

Specifikace softwarového projektu

Iterativní reformulace textového dotazu
do multimediální databáze
s využitím sítě CLIP

Cílem projektu je otestovat a analyzovat vliv automaticky získaných textových popisků mezivýsledků pro následnou možnost reformulace textového dotazu. Popisovaný projekt bude realizován jako webová aplikace, ve které jednotlivé komponenty předzpracování dat budou implementovány v Pythonu za využití nejnovějšího modelu neuronové sítě CLIP, zatímco webové rozhraní zobrazující výsledky bude vyvinuto v PHP a JavaScriptu.

0.1

Zuzana Vopálková

8. srpna 2022

Obsah

1	Základní informace	1
1.1	Popis a zaměření softwarového díla	1
2	Stručný popis softwarového díla	1
2.1	Hlavní funkce	1
2.2	Prostředí aplikace	1
2.2.1	Hledání	1
2.3	Omezení projektu	1
2.4	Dokumentace projektu	1
3	Detailní popis funkcionality	2
3.1	Části projektu	2
3.1.1	Searcher	2
3.1.2	Updater	2
3.1.3	Display	2
3.2	Dataset	3
3.3	Kategorie snímků	3
3.4	Použité technologie	3
4	Experiment	3
4.1	Sběr dat	3
4.2	Průběh experimentu	3
5	Ostatní (mimofunkční) požadavky	3
6	Time-line & Milestones	4

1 Základní informace

1.1 Popis a zaměření softwarového díla

Projekt se zabývá vyhledáváním v multimediální (primárně obrázkové) databázi, kde je možné interaktivně reformulovat textový dotaz. Tato strategie je vhodná v případech, kdy uživatel hledá také vhodná slova popisující vyhledávaný snímek/scénu. Cílem projektu je otestovat a analyzovat vliv automaticky získaných textových popisků mezivýsledků pro následnou možnost reformulace textového dotazu.

Popisovaný projekt bude realizován jako webová aplikace, ve které bude možné interaktivní vyhledávání s využitím výše popsaného postupu a s použitím nejnovějšího modelu neuronové sítě CLIP. Jednotlivé komponenty předzpracování dat budou implementovány v Pythonu, zatímco webové rozhraní zobrazující výsledky hledání a anotace bude vyvinuto v JavaScriptu. Projekt případně využije také existující komponenty z předchozích projektů vedoucího. Kromě softwarové části projekt bude obsahovat ještě analýzu samotného metody opakované reformulace dotazu.

2 Stručný popis softwarového díla

2.1 Hlavní funkce

Hlavní funkcí programu je hledání v multimediální databázi pomocí textového dotazu s následným zobrazením výsledků zařazených do class, což by uživateli mělo pomoci s výběrem slov pro reformulaci dotazu. Kromě textového vyhledávání program umožňuje i podobnostní vyhledávání na základě vybraných obrázků z aktuálně zobrazených.

2.2 Prostředí aplikace

Uživatelé používaná část softwaru bude vyvinuta jako webová aplikace s připojením na server, který bude zpracovávat dotazy do databáze. Webové rozhraní bude sloužit pouze na zobrazování výsledků a na interakci s uživatelem. Webové rozhraní bude vytvářeno pomocí JavaScriptu, přičemž události a zobrazování budou obstarávat pomocí knihovny React a dotazy na server a jejich zpracování bude zajištěno pomocí knihovny Django.

2.2.1 Hledání

Samotné uživatelské hledání je převážně založeno na textových dotazech. Po odeslání textového dotazu je zobrazené omezené množství nejrelevantnějších snímků, u kterých si uživatel může zobrazit kategorie, do kterých byli snímky zařazeny a pomocí toho přereformulovat svůj dotaz. Další možností uživatele je zobrazené snímky využít na podobnostní hledání nebo na zobrazení snímků v kontextu databáze.

2.3 Omezení projektu

Hlavní omezení softwaru bude server, na kterém poběží serverová část softwaru, která má za úkol zpracovávat dotazy. Serverová část bude umožňovat více sessions, aby byl možný přístup více uživatelům naráz.

