

Vypracování týmového projektu

V tomto dokumentu shrnujeme výsledek vypracování našeho výzkumného projektu “**Rozšíření state-of-the-art vyhledávacího systému SOMHunter**”. Bod po bodu, jež byly naším cílem, rozebereme jestli vše proběhlo tak, jak bylo plánováno v zadání, nebo případně k jakým došlo změnám a proč.

Popis a cíle projektu

Proti plánovanému způsobu implementace došlo také k zásadní změně v celkovém návrhu nástroje jako takového. Úplně zmizela vrstva implementovaná v Node.js, která fungovala jako “adaptér” mezi voláními UI a C++ jádrem. Byla zde přítomna z historických důvodů, kdy ještě obsahovala i nějakou logiku.

Po začátku prací jsme si uvědomili, že už zastává skutečně POUZE funkci adaptéru a přidává neúměrné množství potřebné práce pro sestavení softwaru. Tuto vrstvu jsme proto odstranili a samotné vystavení API našeho hlavního C++ jádra jsme udělali přímo v projektu jádra. Nyní můžeme volitelně jádro spustit i s HTTP serverem, který zveřejňuje potřebné volání na definovaných endpointech.

A. Implementace textového dotazování pro libovolný podregion snímku videa, což umožní formulovat dotazy schopné zachytit i poziční vztahy entit ve videu. Současná verze systému umožňuje popis pouze snímku jako celku.

Tento bod byl zpracován v plném rozsahu beze změn. Nástroj nyní nabízí dva boxy, na kterých může uživatel vytvářet dotazy na obdélníkové podregiony jak pomocí dotazu, tak pomocí bitmapy.

B. Relokace dotazů, která dá uživateli možnost iterativně vylepšovat jednotlivé komponenty jeho kombinovaného dotazu na základě okamžité zpětné vazby pro dané modality dotazu. Hypotéza je, že přesnější jednotlivé složky povedou k větší celkové přesnosti dotazu.

Tento bod byl také zpracován beze změn. Uživatel má možnost vedle textového dotazu si otevřít malou SOM a na ní vybrat nejpodobnější obrázek. Poté místo textového dotazu bude použit vybraný obrázek jako dotaz. Ukázalo se, že toto rozšíření nebyla pro uživatele příliš používaná možnost hledání.

C. Reimplementace současného webového uživatelského rozhraní za pomocí lépe vyhovujícího frameworku. Tento bod souvisí s předešlými kroky, protože nové uživatelské rozhraní bude počítat i s rozšířeným dotazováním. Mimo to, výsledkem bude výrazně čitelnější a přehlednější kód front-endu, který umožní snadnější nástup do projektu pro budoucí členy týmu.

Bod splněn. Uživatelské rozhraní je nyní naimplementováno ve frameworku Ember.js.

- D. Implementace softwaru pro sběr “relevance feedback” dat**, který bude následně využit k rozsáhlé uživatelské studii. Tento software bude vytvořen z aktuální verze nástroje. Bude prezentovat zjednodušený postup vyhledávání a implementovat vhodný mechanismus tvorby logů.

Bod splněn. Tento projekt se nachází v samostatném repozitáři `somhunter-collector` a `somhunter-collector-tester`.

- E. Analýza logů a připravení simulace reálného uživatele** na základě nasbíraných dat, pomocí softwaru z předchozího bodu. Tyto simulace slouží k automatickému nastavení nástroje SOMHunter. Výstupem bude analýza výběru obrázků z displeje a algoritmus simulace tohoto výběru pomocí heuristik a strojového učení. Poznatky z této studie budou shrnuty do **publikace** a zaslány do impaktovaného žurnálu MTAP.

Bod splněn, žurnál je zaslán MTAP. Softwarový výstup tohoto bodu je obsažen v repozitářích `somhunter-simulator` a `somhunter-user-analysis`.

- F. Analýza logů ze soutěže LSC 2020** společně s porovnáním, jak by si systém vedl, kdyby již tehdy používal nový model CLIP. Výstupem budou data, grafy a další informativní vizualizace výkonu systému, na základě kterých budeme rozhodovat o budoucím vývoji systému SOMHunter. Na základě provedené analýzy bude sepsána **publikace** shrnující výsledky provedené analýzy a porovnání. Tato publikace bude zaslána také do žurnálu.

Bod splněn s tím, že nebyla analýza provedena na datech z LSC 2020, ale z dat z VBS 2021. V době začátků práce již byly k dispozici novější data v novém formátu, proto nebyl důvod stavět analýzu na těch starších. Také proběhla změna s publikací článku. Zde zmíněné speciální vydání bylo nakonec zrušeno. Proto výstupy této analýzy byly zahrnuty do přijatého článku na VBS 2022 (Video Search with Context-aware Ranker and Relevance Feedback).

- G. Integrace text-to-video modelu CLIP**, který dle našich předběžných měření vykazuje vyšší přesnost na našem test datasetu než momentálně využívaný model W2VV++. Tato zásadní změna by měla přinést výrazně vyšší výkonnost systému, zároveň ale vyžaduje zásahy do celkové architektury systému.

Bod splněn. Evaluace dotazu pomocí CLIP modelu byla implementována mimo jádro systému SOMHunter z důvodu využívání modelu dalšími softwarovými nástroji, tento projekt se nachází v repozitáři `ranking-server`. Díky tomuto designovému kroku je možné využívat tuto službu i z dalších softwarových nástrojů, např. nástroj CVHunter při soutěži VBS 2021.

- H.** Posledním bodem projektu je **zobecnění textového dotazování** na dotazování buď pomocí textu nebo pomocí obrázku dodaného z libovolného zdroje (typicky bitmapa ze

schránky systému). Přičemž uživatel může typ jednotlivě umístěných dotazů (z bodu A) libovolně kombinovat. To dovolí uživatelům popsat i koncepty, na které se těžko hledají slova nebo koncepty, které implementovaný text-to-image model neumí dobře umístit.

Bod splněn beze změn. Nyní SOMHunter podporuje libovolné kombinace textových a bitmapových dotazů na podregiony snímku.

Výstupní publikace

Evaluating a Bayesian-like Relevance Feedback Model with Text-to-Image Search Initialization

Multimedia Tools and Applications journal

Ladislav Peška, Marta Vomlelová, Patrik Veselý, Vít Škrhák and Jakub Lokoč.

Video Search with Context-aware Ranker and Relevance Feedback

International Conference on Multimedia Modeling

Jakub Lokoč, František Mejzlík, Tomáš Souček, Patrik Dokoupil, and Ladislav Peška.

A navíc také participace na žurnálu z VBS 2021:

Interactive Video Retrieval Evaluation at a Distance

International Journal of Multimedia Information Retrieval

Silvan Heller, Viktor Gsteiger, Werner Bailer, Cathal Gurrin, Björn Þór Jónsson, Jakub Lokoč, Andreas Leibetseder, František Mejzlík, Ladislav Peška, Luca Rossetto, Konstantin Schall, Klaus Schoeffmann, HeikoSchuldt, Florian Spiess, Ly-Duyen Tran, Lucia Vadicamo, Patrik Veselý, Stefanos Vrochidis, Jiaxin Wu.