

بسمه تعالی

وزارت ارتباطات و فناوری اطلاعات
پژوهشگاه ارتباطات و فناوری اطلاعات
(مرکز تحقیقات مخابرات ایران)



درخواست ارائه پیشنهاد (RFP) پروژه جویشر مستندات گفتاری

زیرمجموعه: طرح جویشر

نسخه: ۱,۰

آذر ماه ۹۵

سطح دسترسی: عمومی

فهرست مطالب

۱	مقدمه	۲
۲	تعریف پروژه	۳
۳	هدف پروژه	۴
۴	قلمرو اجرای پروژه	۵
۵	مراحل اجرا و شرح خدمات پروژه	۶
۶	روش آزمون کیفیت و صحت خروجی	۹

۱ مقدمه

در عصر وب و فناوری اطلاعات، انواع مختلفی از مستندات هرروزه تولید شده و از طریق وب منتشر می‌شوند. بسیاری از این مستندات، مستندات متنی هستند که اطلاعات را در قالب دنباله‌ای از کاراکترها منتقل می‌سازند ولی مستندات از انواع دیگر هم وجود دارند که حجم زیادی از محتوای وب را اشغال کرده‌اند و شاید تقاضا برای آنها بیشتر از مستندات متنی باشد. این مستندات شامل تصویر، صوت، ویدئو و ... می‌باشند. در این میان مستندات صوتی به دلیل اینکه ساده‌ترین راه ارتباط میان انسان‌ها از طریق صوت انجام می‌گیرد، از اهمیت ویژه‌ای برخوردار است. مستندات صوتی را می‌توان به دو دسته کلی مستندات گفتاری و مستندات غیرگفتاری تقسیم کرد. مستندات گفتاری را نیز می‌توان به دو دسته آهنگین و غیرآهنگین تقسیم‌بندی کرد. مستندات گفتاری غیرآهنگین شامل مکالمات و دیالوگ‌ها، سخنرانی‌ها، مصاحبه‌ها، جلسات، گویندگی^۱ در رسانه‌ها و ... می‌باشد. همچنین موسیقی با کلام، آواز و سرود نیز جزء مستندات گفتاری آهنگین به شمار می‌رود. مستندات صوتی غیرگفتاری شامل موسیقی بدون کلام، اصوات موجود در طبیعت، صداهای مصنوعی مانند افکت‌های صوتی، صدای ابزارها و ... می‌باشند.

با توجه به طیف گسترده مستندات صوتی (به خصوص مستندات گفتاری) و اهمیت آنها، نیاز به جستجو و بازیابی مؤثر این‌گونه مستندات احساس می‌شود. در بسیاری اوقات ما ممکن است در وب یا در یک آرشیو صوتی به دنبال مستند یا مستندات گفتاری باشیم که در آن کلمات یا عبارات خاصی به کار رفته باشد و یا در مورد موضوع موردنظر ما صحبت کرده باشد. در این حالت جستجو در متن وبسایت‌ها لزوماً مشکل را حل نمی‌کند، زیرا متن وبسایت‌ها همیشه دربردارنده تمام محتویات مستندات گفتاری نیست. این موضوع، به خصوص زمانی اهمیت پیدا می‌کند که بخواهیم یک آرشیو مستندات ویدئویی را بر اساس محتوای گفتاری آنها جستجو کنیم. بسیاری از ویدئوها (کلیپ‌ها، فیلم‌ها، برنامه‌های تلویزیونی و ...) دارای یک بخش گفتاری یا صوتی می‌باشند و می‌توان آنها را بر اساس محتوای صوتی آنها جستجو و بازیابی کرد.

در کشور ما چند موتور جستجوی بومی تاکنون توسعه داده شده و هم‌اکنون در طرح جویشگر نیز فعالیت‌ها برای ارتقاء کیفیت و قدرت این جویشگرها در حال انجام است. علاوه‌براین، سرویس‌های مختلفی نیز در این موتورهای جستجو در حال توسعه می‌باشند. با توجه به اینکه تاکنون امکان جستجو و بازیابی مستندات گفتاری در این موتورهای جستجو فراهم نشده است، پروژه فوق با هدف توسعه و ارتقای موتورهای جستجوی موجود به قابلیت جستجو در مستندات گفتاری ارائه می‌گردد.

