

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه اصفهان

دانشکده مهندسی کامپیوتر

گروه مهندسی نرم افزار

پایان نامه کارشناسی

رشته مهندسی کامپیوتر گرایش نرم افزار

عنوان پایان نامه

طراحی و پیاده سازی دستیار صوتی لینوکس

استاد راهنما:

دکتر بهمن زمانی

پژوهشگر:

محمد قربانپور

شهریور ۱۴۰۱



دانشگاه اصفهان

دانشکده مهندسی کامپیوتر

گروه مهندسی نرم افزار

پروژه کارشناسی رشته‌ی مهندسی کامپیوتر گرایش نرم افزار

آقای محمد قربانپور

تحت عنوان

طراحی و پیاده سازی دستیار صوتی لینوکس

در تاریخ / / ۱۳ توسط هیأت داوران زیر بررسی و با نمره به تصویب نهایی رسید.

۱- استاد راهنمای پروژه:

دکتر

امضا

۲- استاد داور :

دکتر

امضا

امضای مدیر گروه

تشکر و قدردانی

سپاس خدای بزرگ را که مرا یاری رساند تا بتوانم مقطع تحصیلی کارشناسی را به پایان رسانده و گامی در راستای اعتلای علم بردارم.

از استاد راهنمای گران‌قدرم جناب آقای دکتر بهمن زمانی که برای نوشتن این پایان‌نامه کمک و راهنمایی‌های ارزنده‌ای کردند، کمال تشکر را دارم.

تشکر از تمامی اساتیدی که توفیق دانشجویی در محضرشان را داشتم.

از دوستان عزیزم که همراهان همیشگی من بوده‌اند و اوقات خوشی را در کنار هم سپری کرده‌ایم، تقدیر و تشکر دارم.

و در پایان از تمامی عزیزانی که در طول انجام این پروژه مرا یاری کرده‌اند کمال تشکر و قدردانی را ابراز می‌نمایم.

چکیده:

در دنیای امروزی زمان اهمیت بسیاری دارد. با انتشار یافتن نرم‌افزارهای جدید برای ارائه خدمات به کاربران، پیچیدگی در استفاده از این قابلیت‌ها نیز افزایش یافته است. بنابراین دنیا به نرم‌افزارهای دستیار صوتی برای کاهش پیچیدگی‌ها و صرفه‌جویی در زمان نیاز پیدا کرده است.

در این پروژه تلاش شده است تا با استفاده از زبان‌های برنامه‌نویسی پایتون، AIML و مدل‌های تبدیل گفتار به متن و برعکس، یک نرم‌افزار دستیار صوتی برای سیستم‌عامل لینوکس پیاده‌سازی کنیم. ابتدا با بررسی دیگر نرم‌افزارهای دستیار صوتی و همچنین نیازهای مورد نظر، نیازمندی‌های کاربردی مورد نیاز برای نرم‌افزار را استخراج می‌کنیم. سپس با توجه به این نیازمندی‌ها، ویژگی‌های نرم‌افزار را گام به گام پیاده‌سازی می‌کنیم. سپس الگوهای زبانی مطابق با قابلیت‌های دستیار صوتی را طراحی می‌کنیم و در نهایت یک نرم‌افزار قابل اجرا تولید می‌کنیم.

فهرست مطالب

صفحه	عنوان
۶.....	فصل اول مقدمه.....
۶.....	۱-۱- هدف پروژه.....
۶.....	۲-۱- کاربردهای پروژه.....
۷.....	۳-۱- ساختار پایان نامه.....
۸.....	فصل دوم معرفی نرم افزارهای مشابه.....
۸.....	۱-۲- مقدمه.....
۸.....	۲-۲- Jarvis.....
۸.....	۳-۲- Braina.....
۱۰.....	۴-۲- DataBot.....
۱۰.....	۵-۲- Cortana.....
۱۰.....	۶-۲- Computer Talker.....
۱۱.....	۷-۲- Mycroft.....
۱۱.....	۸-۲- جمع بندی.....
۱۲.....	فصل سوم پیش زمینه.....
۱۲.....	۱-۳- مقدمه.....
۱۲.....	۲-۳- تکنولوژی های مورد استفاده.....
۱۲.....	۱-۲-۳- زبان برنامه نویسی پایتون.....
۱۲.....	۲-۲-۳- AIML.....
۱۳.....	۳-۳- مدل های تبدیل گفتار به متن فارسی.....
۱۴.....	۱-۳-۳- Persian STT.....
۱۴.....	۲-۳-۳- Wav2Vec2-Large-XLSR-53-Persian.....
۱۵.....	۳-۳-۳- Vosk-Model-Fa.....
۱۶.....	۴-۳- جمع بندی.....
۱۷.....	فصل چهارم شرح پروژه.....
۱۷.....	۱-۴- مقدمه.....
۱۷.....	۲-۴- فهرست نیازمندی ها.....

فهرست مطالب

صفحه	عنوان
۱۹	۳-۴- طراحی
۲۲	۴-۴- پیاده‌سازی
۲۲	۱-۴-۴- قابلیت‌های دستیار صوتی
۲۴	۲-۴-۴- تبدیل گفتار به متن
۲۵	۳-۴-۴- تبدیل متن به گفتار
۲۵	۴-۴-۴- رابط کاربری گرافیکی
۲۶	۵-۴- جمع‌بندی
۲۷	فصل پنجم نتایج
۲۷	۱-۵- مقدمه
۲۷	۲-۵- دستورات دستیار صوتی
۲۹	۳-۵- کار کردن با دستیار صوتی
۳۱	۴-۵- نتایج پیاده‌سازی قابلیت‌ها
۳۳	۵-۵- جمع‌بندی
۳۴	منابع:

