod.

V rámci jazyků dosahuje nejhorších výsledků ruština, přestože pro ni bylo k dispozici nejvíce dat (viz tabulka 5.2). Vzhledem k tomu, že data neprošla žádnou kontrolou, nelze říci, v kolika případech je klasifikace určena jako chybná, přestože chyba je v datech, a ne klasifikaci. V datech pro ostatní jazyky bylo opraveno zhruba 1-2 % chyb (viz tabulka 5.1), což zlepšilo přesnost při klasifikaci. Pro tyto opravované jazyky se k sobě přesnosti blíží, počet dat je pro ně srovnatelný, ale závisí nejen na jejich celkovém počtu, ale i na počtu vět s jednotlivými tázacími zájmeny, nebo na poměrném zastoupení eliptických a vylučovacích otázek, které jsou těmi nejproblémovějšími z hlediska klasifikace.

6.1 Analýza chyb

Přestože dosahují nejlepších výsledků pro všechny jazyky, ani pravidla, ani T5 nejsou bezchybné metody. Chybují hlavně v oblastech, které jsou zmíněny již v kapitole 5.1.1 o korekci dat. V následujících příkladech jsou věty uvedeny s tím tónem, který jim přiřadil daný klasifikátor:

- Hlavním problémem u pravidel byly eliptické otázky, které obsahovaly tázací zájmena.
  - *How come? Td* (správně Tu)
  - *Cože? Td* (správně Tu)

  T5 si s nimi v některých případech ale poradila:
  - *Jaký? Tu*
  - *Od koho? Tu*

  K eliminaci těchto chyb by k pravidlům bylo možné doplnit podmínku, že klesavý tón by se přiřazoval pouze větám, které mají více než dvě slova.
- T5 měl naopak větší problémy, když se ve větách objevovaly některé méně obvyklé tvary tázacích zájmen.
  - *Oč konkrétně jde? Tu* (správně Td)

  V trénovacích datech bylo „Oč“ obsaženo celkem 3x, všechny výskyty se správnou anotací, což znamená, že ho lépe zaklasifikovala nejen ruční pravidla, ale i ta automatická.

  - *Oč jde v současné době? Td*
  - *Oč jde autorům? Td*
  - *Oč v souboji advokátů šlo? Td*

- Pravidla měla problém, pokud se vyskytovala na druhé pozici ve větě spojka o stejném přepisu jako tázací zájmena. Spojku na jakémkoliv pozdějším místě však zaklasifikovala správně

  - *Víte, co to znamená? Td* (správně Tu)
  - *Víte, co to je? Td* (správně Tu)
  - *Věříte všemu, co vám poradenské firmy doporučují? Tu*

  Pravidla tak nejsou schopna rozhodnout, že druhá věta v delším souvětí začíná tázacím zájmenem. Správně tuto větu nezaklasifikovala ani T5, jelikož v trénovacích datech není dostatek příkladů, kdy se tázací zájmeno objevuje uprostřed věty.

  - *Sám jste zařídil dva góly, jak jste viděl obě situace? Tu* (správně Td)

  Pro zlepšení klasifikace těchto vět by bylo vhodné zkombinovat toto rozpoznávání s nějakou metodou morfologicko-syntaktické analýzy, která by předem určila slovní druhy a odlišila by tak zájmena od spojek.

- Uprostřed věty se objevuje i slovo „nebo“, jak ve slučovacím, tak vylučovacím poměru. Pravidla nemají, jak tento typ otázek ovlivnit, když hledají pouze v prvních dvou slovech - všechna „nebo“ jsou klasifikována jako Tu. Hledání by se dalo modifikovat do podoby, kdy se hledá v prvních dvou slovech každé věty souvětí, v češtině je lze někdy odlišit podle čárek, ale například v angličtině se čárky tolik nepoužívají. T5 se snaží rozlišovat mezi „nebo“ v obou poměrech, ale vzhledem k malému výskytu v trénovacích datech také není příliš úspěšná:

  - *To be or not to be? Tu* (správně Td)
  - *Are you blind or something? Td* (správně Tu)
  - *Will you pay by cash or cheque? Tu* (správně Td)

- T5 někdy chybovala u naprosto běžných doplňovacích a zjišťovacích otázek, které však byly velmi dlouhé:

  - *Která z vašich velkých bank nemusí navyšovat rezervy, zvlášť při aplikaci nejpřísnějších zahraničních standardů? Tu* (správně Td)
  - *Are you organizing an upcoming car show, vintage rally, car club meeting, classic car display or other automotive event? Td* (správně Tu)