^۱ narration

۲ تعریف پروژه

در این پروژه جویسگر مستندات گفتاری مدنظر است که امکان جستجو و بازیابی مستندات گفتاری^۲ غیرآهنگین در یک آرشیو از این مستندات (مانند وب) را داشته باشد. همان‌گونه که ذکر شد مستندات گفتاری به دو دسته گفتار آهنگین (مانند موسیقی، آواز و ...) و گفتار غیرآهنگین (سخنرانی، گویندگی اخبار و ...) تقسیم می‌شوند. در این پروژه مستندات گفتاری غیرآهنگین مدنظر می‌باشد. در ادامه این RFP، منظور از مستندات گفتاری، مستندات گفتاری غیرآهنگین می‌باشد که مستندات ویدئویی دارای گفتار را نیز شامل می‌شود.

فرض می‌کنیم یک مجموعه^۳ یا آرشیو از مستندات گفتاری داریم. این آرشیو در حالت کلی می‌تواند هر مکان‌های مختلفی از نظر فیزیکی قرار گرفته باشد (مثلاً آرشیو صوتی صداوسیما یا آرشیو کلاس‌های درس ارائه‌شده در یک دانشگاه مجازی) ولی یکی از بزرگترین مجموعه‌هایی که با آن سروکار داریم آرشیو وب است. در این پروژه با توجه به اینکه هدف، به‌کارگیری خروجی پروژه در جویسگرهای بومی می‌باشد، آرشیو یا مجموعه مستندات موردنظر مجموعه مستندات گفتاری فارسی موجود در وب است.

با فرض وب به عنوان آرشیو مستندات گفتاری، در این پروژه کاربر باید قادر باشد مستندات گفتاری مرتبط با پرس‌وجوی خود را در وب بیابد. به عنوان مثال کاربر ممکن است به دنبال جلسه‌ای از یک دوره^۴ درسی آنلاین^۴ بگردد که در آن از مثلاً ترجمه ماشینی صحبت شده باشد و یا به دنبال و فیلم یا برنامه^۴ رادیو-تلویزیونی خاصی باشد که نام آن را نمی‌داند ولی محتوای آن را در ذهن دارد. در این حالت، کاربر پرس‌وجوی خود را به صورت متن و در قالب مجموعه‌ای از کلمات یا عبارات بیان می‌کند (query by keyword) و جویسگر باید قادر باشد در میان اسناد گفتاری نمایه‌شده، شبیه‌ترین و مرتبط‌ترین اسناد را به پرس‌وجوی کاربر پیدا کند.

در این پروژه از آنجاکه هدف پیدا کردن مستندات گفتاری مرتبط با پرس‌وجو بوده و یافتن دقیق عبارت جستجو و محل آن در مستند گفتاری مدنظر نیست، پرس‌وجو به صورت متنی (query by keyword) مدنظر بوده و پرس‌وجوی گفتاری (query by example) مدنظر نمی‌باشد.

^۲ Spoken Document Retrieval (SDR)

^۳ collection

^۴ online course

۳ هدف پروژه

هدف از این پروژه توسعه یک سامانه جویسگر بومی مستندات گفتاری است که به وسیله آن کاربران جویسگرهای بومی بتوانند مستندات گفتاری فارسی مورد نیاز خود را با جستجوی عبارات کلیدی آن در وب پیدا کنند. به عبارت دیگر زمانی که جستجو در محتویات متنی به نتیجه قابل قبولی منجر نمی شود، این سامانه باید بتواند با جستجو در محتویات صوتی، مستند یا مستندات مورد درخواست کاربر را بازیابی نماید.