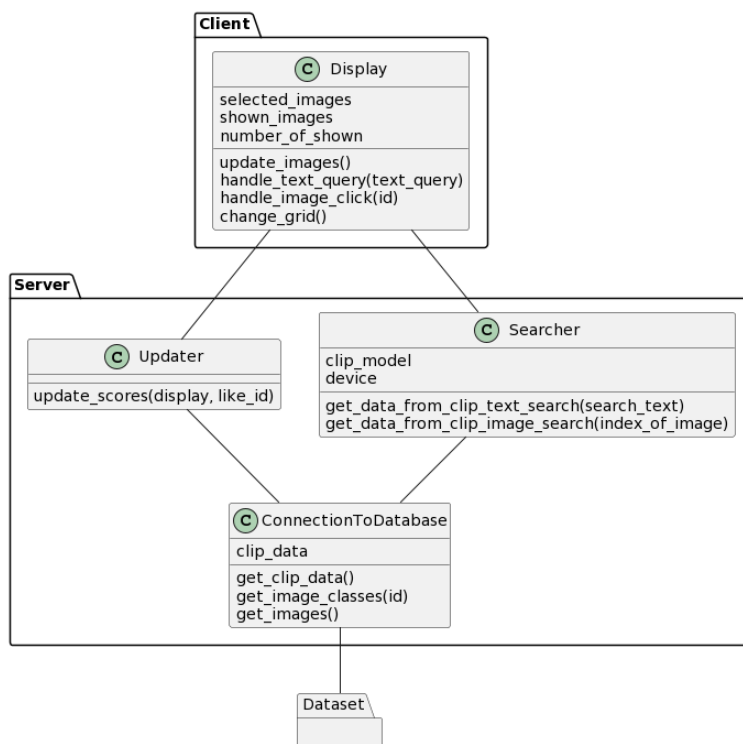
2.4 Dokumentace projektu

Dokumentace tohoto projektu se bude skládat z této specifikace a dále z uživatelské dokumentace (+ příručky) a softwarové dokumentace, která se zabývá samotným kódem a jeho specifiky.

3 Detailní popis funkcionality

3.1 Části projektu

Projekt má dvě hlavní části, kterými jsou server, který bude zpracovávat všechny dotazy do databáze s daty získaných z neuronové sítě Clip, a webové rozhraní, které bude interagovat s uživatelem.



Obrázek 1: Class diagram

3.1.1 Searcher

Tato část programu bude pracovat s daty získaných pomocí neuronové sítě Clip. Vyhledávat bude na základě textového dotazu nebo vybraného snímku, přičemž při každém dotazu vrací snímky zařazené do classes, což jsou běžně užívané anglická podstatná jména (získaná z dostupné databáze).

Data z neuronové sítě Clip, fotky i jejich zařazení do classes je předzpracováno, takže při dotazu jsou již tato data pouze použita.

3.1.2 Updater

Úkolem této části je aktualizace dat na základě označených a zobrazených snímků. I přesto, že je modul nazvaný updater, nebude provádět doslova updates, ale místo toho provádět similarity searcher na základě označených obrázků (z tohoto důvodu budou povoleny označit pouze 2|3 snímky).

3.1.3 Display

V zásadě se jedná o jedinou část, která interaguje s uživatelem. Nejdůležitější částí je tedy zpracování událostí jako je odeslání textového dotazu a označování snímků, které je spojeno se zobrazením

podrobností spojených se snímkem (zařazení do kategorií).

3.2 Dataset

Databáze snímků, s kterou se bude program pracovat bude složena z naparsovaných videí a bude obsahovat 11 870 snímků.

3.3 Kategorie snímků

Každý snímek z databáze bude zařazen do minimálně jedné kategorie, avšak samotné množství kategorií nebude omezeno počtem, ale bude postaveno na blízkosti (kosinová podobnost) dat snímku získaných z neuronové sítě Clipu k samotným slovům.

3.4 Použité technologie

- * JavaScript (knihovny jQuery|React)
- * Python (knihovny Django, clip, numpy, torchvision, ...)
- * neuronová síť Clip

4 Experiment

4.1 Sběr dat

Pro vyhodnocení efektivity iterativní reformulace dotazů budou nasbírána data (ideálně i od do techniky ne úplně zasvěcených uživatelů) ideálně v rozmezí od 100 uživatelských dotazů (vyhledávání konkrétního snímku).

4.2 Průběh experimentu

Po dokončení vývoje programu bude uskutečněn experiment, v kterém uživatelé budou v omezeném čase hledat předem určený snímek. Pokud snímek bude nalezen bude dané "kolo" pro uživatele úspěšně ukončeno, v případě, že uživatel nebude schopný snímek nalézt bude mít možnost se v daném "kole" vzdát/ukončit ho.

Po celou dobu budou sbírány data o rychlosti a postupu nalezení snímku uživateli.

5 Ostatní (mimofunkční) požadavky

Aplikace nebude v konečné verzi sloužit veřejnému používání (kvůli zpracovávání požadavků na serveru - omezená rychlost zpracovávání dotazů).

6 Time-line & Milestones

Deadline	Popis
do 16. 5.	Dokončená první verze podrobné specifikace
do 6. 6.	Zpracování modulu Searcher a zpracování databáze fotek (videí)
do 20. 6.	Zpracování modulů Updater -> dokončení serverové části
do 4. 7.	První verze API (celá funkčnost)
do 11. 7.	Dokončení API části -> první funkční verze softwaru
do 25. 7.	Dopsání dokumentace (uživatelská dokumentace)
do 15. 8.	Doladění detailů (případná úprava designu, testování) + připravení materiálů na testování (výběr videí)
do 12. 9.	Uskutečnění pokusu a zpracování výsledků
do konce září	Odevzdání kompletní ročníkové práce