فهرست شکل‌ها

عنوان	صفحه
شکل ۱-۲: نمای کلی نرم‌افزار جارویس	۹
شکل ۱-۳: نمودار آموزش مدل Persian STT	۱۴
شکل ۲-۳: تعدادی پیش‌بینی مدل Wav2Vec2-Large-XLSR-53-Persian	۱۵
شکل ۱-۴: معماری چندلایه پروژه	۱۹
شکل ۲-۴: نمودار بسته برای معماری چندلایه پروژه	۲۰
شکل ۳-۴: نمودار توالی برای انجام دستورات متنی	۲۰
شکل ۴-۴: نمودار توالی برای انجام دستورات صوتی فارسی	۲۱
شکل ۵-۴: نمودار توالی برای انجام دستورات صوتی فارسی نیازمند پایگاه داده	۲۱
شکل ۶-۴: طراحی وایرفریم واسط کاربری نرم‌افزار	۲۲
شکل ۷-۴: محتوای فایل AIML	۲۵
شکل ۸-۴: عکس دستیار صوتی	۲۶
شکل ۱-۵: صفحه رابط کاربری	۲۹
شکل ۲-۵: صفحه رابط کاربری قبل از ارسال متن	۳۰
شکل ۳-۵: صفحه رابط کاربری بعد از ارسال متن	۳۰
شکل ۴-۵: صفحه رابط کاربری بعد از ارسال متن به صورت صوتی	۳۱
شکل ۵-۵: خروجی قابلیت آمار کرونا	۳۱
شکل ۶-۵: خروجی قابلیت وضعیت سیستم	۳۲
شکل ۷-۵: خروجی قابلیت ترجمه	۳۲
شکل ۸-۵: خروجی قابلیت اخبار	۳۲

فهرست جدول‌ها

صفحه

عنوان

جدول ۵-۱: نمونه دستور فارسی قابلیت‌ها ۲۷

مخفف‌ها:

فصل اول

مقدمه

۱-۱- هدف پروژه

در دنیای امروزی زمان اهمیت بسیاری دارد. با انتشار یافتن نرم‌افزارهای جدید برای ارائه خدمات به کاربران، پیچیدگی در استفاده از این قابلیت‌ها نیز افزایش یافته است. بنابراین دنیا به نرم‌افزارهای دستیار صوتی برای کاهش پیچیدگی‌ها و صرفه‌جویی در زمان نیاز پیدا کرده است. همچنین به دلیل عدم وجود دستیار صوتی‌ای برای سیستم عامل لینوکس که از زبان فارسی پشتیبانی کند، به همین سبب تصمیم گرفتیم یک نرم‌افزار دستیار صوتی برای سیستم عامل لینوکس برای کاربران فارسی و انگلیسی زبان پیاده‌سازی کنیم.

۱-۲- کاربردهای پروژه

پروژه خود به صورت واحد فقط برای کاربران سیستم عامل لینوکس کاربرد دارد اما مدل تبدیل گفتار به متنی که قرار است در این پروژه به کار برده شود به صورت مستقل می‌تواند در دستیار صوتی‌های سایر سیستم عامل‌ها، صنعت رباتیک و یا حتی در صنعت خودروسازی نیز استفاده شود.

۱-۳- ساختار پایان نامه

با توجه به مقدمه ای که ذکر گردید روند ارائه مطالب در این گزارش به این صورت است که در فصل دوم نرم افزارهای مشابه مانند Mycroft و Cortana معرفی می شوند سپس در فصل سوم پیش زمینه مورد نیاز برای پیاده سازی پروژه گفته می شود و در فصل چهارم پروژه شرح داده می شود و نهایتاً در فصل پنجم نتایج پیاده سازی پروژه مطرح می شود.

فصل دوم

معرفی نرم افزارهای مشابه

۲-۱- مقدمه

در این فصل قرار است به معرفی نرم افزارهای مشابه در سیستم عامل های متفاوت بپردازیم و با ویژگی های کلیدی آن ها آشنایی نسبی پیدا کنیم.

۲-۲- Jarvis

جارویس^۱ توسط سازمان غیرانتفاعی The Pennyworth Project ایجاد شده است و برای سیستم عامل ویندوز منتشر شده است. جارویس یک برنامه نرم افزاری عالی برای سنجش زمینه با پشتیبانی از پیش بینی است. از قابلیت یادگیری، پیش بینی ها، قوانین و حسگرها پشتیبانی می کند. در این قسمت ویژگی های کلیدی جارویس آورده شده است:

- هنگامی که برای اولین بار نرم افزار را راه اندازی می کنید، یک نماد کوچک در قسمت اعلان سیستم ظاهر می شود و از این طریق می توانید با جارویس تعامل کنید.
 - اگر بر روی آیکون کلیک راست کنید، منویی ظاهر می شود و تمام قابلیت هایی که این نرم افزار قادر به ارائه آن است را نشان می دهد.
 - شما قادر خواهید بود نرم افزار را آموزش دهید و پیش بینی های زمینه آن را بهبود دهید.
 - با گذشت زمان، نرم افزار در پیش بینی خودکار زمینه شما عالی خواهد شد.
- در شکل ۲-۱ یک نمای کلی از نرم افزار نشان داده شده است.

۲-۳- Braina

Braina مخفف Brain Artificial است و یک دستیار مجازی هوشمند است که به کاربر امکان می دهد از طریق دستورات صوتی با رایانه شخصی خود ارتباط برقرار کند.