اهداف مورد نظر در این پروژه عبارتند از:

- توسعه یک سامانه جویسگر بومی مستندات گفتاری فارسی
- پاسخگویی به نیازهای کاربران فارسی زبان برای یافتن مستندات گفتاری
- تسهیل و تسریع دسترسی کاربران بومی به مستندات گفتاری فارسی
- گسترش محتوای گفتاری فارسی در وب
- حفظ و صیانت و ترویج زبان فارسی

۴ قلمرو اجرای پروژه

مشتری پروژه: طرح جویسگر

قلمرو منطقی: حوزه عمومی

قلمرو فیزیکی: مرکز داده توافق شده با طرح جویسگر

قلمرو فناوری: بازیابی گفتار/اطلاعات، پردازش گفتار، نمایه‌سازی گفتار، پردازش خط و زبان فارسی، هوش

مصنوعی

۵ مراحل اجرا و شرح خدمات پروژه

پیشنهاددهنده می‌بایست مراحل و فازبندی انجام کار را با جزئیات کامل ارائه کند. توجه به موارد زیر در پیشنهاد ارسالی ضروری است:

موارد کلی

- در صورتی که پیشنهاددهنده تمام یا بخشی از سرویس جویشر مستندات گفتاری را به صورت محصول در اختیار دارد، باید وضعیت محصول به طور دقیق ذکر شود. بیان جزئیات محصول باید شامل، مشخصات کلیدی، دقت، کیفیت، تعداد کاربران، سرعت پاسخگویی و معماری سخت‌افزار و نرم‌افزار مورد استفاده باشد.
- در پیشنهاد ارسالی می‌بایست نحوه بهینه‌سازی و ارتقاء کیفیت سامانه در حین اجرای پروژه ذکر گردد.
- پیشنهاد ارسالی محدود به موارد مشخص شده در این مستند نمی‌باشد و در صورت داشتن ایده جدید و یا ویژگی کاربردی در صورت ارائه ارزیابی و لحاظ خواهد شد.
- نیازمندیهای سامانه در تعامل با جویشرهای بومی باید در پیشنهاد ارسالی مشخص شود.
- در پیشنهاد ارسالی می‌بایست حتی‌الامکان از منابع موجود کشور در بخش سخت‌افزار و نرم‌افزار استفاده شود.
- در پیشنهاد ارسالی می‌بایست برنامه توسعه‌ی بازار و طرح تجاری نیز ارسال شود. در طرح تجاری نحوه همکاری با موتورهای جستجو، تسهیم درآمد و جذب کاربران ارائه گردد.

ویژگی‌های فنی

در این پروژه، یک سامانه جستجو و بازیابی مستندات گفتاری برای زبان فارسی به صورت query by keyword توسعه می‌یابد. آزانجاکه معمولاً در این گونه سیستم‌ها دو مرحله کلی نمایه‌سازی و بازیابی گفتار صورت می‌گیرد، پیشنهاددهنده باید برنامه و استراتژی خود را در این راستا در زمینه‌های زیر مشخص نماید:

- معماری سامانه بازیابی مستندات گفتاری (SDR)
- نحوه تبدیل فایل‌های گفتاری به متن
- نحوه نمایه‌سازی گفتار
- روش بازیابی اطلاعات در نمایه‌های گفتاری
- نحوه جداسازی گفتار از اصوات غیرگفتاری (موسیقی، نویز و ...)

در ادامه، مشخصه‌ها و الزامات فنی مدنظر برای سامانه SDR مشخص شده است.

- قابلیت تبدیل مؤثر انواع فایل‌های گفتاری به متن

- متوسط دقت مازول بازشناسی گفتار ۶۵٪

- قابلیت کار با انواع فرمت‌های صوتی
- قابلیت جداسازی مؤثر بخش‌های گفتاری و غیرگفتاری در فایل‌های صوتی
- قابلیت به‌کارگیری روش‌های متداول نمایه‌سازی مستندات گفتاری
- تبدیل و نمایه‌سازی فایل‌های گفتاری خزش‌شده به صورت منظم و دوره‌ای
- به‌کارگیری متدهای رایج بازیابی اطلاعات
- بازیابی مستندات گفتاری با معیار NDCG معادل ۰.۶۰ و معیار P@۱۰ معادل ۰.۸۰
- متوسط زمان بازیابی حداکثر ۱ ثانیه

الزامات نگهداری

- ارائه خدمات پشتیبانی رایگان به مدت دو سال
- نگهداری از سامانه بصورت رایگان به مدت دو سال

