

Posudek disertační práce Mgr. Petry Galuščákové

Information retrieval and navigation in audio-visual archives

Předložená disertační práce se věnuje velice aktuálnímu tématu vyhledávání informací (Information Retrieval – IR) v rozsáhlých archivech obsahujících videonahrávky. V práci je nejprve využit „tradiční“ postup vyhledávání informací, při kterém uživatel specifikuje svoji informační potřebu ve formě textového dotazu. Později je správně argumentováno, že možnost pouhého textového dotazu může být pro vyhledávání informací v archivu obsahujícím též obrazový materiál zbytečně limitující a zaměření práce je přesunuto směrem k tzv. „hyperlinkingu“, kdy je dotaz specifikující informační potřebu uživatele zadán ve formě konkrétního fragmentu videonahrávky (tento přístup je znám také pod pojmem „query-by-example“) a systém vyhledávání pak poskytuje odkazy (hyperlinky) na segmenty tomuto dotazu podobné.

Celá práce je přehledně a logicky členěna do 6 kapitol. V první a druhé kapitole autorka stručně představuje základní principy vyhledávání informací na základě textových dotazů, datové sady využívané pro experimenty a také evaluační míry použité pro vyhodnocení úspěšnosti jednotlivých metod. Kapitola 3 je věnována důležitému problému, se kterým se tvůrce IR systému nevyhnutelně setkává v případě, kdy textová data určená k vyhledávání nesestávají z jednoznačně rozlišitelných jednotlivých dokumentů, ale jsou souvislým proudem textu vzniklým buď manuálním, nebo automatických přepisem. V takovém případě je třeba pro účely efektivního vyhledávání uvedený proud textu rozdělit na uměle vytvořené dokumenty – v ideálním případě tak, aby tyto dokumenty byly tematicky koherentní. Experimenty provedené autorkou disertační práce však ukazují, že dobrých výsledků lze dosáhnout i rozdělením na segmenty s fixní délkou, které se částečně překrývají (détku segmentu a překryv je však nutno pečlivě vyladit). Kapitola 4 je věnována výše zmíněnému hyperlinkingu; jsou zde také navrženy a experimentálně ověřeny metody, které kombinují textovou informaci získanou z přepisů audiostopy s informací akustickou (jako např. prozodická charakteristika řeči či informace o hudební skladbě obsažené na nahrávce) a obrazovou. Stručná kapitola 5 pak popisuje automatický výběr tzv. kotev (anchors), což jsou segmenty vhodné pro následné propojení pomocí hyperlinků. V kapitole 6 je představena implementace většiny v práci otestovaných algoritmů do volně šiřitelného uživatelského rozhraní, které širokému spektru uživatelů umožňuje na ukázkovém archivu vyzkoušet vyhledávání s použitím algoritmů vyhodnocených autorkou práce jako nejperspektivnější.

Text disertační práce ukazuje, že autorka má důkladný přehled o nejnovějších metodách vyhledávání informací s využitím nejen textových, ale i obrazových a do jisté míry i akustických dat. Tato šířka záběru není dle mých zkušeností úplně běžná a je vidět, že autorka v tomto ohledu výborně využila svoje částečné pracovní zapojení do projektu „Centrum pro multimodální interpretaci dat velkého rozsahu (CEMI)“, kde mohla nejmodernější metody zpracování obrazu konzultovat s předními tuzemskými odborníky. Na práci oceňuji především kreativitu při návrhu nových postupů pro vyhledávání informací s využitím různých modalit a pečlivost při jejich experimentálním ověřování. Chtěl bych také vyzdvihnout skutečnost, že se autorka zúčastnila mnoha oficiálních evaluačních kampaní v daném oboru, kde velmi často dosáhla na přední umístění. Práce je psaná dobrou a srozumitelnou angličtinou. Za výborný výsledek práce považují také výše zmíněnou implementaci nejlepších algoritmů ve formě ukázkového vyhledávacího systému.

Celkově mám k práci jen menší výhrady. Pasáže, které popisují klíčové algoritmy předložené práce, by si dle mého soudu zasloužily detailnější formální (matematický) popis. Také mi trochu vadila skutečnost, že zatímco výsledky kampaní, ve kterých autorka práce zvítězila, jsou uvedeny v přehledných grafech zobrazujících výsledky všech účastníků (a často impozantní nárůst autorčina systému), výsledky soutěží, kde byl vyvinutý systém méně úspěšný, jsou jen krátce zmíněny v textu (TRECVID 2015, SAVA Task 2015). Jakkoli je tento přístup lidsky pochopitelný, úplné srovnání by bylo dle mého názoru vhodnější.

Publikační činnost považuji za dostatečnou, přestože jde z velké části o články publikované na workshopech spojených s evaluačními kampaněmi. Na druhou stranu, účast v uvedených soutěžích znamená srovnání se světovou konkurencí v oboru a autorka práce v tomto srovnání obstála nadmíru úspěšně.

Závěrem tedy konstatuji, že předložená disertační práce prokazuje předpoklady autorky k samostatné tvořivé práci a jako takovou ji doporučuji k obhajobě.

Upřesňující dotaz k textu práce:

V kapitole 4.2.1 (Transcript quality) jsou v prvním odstavci uvedeny WER použitých systémů pro automatické rozpoznávání řeči, které se pohybují od 11% do 27.1%. V tabulce 4.2 jsou hodnoty WER pro stejné systémy v rozmezí 58.6 až 71.8%. Víím, že testy v tabulce 4.2 jsou provedeny na jiných datech, ale uvedený nepoměr je zarážející. Prosím o komentář při obhajobě.

V Plzni 21.12.2017

doc. Ing. Pavel Ircing, Ph.D.