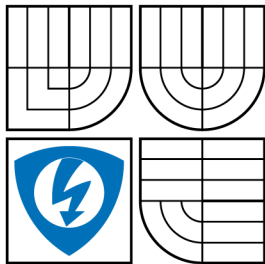


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND
COMMUNICATION
DEPARTMENT OF TELECOMMUNICATION

MODERNÍ METODY MULTIMEDIÁLNÍHO VYUČOVÁNÍ MODERN METHODS OF MULTIMEDIA TEACHING

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

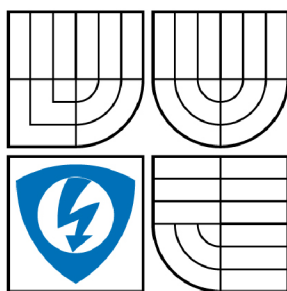
AUTOR PRÁCE
AUTHOR

ZDENĚK MAZAL

VEDOUCÍ PRÁCE
SUPERVISOR

ING. VÁCLAV PFEIFER

BRNO 2008



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav telekomunikací

Bakalářská práce

bakalářský studijní obor

Teleinformatika

Student: Mazal Zdeněk

ID: 77854

Ročník: 3

Akademický rok: 2007/2008

NÁZEV TÉMATU:

Moderní metody multimediálního vyučování

POKYNY PRO VYPRACOVÁNÍ:

Zjistěte současný stav elektronického vyučování v České republice a ve světě. Rozeberte možnosti, výhody a nevýhody elektronické formy vzdělávání a možnosti aplikace tzv. E-learningu v praxi. Dále se zaměřte na problém tzv. multimediálních vyhledavačů, které efektivně vyhledávají na základě určitých znaků v Audio/Video záznamu. Co je to tzv. Pattern ? Co je to indexace obsahu ? Na základě teoretických znalostí navrhnete jednoduchou koncepci multimediálního vyhledavače. Koncepci pak prakticky realizujete ve vhodném programovacím jazyku a provedte vhodná měření, která budou prezentovat úspěšnost vyhledavače. Diskutujte možnosti vylepšení aby se dal systém efektivně využívat v praxi, např. v rámci E-learningového systému.

DOPORUČENÁ LITERATURA:

- [1] Spell, B.: JAVA Programujeme profesionálně. Nakladatelství CPRESS 2002. ISBN: 80-7226-667-5.
- [2] Roček, R.: Moderní metody elektronického vyučování, Bakalářská práce, Ústav Telekomunikací FEKT VUT v Brně.

Termín zadání: 11.2.2008

Termín odevzdání: 4.6.2008

Vedoucí práce: Ing. Václav Pfeifer

prof. Ing. Kamil Vrba, CSc.

předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

LICENČNÍ SMLOUVA

POSKYTOVANÁ K VÝKONU PRÁVA UŽÍT ŠKOLNÍ DÍLO

uzavřená mezi smluvními stranami:

1. Pan/paní

Jméno a příjmení: Zdeněk Mazal
Bytem: Benátská 1018, 57001, Litomyšl - Litomyšl-Město
Narozen/a (datum a místo): 18.8.1985, Litomyšl

(dále jen "autor")

a

2. Vysoké učení technické v Brně

Fakulta elektrotechniky a komunikačních technologií
se sídlem Údolní 244/53, 60200 Brno 2
jejímž jménem jedná na základě písemného pověření děkanem fakulty:
prof. Ing. Kamil Vrba, CSc.

(dále jen "nabyvatel")

Článek 1

Specifikace školního díla

1. Předmětem této smlouvy je vysokoškolská kvalifikační práce (VŠKP):

- disertační práce
- diplomová práce
- bakalářská práce

jiná práce, jejíž druh je specifikován jako

(dále jen VŠKP nebo dílo)

Název VŠKP: Moderní metody multimediálního vyučování

Vedoucí/školitel VŠKP: Ing. Václav Pfeifer

Ústav: Ústav telekomunikací

Datum obhajoby VŠKP:

VŠKP odevzdal autor nabyvateli v:

- tištěné formě - počet exemplářů 1
- elektronické formě - počet exemplářů 1

2. Autor prohlašuje, že vytvořil samostatnou vlastní tvůrčí činností dílo shora popsané a specifikované. Autor dále prohlašuje, že při zpracovávání díla se sám nedostal do rozporu s autorským zákonem a předpisy souvisejícími a že je dílo dílem původním.
3. Dílo je chráněno jako dílo dle autorského zákona v platném znění.
4. Autor potvrzuje, že listinná a elektronická verze díla je identická.

Článek 2
Udělení licenčního oprávnění

1. Autor touto smlouvou poskytuje nabyvateli oprávnění (licenci) k výkonu práva uvedené dílo nevýdělečně užít, archivovat a zpřístupnit ke studijním, výukovým a výzkumným účelům včetně pořizování výpisů, opisů a rozmnoženin.
2. Licence je poskytována celosvětově, pro celou dobu trvání autorských a majetkových práv k dílu.
3. Autor souhlasí se zveřejněním díla v databázi přístupné v mezinárodní síti
 - ihned po uzavření této smlouvy
 - 1 rok po uzavření této smlouvy
 - 3 roky po uzavření této smlouvy
 - 5 let po uzavření této smlouvy
 - 10 let po uzavření této smlouvy(z důvodu utajení v něm obsažených informací)
4. Nevýdělečné zveřejňování díla nabyvatelem v souladu s ustanovením § 47b zákona č. 111/1998 Sb., v platném znění, nevyžaduje licenci a nabyvatel je k němu povinen a oprávněn ze zákona.

Článek 3
Závěrečná ustanovení

1. Smlouva je sepsána ve třech vyhotoveních s platností originálu, přičemž po jednom vyhotovení obdrží autor a nabyvatel, další vyhotovení je vloženo do VŠKP.
2. Vztahy mezi smluvními stranami vzniklé a neupravené touto smlouvou se řídí autorským zákonem, občanským zákoníkem, vysokoškolským zákonem, zákonem o archivnictví, v platném znění a popř. dalšími právními předpisy.
3. Licenční smlouva byla uzavřena na základě svobodné a pravé vůle smluvních stran, s plným porozuměním jejímu textu i důsledkům, nikoliv v tísní a za nápadně nevýhodných podmínek.
4. Licenční smlouva nabývá platnosti a účinnosti dnem jejího podpisu oběma smluvními stranami.

V Brně dne:

.....

Nabyvatel

.....

Autor

ABSTRAKT

Obsahem práce je shrnutí výhod a nevýhod e-learningu, další část se zabývá problematikou vyhledáváním klíčových slov ve zvukových záznamech, kde je přehled používaných metod, fungujících vyhledávačů, jejich rozdělení a možnosti využití. Obsahuje také návrh, realizaci a výsledky úspěšnosti jednoduchého vyhledávače slov ve zvukovém záznamu programovaného v prostředí Matlab.

KLÍČOVÁ SLOVA

E-learning, vyhledávač, Matlab, DTW, HMM

ABSTRACT

The work is a summary of the advantages and disadvantages of e-learning, the next section deals with search keywords in sound record, where the survey methods used, operating search engines, their division and the possibilities of use. It also includes the design, implementation and results of the success of a simple search engine of the words in sound record, programmed in Matlab Environment.

KEYWORDS

E-learning, search engine, Matlab, DTW, HMM

MAZAL Z. *Moderní metody multimediálního vyučování*. Brno: Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií. Telekomunikací, 2008. 38. Bakalářská práce. Vedoucí práce byl Ing. Václav Pfeifer.

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Moderní metody multimediálního vyučování“ jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

V Brně dne

.....