مراحل اجرای پروژه

- اجرای پروژه به مدت هجده ماه در سه مرحله، فاز اول (شش ماه)، فاز دوم (شش ماه) و فاز سوم (شش ماه) انجام می‌گردد.
- فاز اول:
 - توسعه و بهینه‌سازی موتور بازشناسی گفتار برای تبدیل مؤثر انواع فایل‌های صوتی به متن.
 - طراحی ماژول جداسازی بخش‌های گفتاری از غیرگفتار در این فاز انجام می‌گردد.
 - روال تست و ارزیابی با همکاری تیم نظارت و شرکت‌های مجری در این فاز انجام می‌گیرد.
- فاز دوم
 - نمایه‌سازی مستندات حاصل از بازشناسی گفتار و طراحی الگوریتم بازیابی اطلاعات مناسب برای آن
 - ارائه سرویس اولیه بازیابی مستندات گفتاری
 - ارزیابی بر اساس روال تست طراحی‌شده در فاز اول انجام می‌گردد و مجریانی که در ارزیابی، موفق به برآورده نمودن نیازمندی‌های روال تست و ارزیابی شده باشند، امکان ادامه فعالیت در فاز سوم را خواهند داشت.
- فاز سوم
 - بهینه‌سازی و ارتقاء کیفی سرویس بازیابی مستندات گفتاری جهت برآورده نمودن کیفیت مطرح شده در RFP
 - همکاری با یکی از موتورهای جستجوی بومی

- ۷۰ درصد ارزیابی فنی فاز سوم مجریان منوط به ارائه سرویس بازیابی مستندات گفتاری در یکی از موتورهای جستجوی بومی (دارای قرار داد با طرح جویسگر) می باشد.
- تعاملات همکاری سرویس بازیابی مستندات گفتاری با موتورهای جستجوی بومی باید توسط مجری به صورت مستقل انجام گردد.
- مجری موظف به ارائه طرح تجاری در فاز سوم در جهت همکاری با موتور جستجو در دو سال پس از اتمام قرار داد می باشد.

خروجیهای پروژه

- ارائه سرویس بازیابی مستندات گفتاری با دقت و کیفیت ذکر شده در بند ویژگی های فنی، در یکی از موتورهای جستجوی بومی به مدت دو سال پس از تأیید و اتمام قرار داد
- ارائه ابزارهای تهیه شده در حین انجام پروژه شامل موارد زیر:
 - ابزار نمایه ساز مستندات گفتاری
 - ابزار بازیابی اطلاعات گفتاری
 - ...
- ارائه دادگان های تهیه شده در حین انجام پروژه شامل موارد زیر:
 - دادگان متنی پرس و جوهای کاربران
 - آرشیو فایل های صوتی جمع آوری شده
 - ...

۶ روش آزمون کیفیت و صحت خروجی

معیارهای ارزیابی نیازهای کاربردی

برای سنجش کارایی سامانه‌های بازیابی اطلاعات معمولاً از معیاری مانند دقت (precision)، فراخوانی (recall) و میانگین توافقی این دو معیار یعنی معیار F استفاده می‌شود. در مواردی مانند پروژه حاضر که جستجو و بازیابی در وب انجام می‌گیرد و امکان اندازه‌گیری معیار F وجود ندارد از معیارهایی مانند $P@k$ (precision at k) و یا معیار NDCG^۵ استفاده می‌شود. در پروژه حاضر معیارهای NDCG و $P@10$ (به ازای $k=10$) مدنظر است. همچنین معیار نرخ خطای کلمه (WER^۶) نیز برای ارزیابی ماژول بازشناسی گفتار به کار می‌رود.

معیارهای ارزیابی نیازهای غیرکارکردی

- دسترس پذیری
- کارایی
 - زمان پاسخ گویی
 - توانایی در پاسخ گویی به درخواست‌های همزمان
 - زمان پاسخ گویی به درخواست‌های همزمان
- آمار بازدید
 - متوسط تعداد کل کاربران در روز
 - متوسط تعداد کاربران یکتا در روز
 - متوسط تعداد درخواست‌ها در روز
- تحمل پذیری در برابر خطا
- امنیت

^۵ Normalized Discounted Cumulative Gain

^۶ Word Error Rate