¹ Jarvis



شکل ۱-۲: نمای کلی نرم افزار جارویس

در این قسمت ویژگی های کلیدی نرم افزار هوش مصنوعی Braina آورده شده است:

- یک رابط زبان انسانی، تشخیص صدا و نرم افزار اتوماسیون برای سیستم های دارای ویندوز است.
 - این نرم افزار عالی به کاربران اجازه می دهد تا به طور دقیق گفتار را به بیش از ۱۰۰ زبان مختلف تبدیل کنند.
 - با استفاده از Braina، می توان کامپیوتر خود را با استفاده از دستورات زبان طبیعی کنترل کرد.
 - این یک ربات چت ساده نیست و تمرکز اصلی آن این است که بسیار کاربردی باشد و به کاربران در انجام کارهای مختلف کمک کند.
 - کاربر قادر خواهد بود انواع دستورات را تایپ کند و حتی با آن صحبت کند و Braina به خوبی متوجه خواهد شد که از آن چه می خواهد.
 - این نرم افزار یک دستیار دیجیتالی است که مانند مغز انسان قادر به درک، تفکر و همچنین یادگیری از تجربه خود است.
 - Braina همچنین می تواند از مکالمات بیاموزد.
- این یک نرم افزار هوش مصنوعی چند منظوره است که یک محیط تک پنجره ای را برای کنترل رایانه شما و انجام طیف گسترده ای از وظایف مانند دیکته کردن، جستجوی آنلاین داده ها، پخش آهنگ، باز کردن و جستجوی فایل ها در رایانه شخصی، تنظیم آلارم و یادآوری، انجام عملیات ریاضی، به خاطر سپردن یادداشت ها، خودکارسازی کارهای مختلف، خواندن کتاب های الکترونیکی و غیره را فراهم می کند.

۲-۴- DataBot

DataBot یک ربات سخنگو مجازی است که برای ویندوز منتشر شده است. بسته به موضوع با صدای خودش به سوال کاربر پاسخ می‌دهد.

در این قسمت ویژگی‌های کلیدی DataBot آورده شده است:

- این نرم‌افزار قادر است موضوع مورد نیاز شما را به سرعت شناسایی کند.
- این برنامه قادر به پاسخگویی به سوالات مختلف است و گزارش‌هایی ایجاد می‌کند که شامل پاسخ، جزئیات مرتبط، مطالب، لینک‌ها و موارد دیگر می‌شود.
- DataBot به کاربر امکان می‌دهد تا با استفاده از پیام کوتاه، شبکه‌های اجتماعی و ایمیل، پاسخ‌ها را بدون زحمت به اشتراک بگذارد.
- خدمات و ماژول‌هایی که در DataBot موجود است شامل موارد زیر است:
فرهنگ لغت، فال، منشی، موضوعی، جوک، اخبار، صداها، آموزش مغز، معما، نقل قول و غیره.

۲-۵- Cortana

کورتانا^۱ یک ابزار مفید مبتنی بر هوش مصنوعی است که می‌توان از آن برای تنظیم یادآوری، دیکته کردن ایمیل‌ها، یافتن اطلاعات مختلف در اینترنت و غیره استفاده کرد. مایکروسافت^۲ در سال‌های اخیر مجموعه‌ای از ویژگی‌های جدید را به دستیار مجازی خود اضافه کرده است. اکنون می‌توان از آن در سایر سیستم عامل‌ها از جمله آندروید^۳ نیز استفاده کرد. البته برای اینکه کورتانا به درستی کار کند، باید به تمام اطلاعات رایانه شما دسترسی داشته باشد. برخی از کاربران ممکن است احساس کنند که این امر بر حریم خصوصی آن‌ها تأثیر می‌گذارد، بنابراین هنگام استفاده از کورتانا باید این واقعیت را در نظر داشت.

۲-۶- Computer Talker

Computer Talker یک نرم‌افزار تبدیل متن به گفتار برای ویندوز است که توسط Point64 ایجاد شده است. این ابزار ساده‌ای است که هر چیزی را که می‌نویسید تبدیل به گفتار می‌کند تا کامپیوتر آن را با صدای بلند صحبت کند.

در این قسمت ویژگی‌های کلیدی Computer Talker آورده شده است:

^۱ Cortana

^۲ Microsoft

^۳ Android

- این نرم‌افزار دارای یک رابط کاربری ساده است که استفاده از آن را حتی برای مبتدیان آسان‌تر می‌کند.
- این نرم‌افزار قادر به خواندن یا تبدیل متن به گفتار با استفاده از صدای زنانه طبیعی خود است.
- این برنامه به راحتی می‌تواند کلمات انگلیسی را تشخیص دهد و قادر است آن‌ها را به درستی تلفظ کند.

۷-۲- Mycroft

Mycroft اولین دستیار صوتی منبع‌باز جهان است. این نرم‌افزار روی سیستم عامل لینوکس، داخل خودرو و حتی روی Raspberry Pi اجرا می‌شود. می‌توان از آن در هر چیزی از یک پروژه علمی گرفته تا یک برنامه نرم‌افزاری سازمانی استفاده کرد.

در این قسمت ویژگی‌های کلیدی Mycroft آورده شده است:

- قادر است به انواع سؤالات واقعی پاسخ دهد.
- قادر است تاریخ، زمان، آب و هوا را به کاربر بگوید، آلارم تنظیم کند یا اخبار را پخش کند.
- قادر است موسیقی، رادیو، پادکست، یا فیلم و ویدیو را روی صفحه نمایش نزدیک شما پخش کند.
- اگر لوازم متصل به وای‌فای^۱ دارید، می‌تواند از طرف شما با آن‌ها صحبت کند.

۸-۲- جمع‌بندی

در این فصل با چند تا از بهترین نرم‌افزارهای مشابه با پروژه ما آشنا شدید. نرم‌افزارهای بسیاری در این حوزه ساخته شده است و روز به روز تعداد آن‌ها افزایش می‌یابد و پیشرفته‌تر میشوند ولی همانطور که مشاهده کردید اکثر نرم‌افزارهایی که در این فصل معرفی شد برای سیستم عامل ویندوز می‌باشند در نتیجه پروژه ما رقبای کمتری دارد و می‌تواند با دیگر نرم‌افزارها در سیستم عامل لینوکس رقابت کند.

^۱ Wi-Fi

فصل سوم

پیش‌زمینه

۳-۱- مقدمه

در این فصل قرار است به معرفی تکنولوژی‌های عمده مورد استفاده در پیاده‌سازی نرم‌افزار بپردازیم و همچنین با مدل‌های تبدیل گفتار به متن آشنا شویم.

۳-۲- تکنولوژی‌های مورد استفاده

تکنولوژی‌های عمده مورد استفاده در این پروژه زبان برنامه‌نویسی پایتون^۱ و AIML^۲ می‌باشد که در ادامه به معرفی آن‌ها می‌پردازیم.