(podpis autora)

OBSAH

Úvod	12
1 E-learning	13
1.1 Výhody a nevýhody e-learningu	13
1.1.1 Výhody e-learningu	13
1.1.2 Nevýhody e-learningu	15
2 Vyhledávání klíčových slov	18
2.1 Dělení rozpoznávání řeči podle různých kritérií	18
2.2 Podle druhu metody, na které je založeno rozpoznávání	19
2.3 Podle druhu konstrukce systému rozpoznávání řeči	20
2.3.1 LVCSR	20
2.3.2 Detekce klíčových slov s výplňovým modelem	21
2.3.3 Detekce klíčových slov založená na míře důvěry	21
2.3.4 Fonetické rozpoznávání řeči	21
2.4 Přehled systémů na rozpoznávání a vyhledávání řeči	22
2.4.1 Systém pro automatický přepis televizních a rozhlasových pořadů	22
2.4.2 HP Speech Bot	22
2.4.3 Blinkx	23
2.4.4 Nexidia Nexaminer Enterprise	23
2.4.5 CallManager Eureka!	23
2.5 Indexace a pattern	23
2.6 Architektura systému na vyhledávání v řečových databázích	24
2.7 Skryté Markovovy modely (HMM)	24
2.7.1 Vintsyukův typ	25
2.7.2 4-stavový skrytý Markovův model	25
2.8 Algoritmus dynamic time warping (DTW)	26
2.9 Funkce počet průchodů nulou	28
3 Návrh a realizace jednoduchého vyhledávače	30
3.1 Návrh vyhledávače klíčových slov	30
3.2 Realizace vyhledávače slov ve zvukovém záznamu	31
3.2.1 Úspěšnost vyhledávače	33
4 Závěr	35
Literatura	36

Seznam symbolů, veličin a zkratk	37
Seznam příloh	38

SEZNAM OBRÁZKŮ

1.1	Grafické znázornění finanční náročnosti e-learningu na čase [6]	16
1.2	Grafické znázornění výhod a nevýhod jednotlivých forem výuky [5]	17
2.1	Blokové schéma systému detekce klíčových slov	20
2.2	Základní struktura systému pro vyhledávání ve zvukových záznamech [4]	24
2.3	Vintsyukův typ Markovova modelu	25
2.4	4-stavový skrytý Markovův model	25
2.5	Příznaky obrazu A a B	27
2.6	Lokální vzdálenosti obrazů A a B	27
2.7	Akumulované vzdálenosti a minimální cesta	28
2.8	Signál slova jedna a jeho analýza počtem průchodů nulou	29
3.1	Blokové schéma navrhnutého vyhledávače	30
3.2	Architektura práce vytvořeného programu	33

SEZNAM TABULEK

3.1 Ukázka úspěšnosti vyhledávače	34
---------------------------------------------	----

ÚVOD

Stále více škol a komerčních subjektů využívá ke své výuce e-learning, a to především kvůli svým nesporným výhodám, které mají být shrnuty v této práci. Na druhé straně však přináší i řadu problémů, kterými se práce také zabývá. Avšak hlavním cílem je studie vyhledávačů slov ve zvukových záznamech, neboť stále vzrůstá počet zvukových dat, ať už v rámci v rámci e-learningu jako záznamy z přednášek, nebo jiných zvukových záznamů. Proto také vzrůstá potřeba tyto materiály třídit a vyhledávat v nich. Práce se zabývá dělením vyhledávačů dle různých kritérií, metodami a způsoby vyhledávání. Výstupem je pak jednoduchý vyhledávač slov ve zvukovém záznamu, vytvořen v programu Matlab, využívající metodu počtu průchodů nulou a algoritmus dynamického borcení časové osy.

1 E-LEARNING

Rozvoj komunikačních a informačních technologií přispěl k jejich uplatnění v mnoha oborech, nejinak tomu je v oblasti vzdělávání. Aplikací těchto technologií do procesu vzdělávání označujeme jako e-learning, nebo-li elektronické vyučování. E-learningové systémy mohou pomoci v celém okruhu výuky od přípravy materiálů, přes samotnou výuku až k administrativním úkonům. V případě výuky, jde často o výuku, která využívá různých možností, takovým způsobem, aby byl zásah učitele do výuky minimální. Pomocí moderních nástrojů, postupů a procesů se e-learning snaží co nejefektivněji působit na smysly. Na zrak pomocí textu, fotografií a videí, na sluch hudebními ukázkami a čteným slovem. Nebo kombinací vzniknou multimediální ukázky, animace a vizualizace.

1.1 Výhody a nevýhody e-learningu

Každé nové metody a postupy, které se objeví, jsou podrobeny důkladné analýze, srovnáváním a zhodnocením, nejinak tomu je i u e-learningového vyučování. E-learning s sebou přináší jak výhody, tak i nevýhody oproti klasickému vyučování. Záleží pak na finálním řešení a způsobu využití, zda do popředí vystoupí jeho přednosti, nebo naopak způsobí více škod než užitku.

1.1.1 Výhody e-learningu

- **Neomezený přístup k informacím**

Při klasické formě výuky musí být dostatečný počet studentů, školící personál a materiály pro výuku v určitém čase na určitém místě. To prodlužuje dobu nově příchozích na zaškolení. Naopak při e-learningové formě je studium lehce dostupné v jakýkoliv čas a to odkudkoliv. Student si sám volí čas, který mu vyhovuje, což přináší chuť i a vůli studovat, tím se částečně zvyšuje efektivita učení.

- **Efektivnost výuky**

U klasické výuky je efektivnost výuky značně proměnlivá a závislá na kvalitách rektora. Porovnávat však efektivnost výuky klasické a podporované e-learningem je poměrně obtížné. Na univerzitě v Hradci Králové byl proveden výzkum pod dohledem PhDr. Ivany Šimonové, pro zjištění rozdílů efektivity u těchto druhů výuky. Základ studijního prostředí tvořil LMS WebCT, kde byli jednotlivé e-předměty provozovány. Studenti experimentální skupiny měli kromě studijních materiálů v e-podobě, možnost procvičit si naučenou látku prostřednictvím testů a kvízů. Výsledkem této studie bylo zjištění, že skupina

využívající ke studiu e-learningovou podporu výuky dosáhli v oblasti úloh ověřující zapamatování, porozumění poznatkům a použití vědomostí v problémových situacích minimálně stejných studijních výsledků v porovnání se studenty vyučovaných prezenčně.

- **Žádné limity pro počet vyučovaných**

Počet studentů, kteří se účastní výuky touto formou je téměř neomezen, omezení je pouze v technické vybavenosti.

- **Snadná aktualizace**

Elektronický studijní obsah lze snadno měnit a aktualizovat, není tedy nutné opětovné vyměňování studijních materiálů jednotlivým účastníkům a dochází ke snížení nákladu (např. za učebnice).

- **Multimediálnost materiálů**

Prezentace studijních materiálů v e-learningu probíhá v elektronické podobě, proto je zde možnost zabudovat multimediální prvky do studijního obsahu. Studenti přijímají informace současně několika smysly. Poslední výzkumy ukazují, že pomocí zraku vnímáme 80% informací, 12% sluchem, 5% hmatem a zbylými 3% ostatními smysly. V běžném procesu výuky je dle těchto výzkumů předáváno 75% informací zvukem - do procesu vnímání je tedy zapojen zejména lidský sluch. E-learning obsahuje multimediální prvky, které se snaží tento nedostatek odstranit, působit na lidský zrak a současně zapojit do vnímání i sluch. Vyrovnané smyslové vnímání s převahou zraku je pro zapamatování a znovuobnovení informací velice důležitá.

- **Aktivita**

Tradiční studijní materiály, nepodporují ve studentovi dostatečnou aktivitu a vystavují ho do role pasivních příjemců informací. Z pedagogických studií jednoznačně vyplývá, je-li žák vtažen do problému, je-li aktivizován, má-li pohled na problematiku z více stran pak dosahuje mnohem vyšších výsledků.

- **Aktivita**

Při čtení učebnic a skript je kladena minimální aktivita na studenta, u multimediálních materiálů je student daleko více nucen vynaložit nějakou aktivitu, ať už v podobě odpovědí na otázky nebo spouštění ukázkových příkladů.

- **Propracovaná verifikace**

Každý e-learningový systém by měl umožnit pomocí testovacích objektů a řídicích systémů kvalitní verifikaci znalostí studenta. Lze například nastavit požadované cíle a ty pak jednoduše změřit. Formou verifikace bývají často

testy, nebo otevřené úkoly (korespondenční práce, seminární práce). Ohodnocení bývá bodové nebo slovní, přičemž slovní hodnocení má mnohem větší vypovídací hodnotu, ale je obtížnější pro tutorů. E-learning rovněž statisticky hodnotí úspěšnost jednotlivých kurzů, z výsledků je pak patrné, které kurzy je potřeba přepracovat.

- **Komunikační možnosti**

Potřebuje-li student něco prokonzultovat s tutorem nebo s ostatními studenty, má k dispozici nepřeborné množství nástrojů pro komunikaci jak pro synchronní, tak asynchronní. Nejčastější způsob bývá přes e-mail nebo pomocí diskusních příspěvků.

