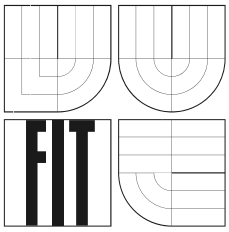
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ



Strukturování dokumentů, nadpisy kapitol, vkládání tabulek a obrázků

Projekt ITY č. 3

2007 **Pavel Smrž**

Abstrakt

Příspěvek se zabývá aplikací moderních metod rozpoznávání mluvené řeči v oblasti e-vzdělávání. Popisuje systém pro indexování a vyhledávání v záznamech přednášek. Výstupem rozpoznávače řeči je acyklický graf hypotéz, takže nelze použít existující řešení pro vyhledávání běžného textu. Pro efektivní vyhledávání v rozsáhlé datové struktuře byl implementován a optimalizován speciální indexační systém. Jednotlivým cestám v grafu, kterým odpovídají posloupnosti slov, jsou přiřazeny váhy odpovídající pravděpodobnostem výskytu daného řetězce slov. Stručně jsou prezentovány výsledky systému a možnosti dalšího použití ve vzdělávání.

Klíčová slova

LaTeX, technologie ve vzdělávání

Obsah

Kapitola 1

Úvod

Záznam přednášek se pomalu stává standardní součástí elektronické podpory vzdělávání na mnoha světových i českých univerzitách. V minulé dekádě se výzkum v této oblasti zaměřil především na automatizaci tvorby multimediálních učebních materiálů kombinujících tradiční výukové podklady (texty, obrázky, multimediální prezentace) se záznamy přednášek.

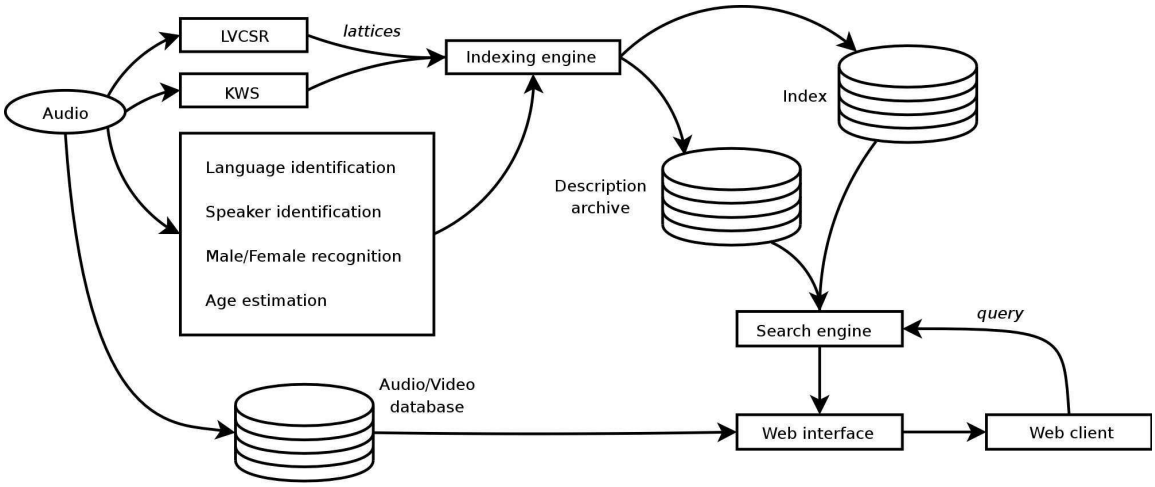
Výzkum v posledních letech se zaměřuje na pokročilé metody dalšího zpracovávání záznamů, jejich indexování, vyhledávání a prezentaci uživatelům. Vyhledávání v rozsáhlých kolekcích dokumentů patří dnes k běžným činnostem mnoha uživatelů Internetu. Standardní je hledání textových dokumentů, pro něž byly implementovány velmi efektivní metody indexování. Vyhledávání obrázků, hudby či videa je často založeno na metadatech, doplňkových informacích, titulcích atd., které mají rovněž textovou podobu. Objevují se však i systémy specializované na konkrétní druh multimediálních materiálů. Této oblasti se věnuje náš příspěvek, který představuje prototyp systému pro vyhledávání v záznamech přednášek.