۳-۲-۱- زبان برنامه‌نویسی پایتون

پایتون یک زبان برنامه‌نویسی شیء‌گرا، تفسیری، سطح بالا و همه منظوره است، که خیدو فان روسوم^۳ آن را طراحی کرده‌است و اولین بار در سال ۱۹۹۱ منتشر شده‌است. فلسفه اصلی طراحی پایتون «خوانایی بالای کد» است و نویسه‌های فاصله خالی در آن معنادار هستند و مکرر استفاده می‌شوند. ساختار زبانی و دیدگاه شیء‌گرا در پایتون به گونه‌ای طراحی شده‌است که به برنامه‌نویس امکان نوشتن کد منطقی و واضح را برای پروژه‌های کوچک و بزرگ می‌دهد.

۳-۲-۲- AIML

AIML مخفف عبارت Artificial Intelligence Markup Language است. AIML یک زبان نشانه‌گذاری مبتنی بر XML است که برای ایجاد برنامه‌های کاربردی هوش مصنوعی طراحی شده است. AIML امکان ایجاد رابط‌های انسانی را فراهم می‌کند در حالی که پیاده‌سازی را ساده برای برنامه‌نویسی، آسان برای درک و بسیار قابل نگهداری نگه می‌دارد. این آموزش اصول اولیه AIML را به شما آموزش می‌دهد.

^۱ Python

^۲ Artificial Intelligence Markup Language

^۳ Guido van Rossum

AIML عناصر زیادی دارد که در ادامه با مهم‌ترین آن‌ها آشنا می‌شویم:

• Category

دسته‌بندی‌ها در AIML یک واحد اساسی را تشکیل می‌دهند. یک دسته‌بندی حداقل از دو عنصر pattern و template تشکیل شده است. در اینجا یک دسته‌بندی ساده آورده شده است:

```
<category>
  <pattern>اسمت چیه</pattern>
  <template>من فنوم هستم</template>
</category>
```

هنگامی که این دسته بارگذاری می‌شود، یک ربات AIML به ورودی "اسمت چیه" با پاسخ "من فنوم هستم" پاسخ می‌دهد.

• Pattern

یک الگو رشته‌ای از کاراکترها است که برای مطابقت با یک یا چند ورودی کاربر در نظر گرفته شده است. یک الگو مانند:

```
<pattern>اسمت چیه</pattern>
```

تنها با یک ورودی مطابقت دارد اما الگوها ممکن است دارای حروف عامیانه نیز باشند که با یک یا چند کلمه مطابقت دارند. یک الگو مانند:

```
<pattern>چیه* </pattern>
```

با تعداد نامتناهی ورودی مطابقت دارد، از جمله "اسمت چیه"، "اسم شما چیه"، "فامیل شما چیه" و غیره.

• Template

یک template پاسخ به یک الگوی منطبق را مشخص می‌کند. یک template ممکن است به سادگی برخی از متن‌های تحت اللفظی باشد مانند:

```
<template>من فنوم هستم</template>
```

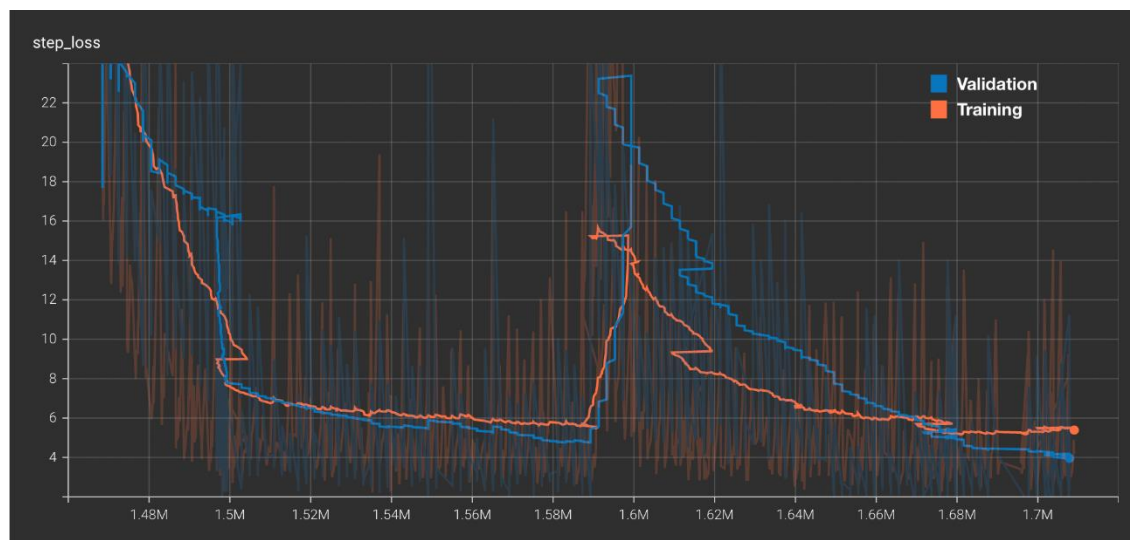
همچنین ممکن است از متغیرهایی استفاده کند که از قبل تعریف شده است و یا به template الگوهای دیگر اشاره بکند.

۳-۳- مدل‌های تبدیل گفتار به متن فارسی

در این بخش با چند مدل تبدیل گفتار به متن فارسی آشنا می‌شویم و آن‌ها را بررسی می‌کنیم.

Persian STT - ۱-۳-۳

این مدل توسط آقای مهدی حاج ملا احمد نراقی آموزش داده شده است. این مدل بر روی دیتاست^۱ Common Voice 9.0 Persian که دارای حدود ۲۷۱ ساعت داده می‌باشد آموزش داده شده است. در شکل ۱-۳ نمودار آموزش این مدل نشان داده شده است.