- **Náklady vs. Příjmy**

U klasického vzdělávání se náklady nesnižují po celý vzdělávací proces (lektori, školící prostory, výroba školních materiálů). Narozdíl od e-learningu, který je finančně náročný především v začátcích zavádění. V dalších fázích již náklady klesají a jsou využity zejména na udržení provozu kurzu. Náklady můžeme rozdělit do 4 základních skupin - náklady na technologie, lidské zdroje, vytváření obsahu a služby. Technologické náklady představují technické prostředky, e-learningové systémy a multimediální zařízení. Vytváření obsahu zahrnuje nákup programů na tvorbu e-learningových kurzů a aktualizaci obsahu, lidské zdroje pak samotnou tvorbu kurzů, programování a převod existujících materiálů do e-learningové podoby. Náklady za služby zahrnují tvorbu konceptu kurzů, řízení a organizaci e-learningových projektů, provozní podporu pro uživatele.

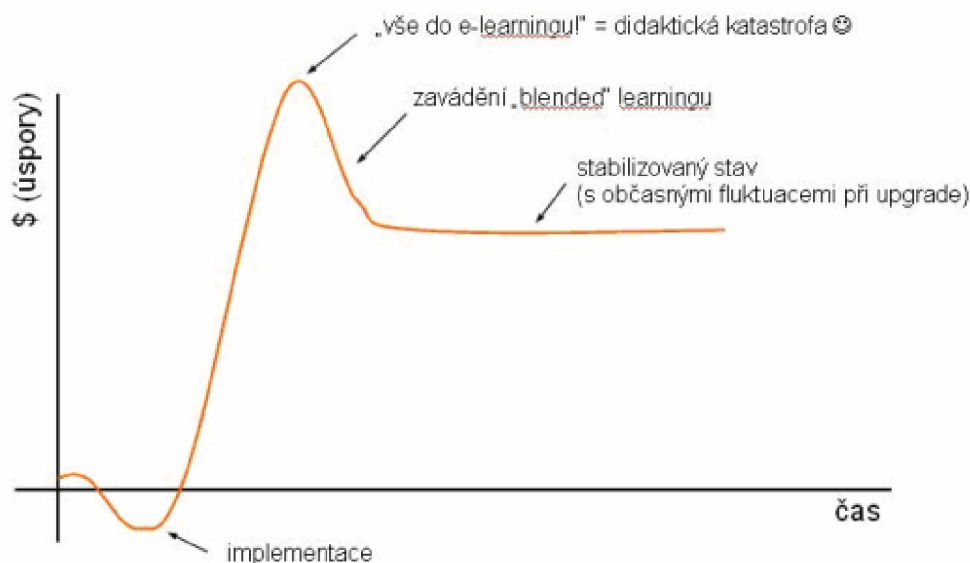
- **Individuální tempo studia**

Při e-learningové výuce si každý vybírá vlastní tempo studia, narozdíl od klasické výuky, kdy rychlost probírání látky se podřizuje pomalejším studentům.

1.1.2 Nevýhody e-learningu

- **Závislost na technologiích**

Při realizaci on-line výuky musí být účastník připojen k internetu či jinou počítačovou sítí a počítač, který je schopen zobrazovat internetové stránky. U některých vzdělávacích prostředích je absence internetu někdy odstraněna pomocí synchronizací dat. Uživatel má stažený studijní obsah ve svém počítači a má k dispozici studijní prostředí bez síťové podpory. Pouze posílá svoje studijní výsledky nebo diskusní příspěvky a nemusí být permanentně připojen k síti.



Obr. 1.1: Grafické znázornění finanční náročnosti e-learningu na čas [6]

- **Složitá tvorba obsahu**

Tvorba je nejenom náročná na čas, ale také na obsah, neboť ten musím mít svou určitou formu a posloupnost.

- **Ne pro každého**

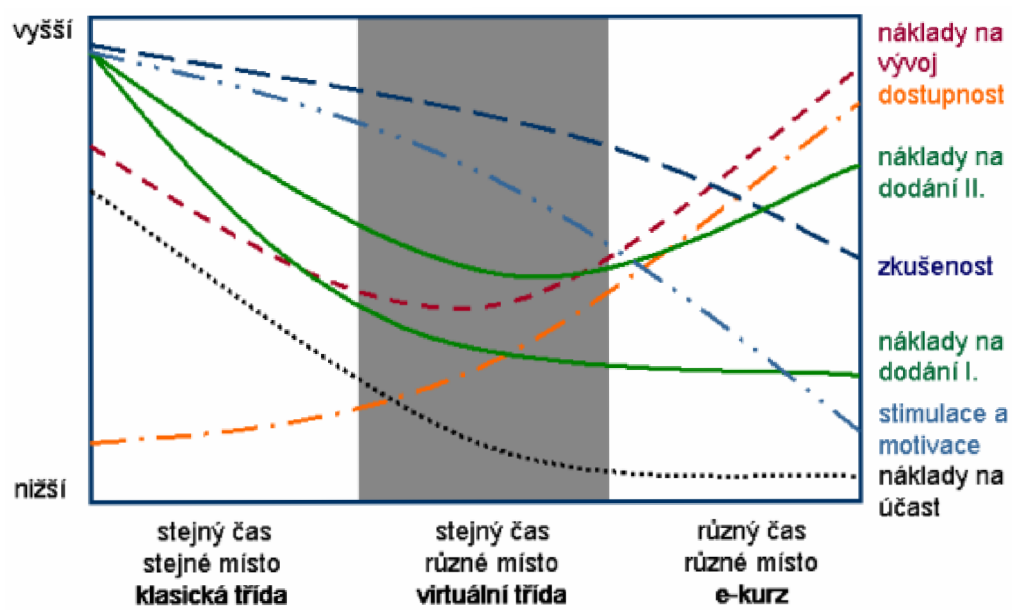
Ne každý student zvládá učení z elektronického textu, někteří potřebují mít k dispozici papírovou podobu, kde si mohou vpisovat poznámky, zvýrazňovat pasáže apod.

- **Nevhodnost pro oblasti vzdělávání**

E-learning je vhodný pouze pro některé oblasti vzdělávání. Nehodí se na výuku, která vyžaduje praktický přístup k aplikaci znalostí, provádění rychlých rozhodnutí, podporu spolupráce s ostatními členy a vnímání řeči těla ostatních studujících.

- **Obtížné vyhledávání**

Vyhledávání multimedií (obrázky, video, zvuky) je samo o sobě obtížné, ale ani text obsažený v kurzu nelze dobře nalézt, neboť většina vývojových nástrojů převádí výsledný kurz do formátu, ve kterém vyhledávat nelze. Vyhledávání se proto omezuje na vyhledávání v popisech kurzů, což je velice neefektivní.



Obr. 1.2: Grafické znázornění výhod a nevýhod jednotlivých forem výuky [5]

2 VYHLEDÁVÁNÍ KLÍČOVÝCH SLOV

V dnešní době jsou velice často součástí výukových materiálů zvukové záznamy z přednášek. Se vzrůstajícím počtem záznamů však vyhledání potřebného záznamu z jeho popisu nebo vyhledání přímo požadovaného úseku záznamu je velice zdoluhavé. Proto vznikají systémy, které dokáží v akustickém signálu nalézt hledané slovo nebo posloupnost slov. Tyto systémy se pak dají použít v mnoha dalších oborech.

2.1 Dělení rozpoznávání řeči podle různých kritérií

Podle složitosti dělíme systémy na rozpoznávání řeči:

- **Izolovaná slova** – mezi jednotlivými slovy jsou pauzy pro jednodušší práci systému, který tak pozná začátek a konec slova. Akustický signál je rozčleněn na menší bloky.
- **Plynulá řeč** – jedná se o běžnou řeč, tedy mezi slovy jsou minimální nebo žádné pauzy a systém má složitější vyhledávání než v prvním případě.
- **Spojená slova (omezený slovník)**

Rozpoznávání řeči dle závislosti na konkrétním řečnickovi:

- **Rozpoznávání závislé na řečnickovi** – systém je schopen rozumět pouze řečnickovi, který ho natrénovával. V tomto případě je pak systém mnohem spolehlivější.
- **Rozpoznávání nezávislé na řečnickovi** – systém rozezná řeč kteréhokoli řečníka, vzory jsou tvořeny zprůměrováním velkého počtu mluvčích.