Přímé indexování a prohledávání obrazové informace zachycované kamerou není pro záznamy přednášek vhodné. Pokud však výzkum v této oblasti pokročí, kterého se náš tým účastní), je možné si představit integraci technologie umožňující např. vyhledání části přednášky, kdy bylo demonstrováno dýchání z úst do úst. Zásadní roli, která je při splnění určitých požadavků již v současnosti zvládnutelná, dnes ale hraje prohledávání a indexování přednášek založené na technologii rozpoznávání řeči.

1.1 Schéma systému

Systém pro vyhledávání v řečových záznamech přednášek zahrnuje celou řadu procesů - moduly segmentace záznamů na ticho a řeč, segmentace řečníků (dotazy posluchačů a diskuse), identifikace jazyka, řečníka, pohlaví, odhad věku, rozpoznávání, odhad pravděpodobnosti rozpoznaných hypotéz, indexování grafu hypotéz a rozhraní pro práci s databází a pro vyhledávání (viz obrázek 1.1).

Indexaci rozpoznaných řečových záznamů (grafů hypotéz) zabezpečuje tzv. indexer. Z každého grafu je vytvořena množina výskytu hypotéz, které jsou uloženy ve vyhledávacím indexu. Každý záznam v této množině obsahuje slovo, pozici v grafu, pravděpodobnost a počáteční a koncový čas. Dále jsou záznamy ve vyhledávacím indexu seřazeny podle identifikátorů slov, čímž vznikne reverzní index. Je také vygenerována tabulka ukazatelů z identifikátorů slov do reverzního indexu. Vyhledávač používá reverzní index, slovník a index dokumentů pro vyhledání zadaného výrazu. Pro vypsání kontextu je používán indexovaný graf hypotéz.



Obrázek 1.1: Základní struktura systému pro vyhledávání v řečových záznamech

1.2 Požadavky na vyhledávací systém a jeho rozhraní

Mezi standardní požadavky na vyhledávací systém patří maximální relevantnost, rychlost hledání a jednoduchost používání. Důležitým aspektem je rovněž prezentace kontextu, v němž se nalezený výraz nachází. Z hlediska uživatelského rozhraní je pro multimediální vyhledávací systém specifická potřeba přehrávat konkrétní část záznamu, v níž systém identifikoval daný výskyt zadaného slova. Vypsání kontextu se také liší od textových systémů ? je nutné vypsat slova na (nejpravděpodobnější) cestě procházející vrcholem reprezentujícím nalezené slovo.

Uživatelské rozhraní vyhledávacího systému musí umožnit snadné zadávání dotazů, nastavování délky požadovaného kontextu, počtu hledaných hypotéz atd. Grafické rozhraní zahrnuje přehrávač video- a audio-proudů, prohlížeč přepisů přednášek, průsvitek a posuvník časové osy. Všechny tyto komponenty musí být vzájemně propojené a synchronizované, tedy např. pohybem po časové ose se aktualizují záznamy, je nalezena relevantní část přepisu a zobrazena příslušná průsvitka.

1.3 Rychlost vyhledávání

Pro testování indexačního a vyhledávacího systému jsme použili rozpoznávač řeči vyvinutý v rámci projektu AMI. Trénování probíhalo na standardní testovací sadě ctstrain04, která je podmnožinou souboru h5train03, definovanou pro účely testování systémů rozpoznávání na cambridgeské univerzitě. Databáze obsahuje okolo 300 hodin anotovaných řečových dat.

I přes značný objem dat je systém schopen najít hledané slovo velmi rychle. Rychlost vyhledávání je samozřejmě silně ovlivněna parametry, které mohou být uživatelsky nastaveny. Omezení okolí nalezeného slova pro zjišťování kontextu udává parametr `--time-delta t`, kde t je čas v sekundách. Uváděný čas je průměrnou hodnotou 10 nezávislých měření.

Kapitola 2

Vyhledávání ve zvukových záznamech

Vyhledávání ve zvukových záznamech přednášek se potýká s mnoha problémy. Na rozdíl od telefonních dialogových systémů, které jsou dnes již poměrně běžné, není například snadné omezit slovník pro rozpoznávání, který často obsahuje odbornou terminologii nepokrytou standardními slovníky. Také zvukové charakteristiky jednotlivých záznamů nebo jejich částí se mohou značně lišit, systémy si např. musí poradit s okolním hlukem, případně se vypořádat s rozpoznáváním více mluvčích (otázky studentů) atd. Existují dva rozdílné přístupy k vyhledávání založenému na řečových technologiích:

1. Metoda LVCSR vytváří přibližný textový přepis záznamu na základě jazykového modelu. Pokud se omezíme na nejpravděpodobnější posloupnost slov (viz dále), dostáváme tedy přímo text a můžeme aplikovat běžné způsoby indexování. Navíc lze použít veškeré inteligentní techniky vyhledávání, založené např. na identifikaci výrazů, které se často vyskytují spolu s hledanými klíčovými slovy nebo frází. Nevýhodou tohoto přístupu je nemožnost rozpoznání slov, která nejsou obsažena ve slovníku. Bohužel právě odborná terminologie spadá často do této oblasti, pokud se jazykové modely nepřizpůsobí konkrétní oblasti přednášky.