شکل ۱-۳: نمودار آموزش مدل Persian STT

بعد از ارزیابی نهایی، معیارهای WER^۲، CER^۳ و loss مدل به صورت زیر شده است:

WER: 0.108172, CER: 0.025068, loss: 1.292882

شایان ذکر است که بسیاری از خطاها به دلیل نحو زبان است. به عنوان مثال، مدل به تنهایی قادر به تمایز بین کلمات زیر نیست:

امیررضا و آهودخت، امیررضا و امیرآرشام، آهودخت و امیرآرشام

خوشبختانه scorer قادر است با توجه به متن، اکثر این موارد را مدیریت کند.

Wav2Vec2-Large-XLSR-53-Persian - ۲-۳-۳

مدل Wav2Vec2-Large-XLSR-53 با استفاده از Common Voice فارسی، توسط آقای مهرداد فراهانی به زبان فارسی Fine-tune شده است. در شکل ۲-۳ تعدادی از پیش‌بینی‌های این مدل را می‌توانید مشاهده کنید.

^۱ Dataset

^۲ Word Error Rate

^۳ Character Error Rate

```

reference: چرا معاشرت با هم ایمانان ما را محفوظ نگه میدارد
predicted: چرا معاشرت با هم ایمانان آ را م حفوظ نگه میدارد
---
reference: بولیوی پس از گویان فقیرترین کشور آمریکای جنوبی است
predicted: بولیوی پس از گویان فقیرترین کشور آمریکای جنوبی است
---
reference: بعد از مدتی اینکار برایم عادی شد
predicted: بعد از مدتی اینکار برایم عادی شد
---
reference: به نظر اون هم همینطوره
predicted: به نظر اون هم همینطوره
---
reference: هیچ مایونزی دارید
predicted: هیچ مایونزی دارید
---
reference: هیچ یک از آنان کاری به سنگ نداشتند
predicted: هیچ شک از آنان کاری به سنگ نداشتند
---
reference: می‌خواهم کمی کتاب شعر ببینم
predicted: می‌خواهم کتاب شعر ببینم

```

شکل ۳-۲: تعدادی پیش‌بینی مدل Wav2Vec2-Large-XLSR-53-Persian

نتیجه ارزیابی به صورت زیر شده است:

WER: 10.36%

Vosk-Model-Fa -۳-۳-۳

Vosk یک جعبه ابزار تشخیص گفتار منبع باز آفلاین است. تشخیص گفتار را برای بیش از ۲۰ زبان و گویش فراهم می‌کند.

Vosk دو نوع مدل دارد. مدل‌های بزرگ و کوچک، مدل‌های کوچک برای برخی کارهای محدود در برنامه‌های تلفن همراه ایده‌آل هستند. آن‌ها می‌توانند روی گوشی‌های هوشمند اجرا شوند. آن‌ها همچنین برای برنامه‌های دسکتاپ توصیه می‌شوند. مدل‌های بزرگ برای تشخیص با دقت بالا می‌باشند.

مدل vosk-model-fa-0.5، یک مدل بزرگ با دایره لغات زیاد می‌باشد و دقت نسبتاً خوبی دارد.

۳-۴- جمع‌بندی

در این فصل با زبان برنامه‌نویسی پایتون و AIML که به طور عمده در پیاده‌سازی پروژه استفاده می‌شوند آشنا شدیم. تبدیل گفتار به متن در این پروژه یک جزء اصلی و مهم می‌باشد به همین دلیل به بررسی و معرفی مدل‌های تبدیل گفتار به متن فارسی پرداخته شد تا با بهترین آن‌ها آشنا شویم. در نهایت این دانش به ما کمک می‌کند تا بهترین مدل موجود را انتخاب کنیم و کارایی پروژه را افزایش دهیم.

فصل چهارم

شرح پروژه

۴-۱- مقدمه

در این فصل ابتدا با نیازمندی‌های این پروژه آشنا می‌شویم سپس در مورد طراحی و پیاده‌سازی اجزای پروژه توضیح داده می‌شود.

۴-۲- فهرست نیازمندی‌ها

۱. این نرم‌افزار باید دستورات کاربر را به صورت صوتی قبول کند.
 - ۱/۱. این نرم‌افزار باید دستورات کاربر را به صورت صوتی با زبان فارسی قبول کند.
 - ۱/۲. این نرم‌افزار باید دستورات کاربر را به صورت صوتی با زبان انگلیسی قبول کند.
۲. این نرم‌افزار باید دستورات کاربر را به صورت متنی قبول کند.
 - ۲/۱. این نرم‌افزار باید دستورات کاربر را به صورت متنی با زبان فارسی قبول کند.
 - ۲/۲. این نرم‌افزار باید دستورات کاربر را به صورت متنی با زبان انگلیسی قبول کند.
۳. این نرم‌افزار باید به صورت صوتی به کاربر جواب دهد.
 - ۳/۱. این نرم‌افزار باید با زبان فارسی به صورت صوتی به کاربر جواب دهد.
 - ۳/۲. این نرم‌افزار باید با زبان انگلیسی به صورت صوتی به کاربر جواب دهد.
۴. این نرم‌افزار باید به صورت متنی به کاربر جواب دهد.
 - ۴/۱. این نرم‌افزار باید با زبان فارسی به صورت متنی به کاربر جواب دهد.
 - ۴/۲. این نرم‌افزار باید با زبان انگلیسی به صورت متنی به کاربر جواب دهد.
۵. این نرم‌افزار باید کار ترجمه از یک زبان به زبان دیگر را انجام دهد.
۶. این نرم‌افزار باید وضعیت آب و هوای شهر درخواستی کاربر را نشان دهد.
۷. این نرم‌افزار باید اطلاعات سیستم را به کاربر نشان دهد.
۸. این نرم‌افزار باید برنامه‌هایی که کاربر بر روی سیستم خود دارد را باز و بسته کند.