Podle počtu slov, které systém rozezná (velikost slovníku):

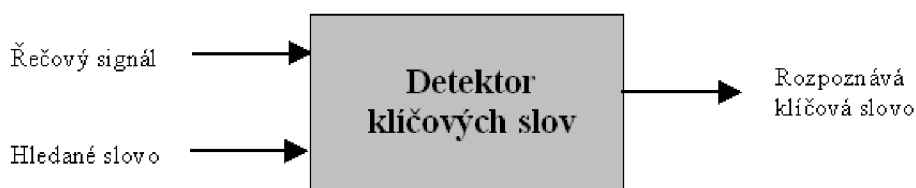
- **Systémy s malým slovníkem** – desítky až stovky slov
- **Systémy se středně velkým slovníkem** – tisíce slov
- **Systémy s velkým slovníkem** – desítky tisíc slov

Z hlediska funkčnosti může detektor klíčových slov pracovat v režimech:

- **Jednoprůchodová** – při každém vyhledávání je prohledáván signál
- **Dvouprůchodová** – akustický signál, je nejprve přepsán do textové podoby (proběhne tzv. indexace), požadovaná slova nebo sousloví se pak vyhledávají pouze v databázích vzniklých indexací.

2.2 Podle druhu metody, na které je založeno rozpoznávání

- **Systémy využívající porovnávání se vzorem** – jedná se o systémy pracující na principu porovnávání se vzorem. Používá se hlavně v případech detekce izolovaných slov. Tato metoda porovnává vzory uložené ve slovníku s akustickým signálem rozpoznávaného slova. Slova ve slovníku mají přiřazeny jeden nebo více vzorových obrazů. Pokud se tato metoda použije na hledání slov ve spojitém signálu, je akustický signál analyzován po jednotlivých úsecích, protože slovo může začínat kdekoliv.
- **Systémy využívající neuronových sítí** – tento druh systému se používal hlavně v devadesátých letech minulého století k rozpoznání mluvené řeči. Byla snaha jej aplikovat i na vyhledávání klíčových slov [1]
- **Systémy pracující se statistickým modelem** – statistické metody mohou být uplatněny pouze v případě, že je k dispozici dostatek trénovacích dat, ze kterých metody statistické indukce mohou odvodit obecnější závěry. Ty se pak vztahují nejen na dosud pozorovaná data, ale i na data nová. Naštěstí v oblasti rozpoznávání řeči lze trénovací data pořídit relativně snadno jejich namluvením. Různorodost řečového signálu je poměrně široká, takže k statistickému popisu je třeba použít statistické modely s velkým počtem parametrů. Čím více parametrů statistické modely mají, tím větší množství trénovacích dat musí být k dispozici a tím pracnější je i nastavení klasifikátorů.[1]



Obr. 2.1: Blokové schéma systému detekce klíčových slov

2.3 Podle druhu konstrukce systému rozpoznávání řeči

2.3.1 LVCSR

Pro rozpoznávání spojitě řeči tato metoda využívá velký slovník (LVCSR - Large Vocabulary Continuous Speech Recognition), s jehož pomocí a na základě jazykového modelu přepisuje akustický záznam do textové podoby. Po této fázi máme k dispozici text, na nějž můžeme použít nejenom klasické způsoby indexace, ale také různé inteligentní techniky vyhledávání, založené například na identifikaci výrazů, které se nacházejí pospolu nebo ve frázi. V případě, že vyhledávané slovo se nenachází ve slovníku (tzv. OOV - Out of Vocabulary), pak toto slovo není možné vyhledat. Tento nedostatek se často řeší kombinací s jinými metodami.[2]

Akustický model – je trénován na řečových datech s odpovídajícími slovními přepisy. Ty jsou pomocí výslovnosti slovníku převedeny na sekvence fonémů a na nich je pak natrénována sada akustických modelů pro jednotlivé základní zvuky.

Jazykový model – hlavním cílem je určit pravděpodobnost určité promluvy v daném jazyce.

Výhody:

- Rychlé vyhledávání
- Slova obsažená ve slovníku jsou nalezena s velkou pravděpodobností
- Jednoduchá indexace

Nevýhody:

- Jazykové omezení

- Omezená slovní zásoba slovníku
- Velká výpočetní náročnost
- Nepřesné u slov mimo slovník

2.3.2 Detekce klíčových slov s výplňovým modelem

Princip je podobný jako u LVCSR. HMM (Hidden Markov Model - skryté Markovy modely) modely klíčových slov soutěží v rozpoznávání síti s modely neklíčové části, které jsou paralelně spojeny s modely klíčových slov. Nevýhodou je pomalost této metody neboť zde chybí možnost indexace, naopak výhodou je univerzálnost, přesnost a není zde problém s neznámým jazykem. [2]

Mezi často používané výplňové modely patří celoslovní model, který se skládá z několika počtu modelů neklíčových částí, fonémový model, který vzniká paralelním spojením modelů jednotlivých fonémů a model slabik. Velkou výhodou této detekce je přesnost, avšak nevýhodou je pomalost a nenemožnost indexace.

2.3.3 Detekce klíčových slov založená na míře důvěry

Neznámá slova a šумы, jenž se objevují před a za klíčovým slovem, neruší detekční algoritmus. Není zde kladen důraz na správnou formulaci promluv, neboť zde není použit žádný jazykový model. Taktéž není potřeba natrénovat jazykový model nebo naladění parametrů výplňového modelu. Hlavní princip je takový, že jestliže míra pravděpodobnosti daného slova je větší než předem definovaný práh, můžeme přijmout hypotézu, že se jedná o hledané slovo.[2]

2.3.4 Fonetické rozpoznávání řeči

Často se používá v kombinaci s LVCSR neboť umožňuje nalezení slov, které nebyly pomocí LVCSR rozpoznány. Vyhledávaná klíčová slova jsou foneticky přepsána a systém se snaží nalézt přepsanou posloupnost fonémů v záznamu.

Výhody

- Možnost indexace sekvencí fonémů
- Možnost kombinace s LVCSR
- Vyhledá i cizí slova a slova mimo slovník LVCSR

Nevýhody

- Pomalejší než vyhledávání slov obsažených v LVCSR

- Složitá vyhledávací metoda
- Nepřesné

2.4 Přehled systémů na rozpoznávání a vyhledávání řeči

První systémy detekovaly a určovaly pouze jednotlivá slova, tedy každé slovo muselo být zřetelně odděleno pauzou ve výslovnosti od následujícího. Další systémy pak dokázaly rozložit slovní spojení na jednotlivá slova. Nejsložitější a nejvýkonnější systémy slouží k rozpoznávání plynulé řeči, které musí pro správné rozpoznání jednotlivých slov určit začátek a konec každého slova. Hlavní parametry udávající kvalitu systému jsou procentuální úspěšnost rozpoznání slov a rychlost zpracování. Systémy na vyhledávání a rozpoznávání řeči našly uplatnění v mnoha oborech.

2.4.1 Systém pro automatický přepis televizních a rozhlasových pořadů

Systém vznikl v Laboratoři počítačového zpracování řeči v Liberci. Systém slouží pro automatický přepis televizních a rozhlasových pořadů, ve kterém pak už lze jednoduše vyhledávat. Přepis je rozdělen do několika kroků, v první části je záznam rozdělen na části, které obsahují řeč a na části bez řečových slov - například znělky, hudba a podobně. Systém dále rozezná podle charakteru akustického signálu části mluvené různými osobami. Což pak velice zpřesní překlad, neboť systém je natrénován pro jednotlivé moderátory. Příspěvky jsou pak odeslány do modulu rozpoznávání řeči, výstupem z tohoto modulu je textový přepis. Čas potřebný pro přepis se pohybuje okolo čtyřnásobku délky překládaného záznamu. Úspěšnost u televizního zpravodajství se pohybuje okolo 70% , v případě rozhlasových zpráv je to 80%.

2.4.2 HP Speech Bot

Patří mezi první aplikace pro vyhledávání v multimediálních záznamech nasazené v reálném provozu. Používá se na rádiových a televizních stanicích, které vysílají po internetu, kde indexuje jejich vysílání. Využívá metody LVCSR, která převádí záznam na text. Mezi hlavní přednosti patří rychlost a přesnost vyhledávání.

2.4.3 Blinkx

Využívá technologie Autonomy a SoftSound, které byly zkoumány na univerzitě Cambridge. Slouží na vyhledávání videa na stránkách www.blinkx.tv. Prozatím bylo indexováno přes 18 miliónů hodin videa.