2. Indexování fonetických jednotek namísto slov umožňuje vyhledávání výrazů, které by nebyly pomocí LVCSR rozpoznány. Vyhledávaná klíčová slova jsou foneticky přepsána a systém se snaží nalézt výskyty dané posloupnosti fonémů v záznamu. Nevýhodou může být náročnější algoritmus vyhledávání, časové nároky je však možné redukovat vytvořením vhodných indexovacích struktur. Cenou, kterou v tomto případě platíme, je čas potřebný pro předzpracování záznamů. To může být limitujícím faktorem např. při potřebě okamžité indexace přednášek pro on-line vysílání (streaming).

Na rozdíl od systémů pracujících s běžnými textovými dokumenty je u vyhledávání ve zvukových záznamech často problematické rozhodnout, zda se hledaný výraz v záznamu skutečně vyskytl. Oba přístupy uvedené výše obecně produkují síť (orientovaný acyklický graf) hypotéz, udávající, jaké kombinace slov, resp. kombinace fonémů se s danou pravděpodobností v záznamu nacházejí. Klasické rozpoznávání mluvené řeči řeší tuto situaci výběrem nejpravděpodobnější posloupnosti slov. Takový přístup postačuje pro vyhledávání ve velmi kvalitních řečových záznamech (viz např. systém HP Speech Bot popsaný níže). V případě méně kvalitní nahrávky nastávají však situace, kdy se hledané klíčové slovo v nahrávce skutečně vyskytuje, avšak je ohodnoceno jako (v daném kontextu) méně pravděpodobné. Na nejlepší cestě se tedy vyskytuje nesprávně rozpoznané slovo. Vhodnější je potom implementovat vyhledávání přímo nad grafem hypotéz. Každé slovo se tak v daném místě vyskytuje s určitou pravděpodobností, kterou je potřeba vypočítat a setřídění výskytů provést na základě tohoto ohodnocení. Je samozřejmě také nutné optimalizovat indexy pro

Gramatika	HD	iHD	% HD	sHD	% HD
ATIS	882 673	793 370	89.8 %	390 860	44.2 %
ATIS (H)	401 782	362 568	90.2 %	139 935	34.8 %
PT	1 227 500	510 175	41.5 %	456 736	37.2 %
CT	638 276	606 591	95.0 %	381 115	59.7 %
CZ	994 402	915 004	92.0 %	496 129	49.8 %

Tabulka 2.1: Výsledky vyhledávání pro různé implementované metody

vyhledávání. Tento přístup jsme zvolili i při implementaci našeho systému. Tabulka 2.1 shrnuje výsledky předchozích experimentů.

Kapitola 3

Závěr a směry dalšího vývoje

Prototyp systému pro indexování a prohledávání záznamů přednášek bude v nejbližší době využit pro procházení videozáznamů FIT VUT v Brně. Současně předpokládáme vznik metodologie pro přípravu vzdělávacích materiálů, vytváření a další zpracovávání záznamů a jejich provazování s průsvitkami, případně studijními texty.

Pro plné nasazení systému bude potřeba dokončit integraci s uživatelským rozhraním pro interaktivní přehrávání záznamů Jferret, vytvořený v rámci projektu AMI. Zaměříme se rovněž na lepší integraci systému KWS pro detekci terminologie, která se nenachází ve slovníku, a na optimalizaci velikosti ukládaných záznamů pomocí bitových polí. Pro reálný provoz bude důležité rovněž asynchronní indexování, kdy nejprve vznikne dopředný index, na jehož bázi jiný proces vytvoří slovník a reverzní index. Bude také testována možnost přímé indexace obrazového záznamu, metody zpracování videa by již nyní měli poskytnout např. možnost výběru nejlepšího pohledu, pokud je přednáška snímána více kamerami.