۹. این نرم افزار باید عملیات جست و جو در گوگل^۱ را انجام دهد.
۱۰. این نرم افزار باید عملیات ماشین حساب را انجام دهد.
۱۱. این نرم افزار باید عملیات سیستمی را انجام دهد.
- ۱۱/۱. این نرم افزار باید صدا را تغییر دهد.
- ۱۱/۱/۱. این نرم افزار باید صدا را کم و زیاد کند.
- ۱۱/۱/۲. این نرم افزار باید صدا را قطع و وصل کند.
- ۱۱/۲. این نرم افزار باید سیستم را خاموش و Restart کند.
- ۱۱/۳. این نرم افزار باید نور صفحه را کم و زیاد کند.
- ۱۱/۴. این نرم افزار باید وای فای را خاموش و روشن کند.
- ۱۱/۵. این نرم افزار باید بلوتوث^۲ را خاموش و روشن کند.
۱۲. این نرم افزار باید قابلیت درست کردن یادداشت را داشته باشد.
۱۳. این نرم افزار باید قابلیت تنظیم زنگ هشدار را داشته باشد.
۱۴. این نرم افزار باید اخبار روزانه را به کاربر نشان دهد.
۱۵. این نرم افزار باید ساعت و تاریخ زمان حال را به کاربر نشان دهد.
۱۶. این نرم افزار باید موسیقی را پخش و متوقف کند.
۱۷. این نرم افزار باید تصویر پس زمینه را تغییر دهد.
۱۸. این نرم افزار باید عملیات برنامه ریزی را برای کاربر انجام دهد.
۱۹. این نرم افزار باید برای کاربر لطیفه تعریف کند.
۲۰. این نرم افزار باید متنی که کاربر مشخص می کند را بخواند.
۲۱. این نرم افزار باید آمار کنونی تعداد مبتلایان و فوت شدگان کرونا در در جهان و کشوری که کاربر درخواست می کند را نشان دهد.
۲۲. این نرم افزار باید BMI^۳ بدن کاربر را حساب کند و در مورد نتیجه ای که از آن حاصل شده، اطلاعات مورد نظر را به کاربر بدهد.
۲۳. این نرم افزار باید نقشه شهرهای مختلف را به کاربر نشان دهد.
۲۴. این نرم افزار باید پکیج^۴ های مورد نیاز کاربر را نصب کند.
۲۵. این نرم افزار باید با کاربر به صورت چت^۱ تعامل کند.

^۱ Google

^۲ Bluetooth

^۳ Body Mass Index

^۴ Package

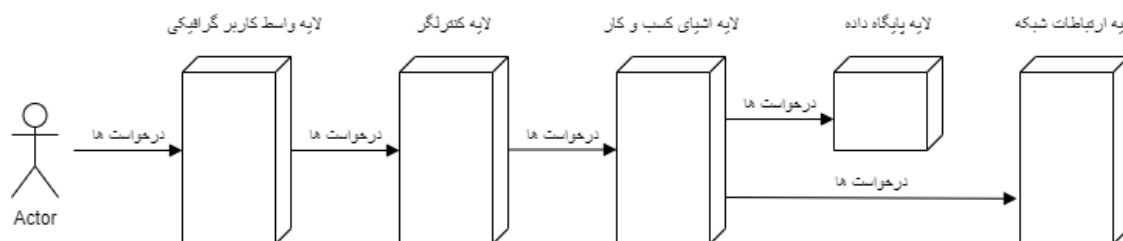
۲۶. این نرم‌افزار باید دارای یک رابط کاربری گرافیکی باشد.

۳-۴- طراحی

با توجه به نیازمندی‌های پروژه و تعاملی بودن سیستم، معماری چندلایه برای پروژه در نظر گرفته شده است. در طراحی معماری، سیستم به لایه‌های زیر تقسیم شده است و وظیفه هر لایه به صورت خلاصه توضیح داده شده است:

۱. **لایه واسط کاربر گرافیکی:** وظیفه این لایه نمایش واسط کاربر گرافیکی به کاربر و ارسال درخواست‌های کاربر به لایه کنترلگر می‌باشد.
۲. **لایه کنترلگر:** این لایه وظیفه کنترل درخواست کاربر را دارد و مسئول ارتباط با لایه‌ی اشیای کسب‌وکار برای انجام درخواست کاربر می‌باشد.
۳. **لایه‌ی اشیای کسب‌وکار:** در این لایه، تبدیل گفتار به متن، انجام درخواست کاربر و تبدیل متن به گفتار صورت می‌گیرد و با توجه به درخواست کاربر، در صورت نیاز از لایه‌ی پایگاه‌داده و ارتباطات شبکه نیز استفاده می‌کند.
۴. **لایه‌ی پایگاه‌داده:** این لایه وظیفه مدیریت پایگاه‌داده‌های مربوط به قابلیت‌های پروژه را دارد.
۵. **لایه‌ی ارتباطات شبکه:** این لایه برای ارتباط با API^۲ بعضی از پکیج‌های مورد استفاده در پروژه می‌باشد.

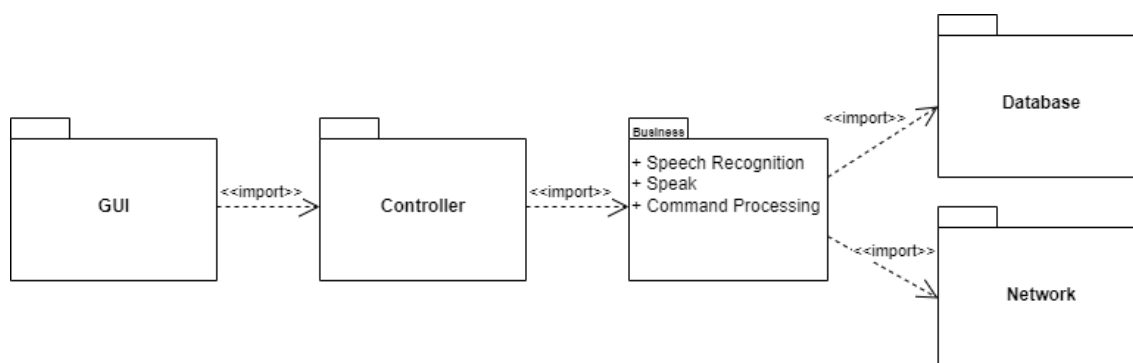
در شکل ۴-۱ معماری طراحی شده برای پروژه نشان داده شده است. همچنین نمودار بسته برای معماری پروژه نیز طراحی شده است که در شکل ۴-۲ قابل مشاهده می‌باشد.