2.4.4 Nexidia Nexaminer Enterprise

Technologie umožňuje nalezení dat ze zvukového záznamu použitím fonetického rozpoznávání. Technologie efektivně vyhledává bez ohledu na pohlaví, věk, nářečí nebo mluvicí styl, rozeznává i vlastní jména. Rychlost analýzy je přibližně 83krát menší než reálná část prohledávaného záznamu. V současné době je podporováno 33 jazyků, není však problém jednoduše doinstalovat další. Systém je využíván hlavně vládními institucemi, zdravotnickou a záchrannou službou.

2.4.5 CallManager Eureka!

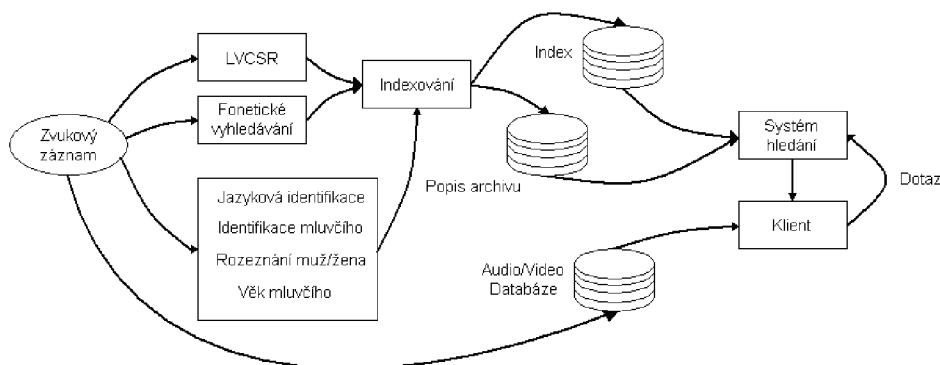
Používá se hlavně k objevení významu každého volání a zjištění zákaznických problémů a smysluplnou reakci na jednotlivé hovory v call centrech. Analýza probíhá ve dvou krocích, nejdříve jsou identifikovány jednotlivé úseky akustických dat (slova, ticho, fráze) a v druhém analyzuje význam těchto úseků.

2.5 Indexace a pattern

Protože vyhledávání dle různých parametrů přímo v prohledávaných souborech je časově velmi náročné, vznikly databáze, které mají hledání výrazně urychlit a zjednodušit. Základním prvkem databáze je index. Speciální programy prochází soubory k vyhledávání a indexují jejich obsah, tedy např. určují výskyt jednotlivých slov, prochází metadata nebo vytváří vektory rysů u obrázků (barevné histogramy apod.). Jednotlivým indexům jsou tedy přiřazeny názvy zástupce, jeho popis a umístění. Velmi jednoduchou metodou pro indexování textových dat, vzniklých například přepisem ze zvukových záznamů, je frekvenční tabulka. Kdy řádky představují klíčová slova a sloupce nálezy slova v dokumentech. Vyhledávací algoritmy musí vědět, co vyhledávají, k čemuž slouží tzv. patterny nebo-li vzory. Podobnosti těchto vzorů se vzory z prohledávaných souborů se pak vyhodnocují.

2.6 Architektura systému na vyhledávání v řečových databázích

Ucelený systém pro vyhledávání v řečových databázích by měl obsahovat procesy na segmentaci záznamů na ticho a řeč, identifikaci jazyka, řečníka, odhad jeho věku a pohlaví.



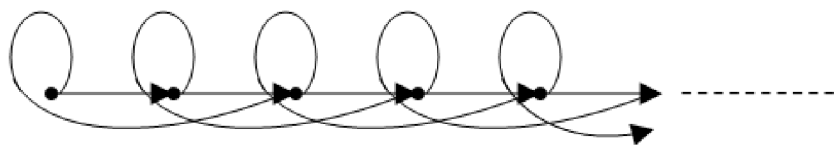
Obr. 2.2: Základní struktura systému pro vyhledávání ve zvukových záznamech [4]

2.7 Skryté Markovovy modely (HMM)

Nejpoužívanější a také nejúspěšnější metodou pro rozpoznávání a vyhledávání slov je metoda skrytých Markovových modelů. Vychází z principu skutečného vytváření řeči. Jestliže řeč rozdělíme na jednotlivé krátké úseky - mikrosegmenty, (které trvají přibližně 20ms), pak v každém úseku dosahuje artikulační konfigurace konečného počtu stavů (vytváří se foném). V tomto úseku se vytváří signál, jehož charakteristika je závislá na stavu artikulačního ústrojí. Z těchto charakteristik je pak vytvořena kódová kniha. Pomocí kódové knihy pak lze popsat jakoukoliv charakteristiku. Při modelování řeči se generují dvě časově na sobě závislé posloupnosti, a to řetězec spektrálních vzorů a podpůrný Markovův řetězec. Podpůrný Markovův řetězec mění své stavy podle matice pravděpodobností. Tyto stavy ale nejsou z vnějšího pohledu patrné, viditelné jsou pouze výstup náhodných funkcí, odtud také odvozen název metody. Levo-pravý Markovův model, využívající se při modelování řeči, probíhá od příchodu prvního spektrálního vzoru z počátečního stavu modelu a končí příchodem posledního vzoru. Se vzrůstajícím časem se indexy stavů buď zvyšují nebo zůstávají stejné.

2.7.1 Vintsyukův typ

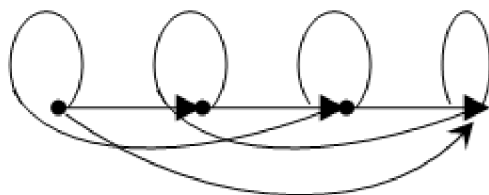
Na každé slovo připadá asi 40-50 stavů. Tento systém pak potřebuje velké množství trénovacích dat.



Obr. 2.3: Vintsyukův typ Markovova modelu

2.7.2 4-stavový skrytý Markovův model

Protože u vintsyukově typu si sousední stavy byly velice podobné, došlo k jejich sloučení a redukcí počtu stavů a to bez větší ztráty úspěšnosti metody. Trénování u tohoto modelu bylo mnohem jednodušší.



Obr. 2.4: 4-stavový skrytý Markovův model

Hlavní procesy skrytých Markovových modelů:

- **Určení pravděpodobnosti promluvy.** V tomto kroku dojde ke zpracování promluvy a určení pravděpodobnosti. U jednotlivých segmentů jsou určeny vektory příznaků, tyto vektory jsou pak nahrazeny indexy vzoru z kódové knihy, kterému se nejvíce podobají. Pro určení pravděpodobností se používají hlavně tyto metody: Viterbiho algoritmus, výpočet odpředu (forward) a výpočet odzadu (backward).
- **Trénování parametrů modelu.** Pro každé slovo je natrénován jeden Markovův model.

- **Rozhodovací kritérium.** Proces vyhodnocuje data na principu maximální pravděpodobnosti. Pro slova je určena pravděpodobnosti s jakou by stavy generoval markovovův natrénovaný model a vybereme slovo s nejvyšší pravděpodobností.

2.8 Algoritmus dynamic time warping (DTW)

Stejná slova vyslovená jedním řečníkem mají různou délku trvání a to nejenom celkovou, ale rozdíly jsou i uvnitř slova - různá délka odpovídajících si fonémů. Jak už z názvu algoritmu vyplývá, dochází k borcení jedné z časových os (nelineární časová normalizace) a tím jsou minimalizovány časové rozdíly porovnávaných obrazů.