- Neuronová síť měla navíc problémy u vět začínajících slovem „A“, pravděpodobně kvůli tomu, že jím v trénovacích datech často začínají eliptické otázky.
  - *A co říkal na výkon rozhodčích? Tu* (správně Td)
  - *A jak platí ředitele škol, které osm let degradují maturitu? Tu* (správně Td)

Z výše uvedené analýzy vyplývá, že nejčastějšími chybami byly eliptické a vylučovací otázky, klasifikátory také neumí rozeznat tázací zájmena uprostřed vět od spojek. Pro lepší výsledky T5 (a i SVM a jiných klasifikátorů) by muselo být dodáno větší množství dat, které by pokrývalo tyto spornější oblasti rozpoznávání. Mezi pravidla by se dalo zařadit nové, které by bralo v potaz celkovou délku věty a všechny velmi krátké by byly označeny jako eliptické. Pomocí jednoduchých pravidel by však nešlo rozlišit spojky uprostřed vět od tázacích zájmen a rozlišit, v jakém poměru je použito slovo nebo. Pomoci by mohlo automatické zjišťování slovních druhů.

Většina těchto problémů by nenastala, pokud by se z dat na začátku odfiltrovaly eliptické a vylučovací otázky. Rozpoznávání by se pak zaměřovalo jen na zjišťovací a doplňovací otázky, kterých je v datech výrazná většina, a všechny přístupy by pak dosahovaly větší úspěšnosti klasifikace.

I přes uvedené chyby je T5 model nejlepším automatickým přístupem k rozpoznávání typů otázek, a to pro všechny testované jazyky. Úspěšnost toho přístupu je v průměru 95 % (viz tabulka 6.1). Pokud by se tudíž T5 model použil v TTS systému, pouze 1 otázka z 20 by měla špatnou intonaci.

Kapitola 7

Závěr

Tato práce se zabývala problematikou předzpracování textu pro TTS systémy, které produkují syntetickou řeč z běžného textu, konkrétně rozpoznáváním koncových tónů otázek, které jsou potřeba při generování prozodie. Úvodní kapitoly se proto zabývají obecně syntézou řeči a intonací. Hlavním cílem práce byla ruční kontrola zpracovávaných dat a jejich využití k navržení postupů pro detekci typů otázek.

Kapitola 2 nejprve rozebírá jednotlivé přístupy k syntéze řeči podle typu modelu, poté přibližuje postupný historický vývoj syntetizéru řeči. Nakonec popisuje jednotlivé fáze zpracování přirozeného jazyka pro následnou syntézu, kde se zabývá morfologicko-syntaktickou analýzou, fonetickou transkripcí a generováním prozodie.

Na generování prozodie navazuje další kapitola, věnovaná nejdůležitější prozodické charakteristice: intonaci. Zkoumá hlavně větnou intonaci, jak se projevují typy vět z hlediska tónu, což je nejdůležitější pro jejich rozpoznávání. Tuto informaci ověřuje pro všechny jazyky, u kterých se měl typ otázek z hlediska tónu automaticky rozpoznávat, a to češtinu, angličtinu, němčinu, španělštinu a ruštinu. Navíc zvýrazňuje některá specifika jazyků a ukazuje, jak se může intonace u člověka odklonit od pravidel na základě různých emocionálních rozpoložení.

Pátá kapitola se zabývá samotným rozpoznáváním typů otázek. Začíná krátkým úvodem o úpravě dat do formátu, ve kterém s nimi bylo pracováno. Data byla rozdělena na trénovací, ze kterých se natrénoval klasifikátor, validační, na kterých se optimalizoval vždy daný klasifikátor, a testovací, na nichž se různé klasifikátory porovnaly mezi sebou. Postupně v sekcích 5.2-5.4 byl představen průběh nastavení znalostního systému s ručně a automaticky odvozenými pravidly, klasifikátorů z knihovny Sci-kit Learn a neuronové sítě T5 a jejich využití při rozpoznávání tónů otázek.

Výsledky byly následně porovnány. Za nejlepší metodu byly označeny ručně vytvořená pravidla a T5 neuronová síť díky jejich vhledu do problematiky - pravidla z pohledu člověka a T5 z obrovského množství dat, na kterých byla předtrénována.