شکل ۴-۱: معماری چندلایه پروژه

^۱ Chat

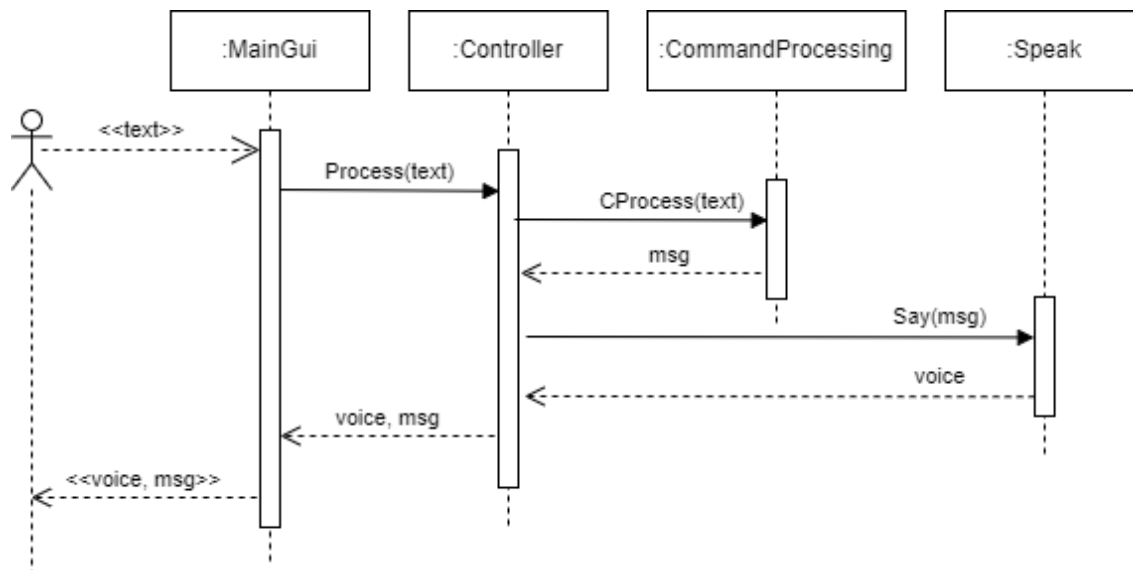
^۲ Application Programming Interface



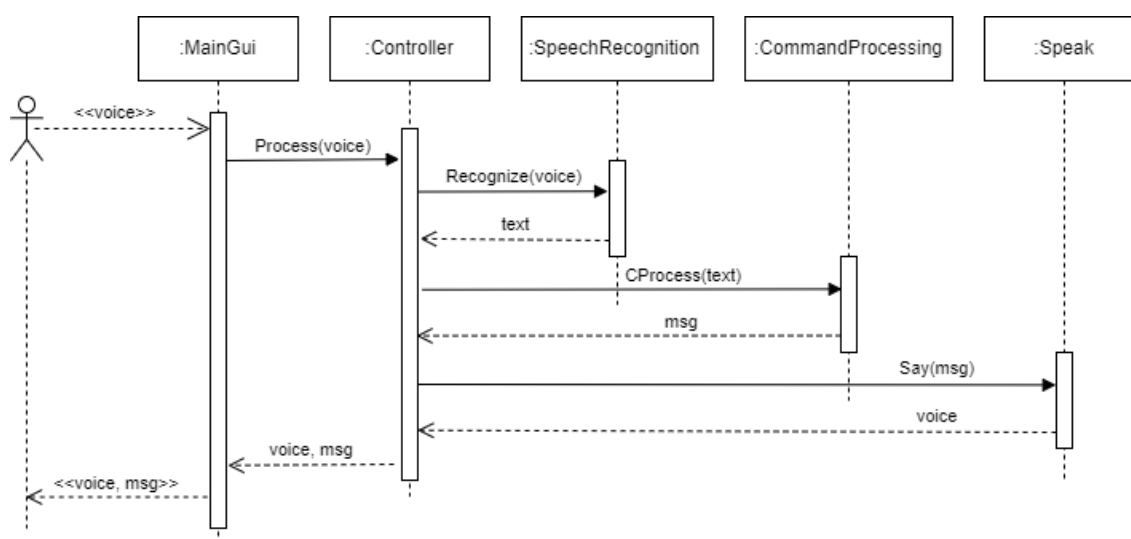
شکل ۴-۲: نمودار بسته برای معماری چندلایه پروژه

همان‌طور که در شکل ۴-۲ مشاهده می‌شود لایه‌ی اشیای کسب‌وکار دارای کلاس Speech Recognition برای تبدیل گفتار به متن، کلاس Speak برای تبدیل متن به گفتار و کلاس Command Processing برای پردازش دستور کاربر می‌باشد.