Postup výpočtu:

Máme dva obrazy (posloupnosti příznaků) hledaného (referenčního) slova:

$$A = \{a_{(1)}, a_{(2)}, \dots, a_{(i)}\},$$

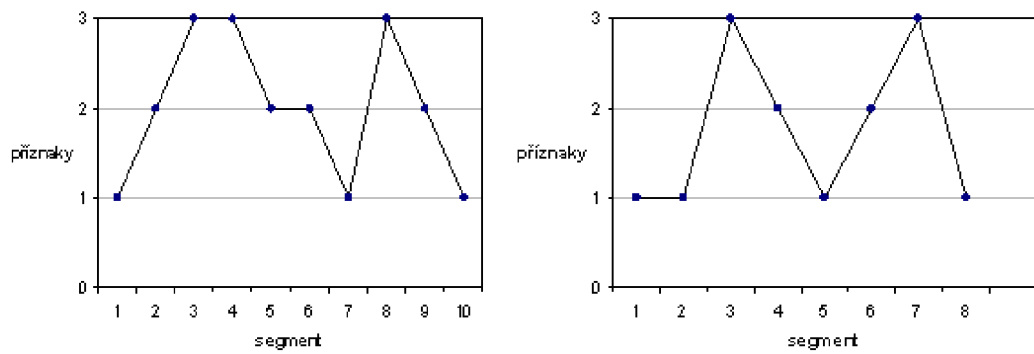
a obraz (testovaného) slova z vyhledávaného záznamu

$$B = \{a_{(1)}, a_{(2)}, \dots, a_{(j)}\}$$

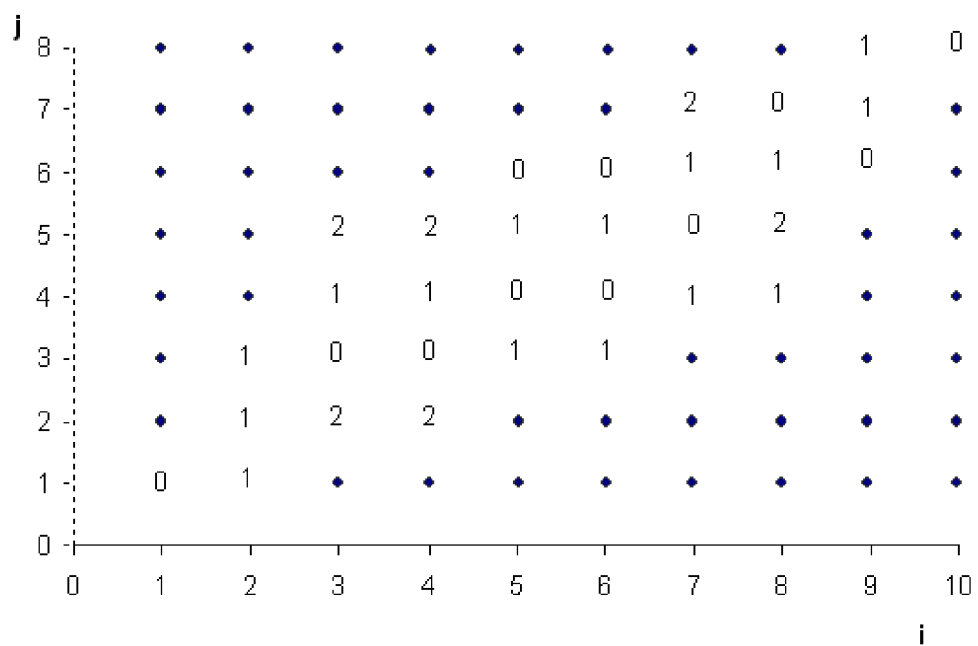
kde $a_{(i)}$ je i -tý příznak referenčního obrazu a $a_{(j)}$ je j -tý příznak testovaného obrazu. Z lokálních vzdáleností se vytvoří matice $I \times J$ lokálních vzdáleností obou obrazů. Uvnitř této matice se pak hledají cesty podle určitých kritérií a omezení. Cesta vede z prvku matice o souřadnicí $(1, 1)$ k prvku (I, J) . Cesta může ve směru vodorovné osy postupovat pouze o jeden krok, ve svislém směru pak o nula, jeden či dva kroky, avšak nikdy ne o krok zpět. Optimální cesta je pak stanovena na základě akumulovaných vzdáleností, tedy součet hodnot vzdáleností podél optimální cesty z počátečního bodu $(1,1)$ do bodu (I, J) .

$$D_{(i,j)} = d_{(i,j)} + \min [D_{(i-1,j)}, D_{(i-1,j-1)}, D_{(i-1,j-2)}]. \quad (2.1)$$

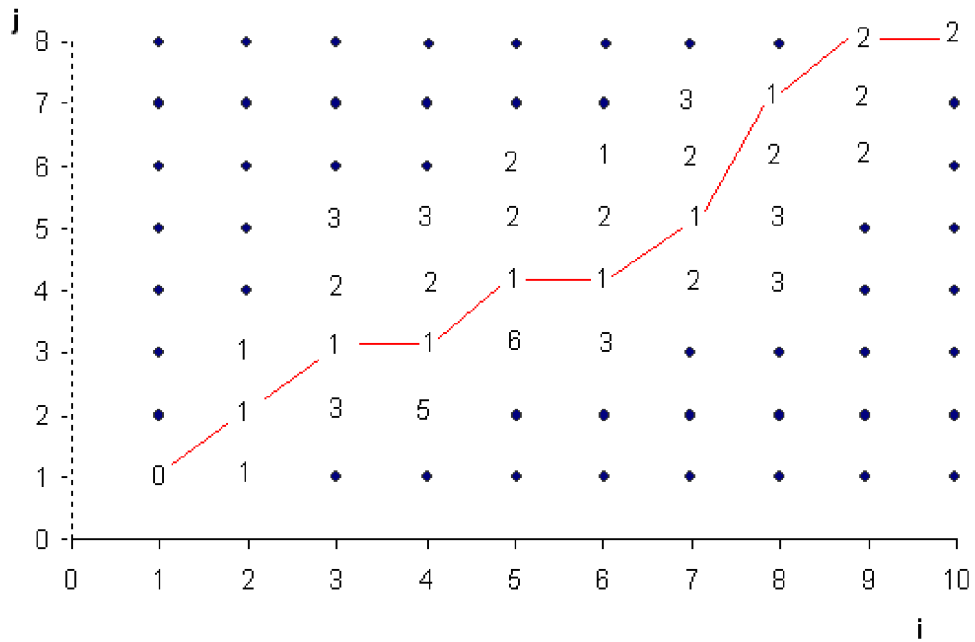
Schématické znázornění výpočtu:



Obr. 2.5: Příznaky obrazu A a B



Obr. 2.6: Lokální vzdálenosti obrazů A a B



Obr. 2.7: Akumulované vzdálenosti a minimální cesta

2.9 Funkce počet průchodů nulou

Jedná se o charakteristiku popisující vlastnost signálu. Průchod nulou u signálu nastane, jestliže dva sousední vzorky mají rozdílné znaménko, tedy kolikrát za určitý úsek projde signál nulou. Znělé hlásky mají menší počet průchodů nulou než neznělé hlásky. Tuto funkci lze definovat jako:

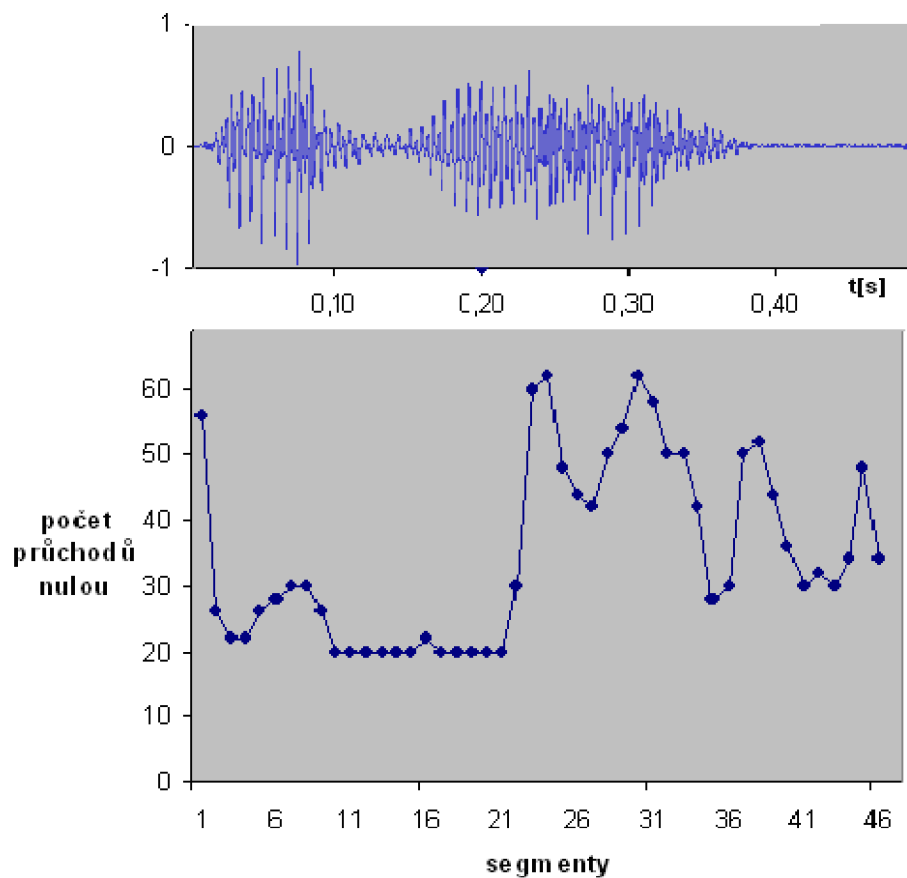
$$Z_i = \sum_{k=-\infty}^{\infty} |[\text{sgn}(k)] - \text{sgn}[s(k-1)]| w_i(k), \quad (2.2)$$

kde:

$$\text{sgn}[s(k)] = \begin{cases} 1 & \text{pro } s(k) \geq 0 \\ -1 & \text{pro } s(k) < 0 \end{cases},$$

$$w_i(k) = \begin{cases} 1 & \text{pro } \frac{N(i-1)}{2} + 1 \leq k \leq \frac{N(i+1)}{2} \\ 0 & \text{jinak} \end{cases},$$

k je aktuální vzorek, N počet rámců a w_i je rámeček, kterému přiřadíme příznak.

