

بسم الله الرحمن الرحيم

گزارش اولیهء پروژه "استخراج کلمات

کلیدی از منابع صوتی و ویدئویی

با استفاده از هوش مصنوعی"

فاطمه وحیدیونسی

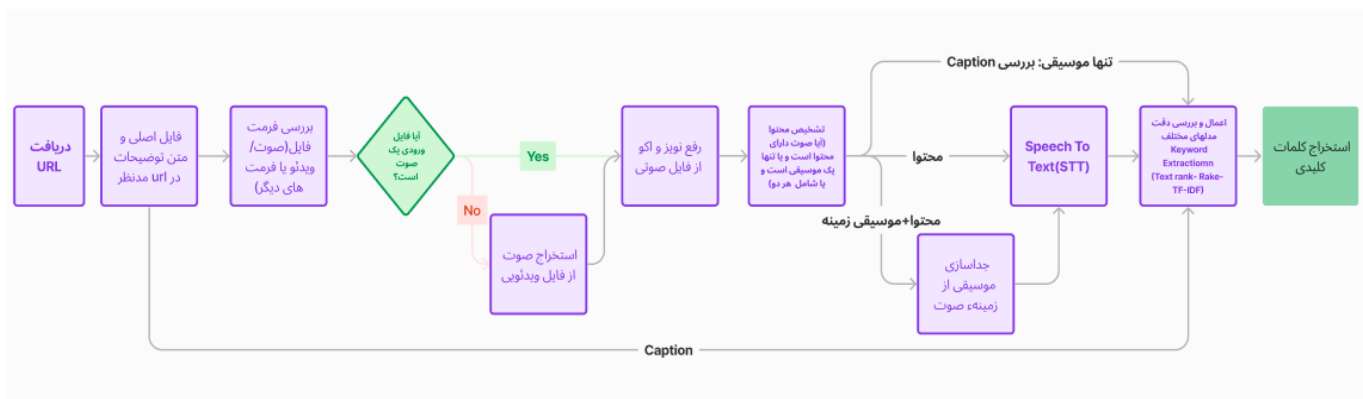


مرکز نوآوری علوم و فناوری های شناختی

دانشگاه علم و صنعت ایران

تاریخ تدوین: 1402/10/29

هدف از این پروژه ساخت یک فریمورک از مازول های مختلف جهت رسیدن به مقصود "استخراج کلمات کلیدی با دقت بالا از فایل های صوتی و ویدئویی" از منابعی مانند اینستاگرام است. فلوچارت قرار داده شده در زیر، مراحل انجام پروژه را نشان می دهد:



نمودار 1- فلوچارت مراحل انجام پروژه

جهت درک بهتر، مراحل پروژه برای نمونه استخراج کلمات کلیدی از "پست های اینستاگرام" در زیر تشریح می شود:

- 1- فایل اصلی پست شده به همراه توضیحات (Caption) موجود در پست دریافت می شود. در این پروژه از هر دو سورس (متن توضیحات کپشن و متن ناشی از تبدیل صوت (STT) جهت استخراج دقیق تر کلمات کلیدی استفاده خواهد شد).
- 2- فایل ورودی از جهت فرمت، بررسی شده، در صورتی که فایل ورودی ویدئویی داشته باشد، مرحله استخراج صوت از ویدئو اجرا می شود.
- 3- نویز، اکو و صداهای زمینه که کیفیت مراحل بعدی از جمله دقت STT به آنها بسیار وابسته است، کاهش داده می شود.
- 4- تشخیص "محتوای" فایل صوتی: در صورتیکه فایل تنها یک موسیقی است و فاقد محتوای صوتی، وارد مرحله استخراج کلمات کلیدی از سورس دیگر، متن caption، می شویم. در صورتیکه صدای یک گوینده به همراه موسیقی بی کلام در بکگراند است، عملیات مرحله 6 ام آغاز می شود.
- 5- در صورتیکه در پس زمینه صدا، موسیقی وجود دارد، حذف موسیقی از بکگراند و ذخیره محتوای اصلی جهت آماده سازی برای مرحله STT اجرا می شود.
- 6- استفاده از متد های مختلف STT جهت تبدیل صدای نهایی به متن، ارزیابی دقت آنها و انتخاب روش بهینه
- 7- استفاده از متد های مختلف Keyword Extraction برای متن Caption و متن حاصل از صوت

8- یافتن ضریب بهینه‌ء بتا در استخراج کلیدواژه از دو سورس "متن کپشن" و "متن حاصل از صوت"

در زیر، جدول وضعیت اجرای مراحل فوق، تا کنون، قرار داده شده است:

وضعیت	مرحله
تا تاریخ 28 دی، UI آپلود فایل انجام شده. حالت "دانلود از url" بعدها به صورت مسئله پروژه اضافه گردیده که در دست انجام است.	1- دریافت فایل اصلی پست شده به همراه توضیحات (Caption) موجود در پست
انجام شده	2- بررسی فرمت فایل ورودی، در صورتی که فایل ورودی ویدئویی داشته باشد، استخراج صوت
در دست انجام	3- کاهش نویز، اکو و صداهای زمینه که کیفیت مراحل بعدی از جمله دقت STT به آنها بسیار وابسته است
در حال بررسی Literature	4- تشخیص "محتوای" فایل صوتی (موسیقی باکلام است یا موسیقی بی کلام در زمینه به همراه محتوای اصلی توسط یک گوینده است).
در حال انجام	5- در صورتیکه در پس زمینه صدا، موسیقی وجود دارد، حذف موسیقی از بکگراند و ذخیره محتوای اصلی جهت STT
با یک مدل تست انجام گرفت در حال ارزیابی متدهای مختلف	6- استفاده از متدهای مختلف STT جهت تبدیل صدای نهایی به متن و ارزیابی دقت آنها
تا اینجا استخراج کلمات کلیدی با روش Text-Rank انجام شده، با توجه به پیاده سازی سابق ماژول استخراج کلمات کلیدی، از این جا به بعد، بر روی مراحل 1 تا 6 و هشتم تمرکز خواهد شد و برای این مرحله از Api های ارزیابی شده موجود استفاده خواهد شد.	7- استفاده از متدهای مختلف Keyword Extraction برای متن Caption و متن حاصل از صوت
در دست انجام	8- یافتن ضریب بهینه‌ء بتا در استخراج کلیدواژه از دو سورس "متن کپشن" و "متن حاصل از صوت"

همانگونه در جدول فوق توضیح داده شده است، تا تاریخ 28 ام دی ماه (سه روز پس از استارت پروژه)، یک نمونه اولیه شامل یک UI دریافت فایل، تشخیص فرمت فایل ورودی، تبدیل ویدئو به صوت، اجرای یک مدل STT (صوت به متن) و استخراج کلمات کلیدی از متن حاصل، با روش Text-Rank انجام گرفته است.

با توجه به هماهنگی ها و صحبت های انجام گرفته، در ادامهء پروژه تمرکز بر روی مراحل 1 تا 6 و مرحلهء هشتم خواهد بود و مرحله 7 با استفاده از Api های موجود و ارزیابی شده انجام خواهد گرفت که نیازمند پیگیری است.