Hlavním přínosem této bakalářské práce bylo dotrénování T5 modelů, které jsou schopny zcela automaticky s vysokou přesností na základě pouhého textu otázek rozpoznat jejich typ a přiřadit jim tak správnou koncovou intonaci. Tuto informaci pak lze

využít pro syntézu řeči v TTS systémech pro zvýšení srozumitelnosti a přirozenosti syntetizovaných promluv.

Literatura

[1] Adam Frémund. *Porozumění řeči založené na neuronových sítích*. Diplomová práce, KKY ZČU Plzeň, 2021.

[2] Daniel Hirst a Albert Di Cristo. *Intonation Systems: A Survey of Twenty Languages*. Cambridge University Press, 1998. ISBN: 9780521395137.

[3] Lauri Juvela, Bajibabu Bollepalli, Junichi Yamagishi a Paavo Alku. "Waveform Generation for Text-to-speech Synthesis Using Pitch-synchronous Multi-scale Generative Adversarial Networks". In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2018), s. 6915–6919. URL: https://api.semanticscholar.org/CorpusID:53116388.

[4] Martin Kříž a Li Xiaokun. *Učebnice čínského jazyka I.* 1. Praha: Oeconomica, 2011. ISBN: 978-80-245-1768-1.

[5] Frank Kügler. *Do we know the answer? - Variation in yes-no-question intonation*. Led. 2003.

[6] *Major Kernel Functions in Support Vector Machine (SVM)*. Ún. 2022. URL: https://www.geeksforgeeks.org/major-kernel-functions-in-support-vector-machine-svm/.

[7] Vladimír Mařík, Olga Štěpánková a Jiří Lažanský. *Umělá inteligence (1)*. Academia, 1993.

[8] Vladimír Mařík, Olga Štěpánková a Jiří Lažanský. *Umělá inteligence (5)*. Academia, 2007.

[9] O.Karaali, G. Corrigan a I. A. Gerson. "Speech Synthesis with Neural Networks". In: *CoRR* cs.NE/9811031 (1998). URL: https://arxiv.org/abs/cs/9811031.

[10] Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alexander Graves, Nal Kalchbrenner, Andrew Senior a Koray Kavukcuoglu. "WaveNet: A Generative Model for Raw Audio". In: *Arxiv*. 2016. URL: https://arxiv.org/abs/1609.03499.

[11] Zdena Palková. *Fonetika a fonologie češtiny*. Karolinum, 1994.

[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot a E. Duchesnay. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), s. 2825–2830.

[13] Kathryn Pruitt a Floris Roelofsen. "The Interpretation of Prosody in Disjunctive Questions". In: *Linguistic Inquiry* 44 (říj. 2013), s. 632–650. DOI: 10.1162/LING_a_00141.

[14] Josef Psutka, Luděk Müller, Jindřich Matoušek a Vlasta Radová. *Mluvíme s počítačem česky*. Prague: Academia, 2006. ISBN: 80-200-1309-1. URL: http://www.kky.zcu.cz/en/publications/PsutkaJ_2006_Mluvimes.

[15] Josef V. Psutka. "Přednášky z předmětu KKY/ZSUR".

[16] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li a Peter J. Liu. *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*. 2019. DOI: 10.48550/ARXIV.1910.10683. URL: https://arxiv.org/abs/1910.10683.

[17] Markéta Řezáčková, Jan Švec a Daniel Tihelka. "T5G2P: Using Text-to-Text Transfer Transformer for Grapheme-to-Phoneme Conversion". In: *Interspeech*. 2021. URL: https://api.semanticscholar.org/CorpusID:239702895.

[18] Indriani Sitorus. *Introduction to SVM and kernel trick - part 1 (theory)*. Zář. 2020. URL: https://medium.com/analytics-vidhya/introduction-to-svm-and-kernel-trick-part-1-theory-d990e2872ace.

[19] Radek Skarnitzl, Pavel Šturm a Jan Volín. *Zvuková báze řečové komunikace : fonetický a fonologický popis řeči*. Prague: Univerzita Karlova v Praze, Nakladatelství Karolinum, 2016. ISBN: 9788024632728.

[20] Marie Sochrová. *Český jazyk v kostce pro SŠ*. Fragment, 2010.

[21] *Support Vector Machines*. Ún. 2022. URL: https://scikit-learn.org/stable/modules/svm.html.

[22] Jan Švec. *t5s - T5 made simple*. URL: https://github.com/honzas83/t5s.