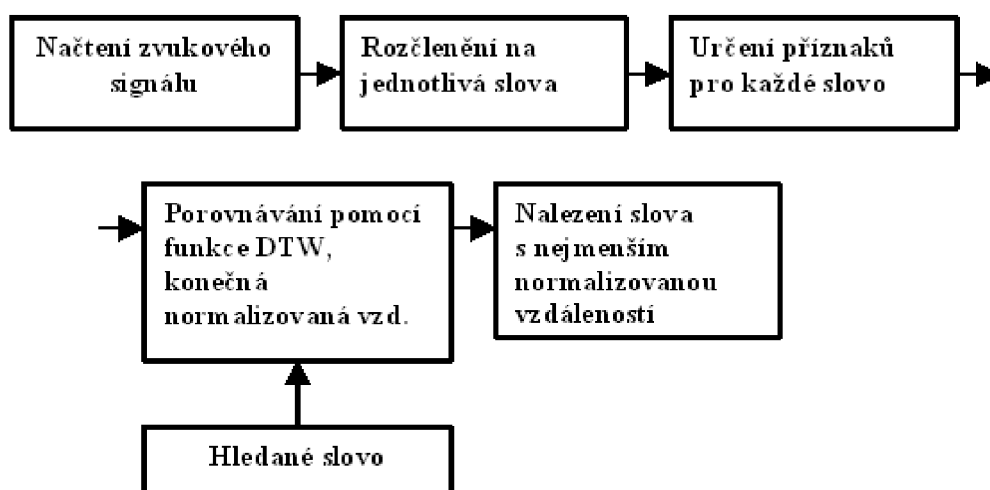


Obr. 2.8: Signál slova jedna a jeho analýza počtem průchodů nulou

3 NÁVRH A REALIZACE JEDNODUCHÉHO VYHLEDÁVAČE

3.1 Návrh vyhledávače klíčových slov

Ve svém návrhu vyhledávače v audio záznamech jsem použil konstrukci systému založeného na porovnávání se vzorem, dále byla použita funkce DTW (Dynamic Time Warping) a funkce středního počtu průchodů signálu nulou.



Obr. 3.1: Blokové schéma navrhnutého vyhledávače

Postup vyhledávání systému:

V prvních krocích je akustický signál rozdělen na jednotlivá slova. V dalším kroku jsou určeny příznaky daných slov pomocí krátkodobé funkce středního počtu průchodů signálu nulou. Slova se rozdělí na mikrosegmenty (přibližně 10-40 ms) a příznak Z_i pro i -tý mikrosegment vočteme následovně:

$$Z_i = \sum_{k=-\infty}^{\infty} |[\text{sgn}(k)] - \text{sgn}[s(k-1)]| w_i(k), \quad (3.1)$$

kde:

$$\text{sgn}[s(k)] = \begin{cases} 1 & \text{pro } s(k) \geq 0 \\ -1 & \text{pro } s(k) < 0 \end{cases} ,$$

$$w_i(k) = \begin{cases} 1 & \text{pro } \frac{N(i-1)}{2} + 1 \leq k \leq \frac{N(i+1)}{2} \\ 0 & \text{jinak} \end{cases} .$$

Porovnávání příznaků pomocí algoritmu DTW:

Soubor příznaků hledaného slova je porovnáván se souborem příznaků slov v audio záznamu pomocí DTW funkce, která rekurentně vytvoří prvky $g(i, j)$ matice $\mathbf{G}(I \times J)$ podle:

$$g_{(1,1)} = d(A_{(1)}, B_{(1)}) , \quad (3.2)$$

$$g_{(i,j)} = \min \begin{cases} g_{(i,j-1)} + d(A_{(i)}, B_{(j)}) \\ g_{(i-1,j-1)} + 2d(A_{(i)}, B_{(j)}) \\ g_{(i-1,j)} + d(A_{(i)}, B_{(j)}) \end{cases} ,$$

$d(A_{(i)}, B_{(j)})$ je vzdálenost mezi i -tým příznakem hledaného (vzorového) slova A a j -tým příznakem slova v záznamu B , kterou definujeme vztahem Konečná normalizovaná vzdálenost pro slovo A a slovo B :

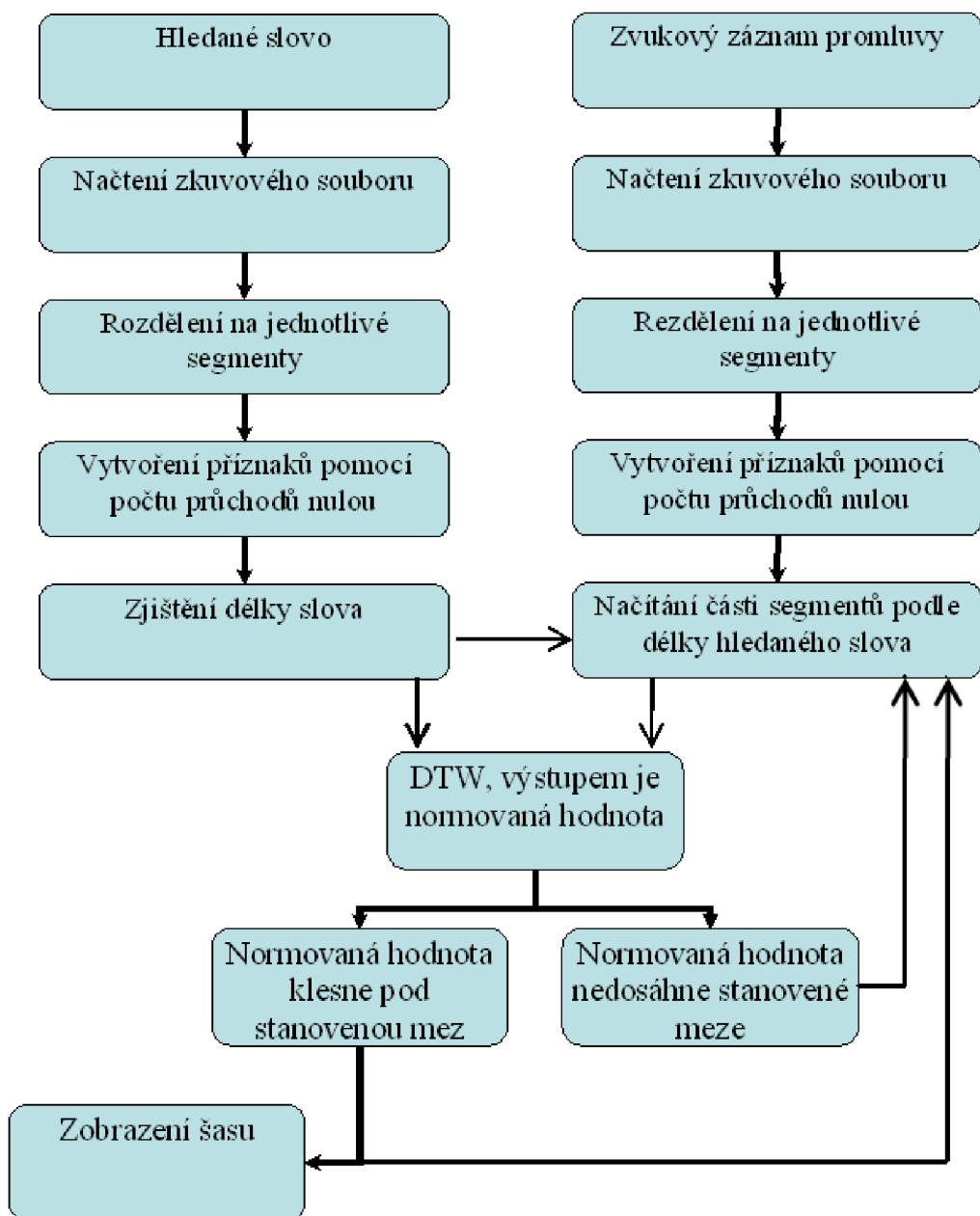
$$D_{(A,B)} = \frac{g_{(I,J)}}{I + J} . \quad (3.3)$$

Výsledkem jsou konečné normalizované hodnoty, pak slovo, které má nejnižší hodnotu této normalizované vzdálenosti je pravděpodobně hledané slovo.

3.2 Realizace vyhledávače slov ve zvukovém záznamu

Program pro vyhledávání slov ve zvukovém záznamu je vytvořen v programovém prostředí Matlab. Celá architektura výpočtů je rozdělena do třech M-souborů (`DTW.m`, `pruchody_0.m` a `hledej.m`). V první fázi

práce programu dojde k načtení zadaného zvukového záznamu slova, které chceme vyhledávat a zvukového záznamu, ve kterém budeme vyhledávat. Jediný podporovaný zvukový formát je wav, oba zvukové soubory musí být ve stejné vzorkovací frekvenci, aby byly zachovány stejné délky mikrosegmentu. Po načtení souborů dochází k jejich rozložení na jednotlivé mikrosegmenty (délka jednoho mikrosegmentu pro vzorkovací kmitočet zvuku 8kHz je 20ms). Pro každý mikrosegment je pomocí funkce počet průchodů nulou, stanoven příznak, charakterizující mikrosegment. Jak hledané slovo, tak i záznam jsou popsány posloupností hodnot. Dalším krokem je zjištění délky hledaného slova, aby mohlo být porovnáváno s podobně velkými úseky ze záznamu. Po nastavní délky, jsou ze záznamu postupně vyjímány úseky, které odpovídají délce hledaného slova. Vyjmutý úsek je porovnáván s hledaným slovem pomocí funkce dynamického borcení časové osy. Výstupem funkce je normovaná hodnota, čím je hodnota nižší, tím více si jsou porovnávané obrazy podobné. Pokud normovaná hodnota klesne pod určitou mez, dojde k výpočtu času a následného zobrazení předpokládaného výskytu slova. Tímto způsobem se prochází celý záznam do posledního segmentu.



Obr. 3.2: Architektura práce vytvořeného programu

3.2.1 Úspěšnost vyhledávače

Hlavním problémem při určování úspěšnosti vyhledávače je stanovit mez normované hodnoty, kdy je slovo určeno jako vyhledávané. Při testování zvuků zaznamenaných na různých mikrofonech, mezní hodnota měla pro stejné slova jinou hodnotu, což může být způsobeno různou úrovní šumu na výstupu mikrofónů. Při testování byla použita

vždy nejvyšší hodnota hledaného slova a všechny slova, která měla nižší normovanou hodnotu byla označena jako slova nalezená. Rychlost vyhledávače je na minutu záznamu jedna hodina výpočtu, avšak při přeskokování jednoho segmentu poloviční, aniž by došlo ke snížení úspěšnosti. Zvukový záznam byl nahrán mikrofonem genius HS-04V. Vzorkovací frekvence záznamu byla 8kHz, tedy jeden segment měl velikost 20ms. Testovaný zvukový soubor obsahoval téměř 200 slov a 1300 znaků. Výsledky vyhledávače jsou zobrazeny v následující tabulce.

Tab. 3.1: Ukázka úspěšnosti vyhledávače

slovo	nor. hodnota	celkový počet	správně	špatně	celkem nalezeno
praxe	29,30	1	1	2	3
systém	29,50	5	4	5	9
teoretické	27,50	3	2	2	5
síťových	28,30	7	5	4	9
telekomunikační	26,60	1	0	3	3
operačních	24,90	2	2	3	5
absolvent	24,53	2	2	3	5
provoz	25,20	2	2	3	5
moderní	27,30	1	0	3	3
řízení	27,25	1	1	1	2

4 ZÁVĚR

Cílem této práce bylo v teoretické části shrnutí výhod a nevýhod e-learningové formy vyučování. Podrobnější rozbor problematiky vyhledávačů slov ve zvukovém záznamu, jejich rozčlenění dle různých kritérií, metody a způsoby vyhledávání, včetně přehledu existujících systémů pro práci se zvukovými záznamy a jejich využití. Cílem praktické části bylo navržení jednoduchého vyhledávače slov ve zvukovém záznamu a jeho následná realizace. Navržený vyhledávač byl vytvořen pomocí programového prostředí Matlab. Jádro programu tvoří několik funkcí, které využívají algoritmus dynamického borcení časové osy pro porovnání obrazů a funkce počtu průchodů nulou, zajišťující analýzu obrazů. Jak je patrné z tabulky 1, nalezení vyhledávaných slov bylo poměrně úspěšné, ale výsledek obsahoval také velký počet špatně označených slov, což může být způsobeno například šumem, na který je funkce počet průchodů nulou poměrně náchylná. Řešením odstranění tohoto problému, by mohlo být použití další funkce na popis signálu, která by špatně označená slova výrazně redukovala, snížení šumu jeho odfiltrováním nebo použitím kvalitnějšího mikrofonu. Jako hlavní problém však spatřuji délku času, který je potřeba na prohledání záznamu. Je možné částečně snížit časovou náročnost výpočtu přeskokování sousedních segmentů, avšak pouze o dva segmenty, při delších skocích už dochází ke snížení úspěšnosti vyhledávání. Dalšími možnostmi rozšíření by mohlo být vytvoření samostatně spustitelného programu a jeho nezávislost na Matlabu.

LITERATURA

- [1] ROČEK, Radovan. Nové možnosti a trendy v oblasti elektronického vyučování. Brno, 2007. 34 s. Bakalářská práce.
- [2] JURČÍČEK, Filip. Dekodér systému rozpoznávání souvislé mluvené řeči s velkým slovníkem (LVCSR) s n-gramovým jazykovým modelem. Plzeň, 2003. 44 s. Diplomová práce.
- [3] KOPECKÝ, Kamil. E-learning (nejen) pro pedagogy. Olomouc : Hanex, 2006. 130 s. ISBN 80-85783-50-9.
- [4] SMRŽ, Pavel. Vyhledávání v záznamech přednášek. [s.l.] : [s.n.], 2006. Dostupné z URL:
<<http://acm.osu.cz/paper/tpev2005/TPEV2005-sbornik.pdf>>.s. 75.
- [5] PRAVDA, Václav. Fenomén e-learningu v současném vzdělávání. 2006. Praha : [s.n.], 2006. Dostupné z URL:
< <http://www.e-univerzita.cz/old/2003/b0529002.doc>>.
- [6] BALARIN, David. Náklady a efektivita e-learningu [online]. 2006. Praha : 2006 [cit. 2008-01-18]. Dostupné z URL:
< www.hrzive.cz/data/2006-05-05/david-balarin.pptxKqc78q93sl9fFg>.
- [7] PSUTKA, Josef. Komunikace s počítačem mluvenou řečí. 1995. vyd. Praha : Academia, 1995. 295 s. ISBN 800-200-0203-0.
- [8] KOVAŘÍK, Martin. Počítačové zpracování dat v programu MATLAB. Bučovice : Nakladatelství Martin Stříž, 2005. 278 s. ISBN 978-80-87106-09-9.

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

$D_{(i,j)}$ celková lokální vzdálenost

$d_{(i,j)}$ lokální vzdálenosti

$D_{(A,B)}$ normovaná vzdálenost

DTW dynamické borcení časové osy – Dual Time Warping

HMM skryté Markovovy modely – Hidden Markov Models

k aktuální vzorek

LMS řídicí vyukový systém – Learning Management System

LVCSR velký slovník rozpoznávání spojitě řeči – Large Vocabulary
Continuous Speech Recognition

N počet rámců

OOV slovo se nenachází ve slovníku – Out of Vocabulary

Z_i i -tý příznak

SEZNAM PŘÍLOH

CD s vytvořenými M-fily, návodem a ukázkovými zvukovými soubory