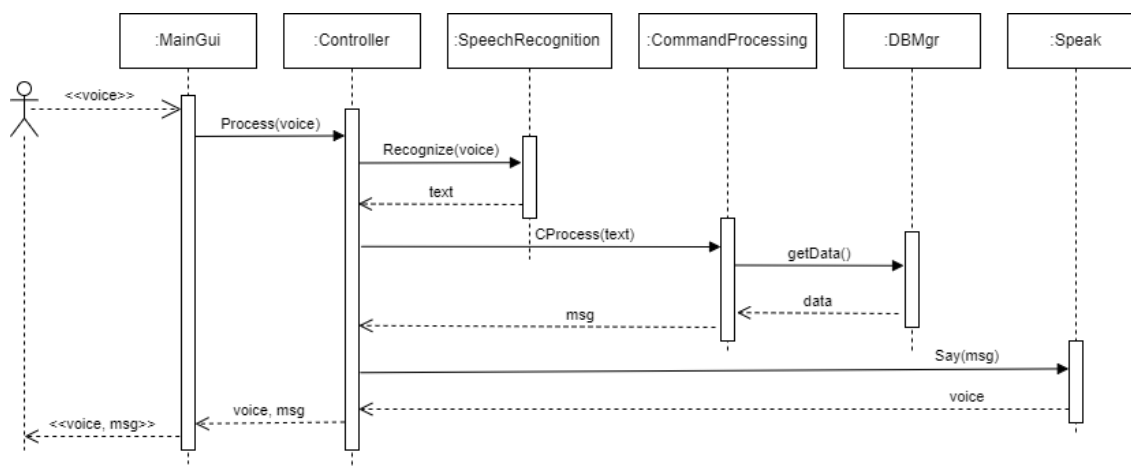
چند نمونه نمودار توالی ساده برای درک نحوه ارتباطات اجزای پروژه با یکدیگر، طراحی شده است. شکل ۴-۳ نمودار توالی برای انجام دستورات متنی کاربر را نشان می‌دهد، شکل ۴-۴ هم نمودار توالی برای انجام دستورات صوتی کاربر به زبان فارسی را نشان می‌دهد و در شکل ۴-۵ هم نمودار توالی برای انجام دستورات صوتی کاربر به زبان فارسی که نیازمند پایگاه‌داده می‌باشند، آورده شده است.



شکل ۴-۳: نمودار توالی برای انجام دستورات متنی



شکل ۴-۴: نمودار توالی برای انجام دستورات صوتی فارسی



شکل ۴-۵: نمودار توالی برای انجام دستورات صوتی فارسی نیازمند پایگاه داده

در مرحله طراحی نرم افزار، یک طراحی وایرفریم^۱ برای رابط کاربری نرم افزار نیز کشیده شده است که در شکل ۴-۶ آورده شده است.

^۱ Wireframe



شکل ۴-۶: طراحی وایرفریم واسط کاربری نرم افزار

۴-۴- پیاده سازی

اجزای کلی پروژه به صورت زیر می باشد:

۱. قابلیت های دستیار صوتی
۲. تبدیل گفتار به متن
۳. تبدیل متن به گفتار
۴. رابط کاربری گرافیکی

در این بخش به توضیح نحوه پیاده سازی هر کدام از این اجزا به طور جداگانه و به صورت خلاصه پرداخته خواهد شد.

۴-۴-۱- قابلیت های دستیار صوتی

برخی از این قابلیت ها با استفاده از پکیج های لینوکس از طریق ترمینال^۱ انجام می شود و برای دیگر قابلیت ها به طور جداگانه برای هر کدام یک برنامه ی پایتون نوشته شده است که در آن ها نیز از پکیج های مورد نیاز پایتون استفاده شده است. پیاده سازی هر کدام از قابلیت ها به صورت خلاصه در اینجا توضیح داده شده است:

۱. **تنظیم زنگ هشدار:** این کار با استفاده از دستور at در ترمینال لینوکس انجام می شود به این صورت که دستور پخش صدای هشدار و زمان آن را به این دستور داده می شود.
۲. **بلوتوث:** خاموش و روشن کردن بلوتوث با استفاده از دستور rfkill در ترمینال لینوکس انجام می شود.
۳. **BMI:** برای محاسبه BMI یک برنامه پایتون نوشته شده است که با استفاده از وزن و قد، مقدار

^۱ Terminal

- BMI بدن را محاسبه می‌کند.
۴. **تغییر روشنایی:** افزایش، کاهش و یا تنظیم روشنایی صفحه با استفاده از دستور `brightnessctl` در ترمینال لینوکس انجام می‌شود.
۵. **ماشین حساب:** انجام محاسبات ریاضی با استفاده از دستور `gnome-calculator` انجام می‌شود.
۶. **آمار کرونا:** برای به دست آوردن آمار کرونا یک برنامه پایتون نوشته شده است که در آن با استفاده از پکیج `covid`، آمار کلی کرونا و یا یک کشور به دست آورده می‌شود.
۷. **ساعت و تاریخ:** ساعت و یا تاریخ با استفاده از دستور `date` در ترمینال لینوکس نشان داده می‌شود.
۸. **باز کردن برنامه:** این کار با استفاده از دستور `gtk-launch` در ترمینال لینوکس انجام می‌شود.
۹. **نقشه:** ابتدا آدرس URL گوگل مپ مکان مورد نظر ایجاد می‌شود سپس با استفاده از دستور `xdg-open` در ترمینال لینوکس، این آدرس درون مرورگر باز می‌شود.
۱۰. **موسیقی:** عملیات مربوط به پخش موسیقی با استفاده از دستور `mocp` در ترمینال لینوکس انجام می‌شود.
۱۱. **اخبار:** برای به دست آوردن اخبار یک برنامه پایتون نوشته شده است که در آن با استفاده از پکیج‌های `beautifulsoup4` و `urllib`، سایت آخرین خبر کراول^۱ می‌شود و آدرس اخبار به دست آورده می‌شود و در نهایت با استفاده از پکیج `newspaper`، اخبار دانلود^۲ و نشان داده می‌شود.
۱۲. **یادداشت:** برای ایجاد و مدیریت یادداشت‌ها یک برنامه پایتون نوشته شده است که در آن با استفاده از پکیج `sqlite3`، یک پایگاه داده برای نگهداری یادداشت‌ها ایجاد شده است و می‌توان یادداشت‌ها را ایجاد یا حذف کرد و یا نمایش داد.
۱۳. **خاموش و Restart:** این کار با استفاده از دستورهای `shutdown` و `reboot` در ترمینال لینوکس انجام می‌شود.
۱۴. **خواندن:** برای خواندن متن انتخاب شده یک برنامه پایتون نوشته شده است که در آن با استفاده از پکیج `os`، متن انتخاب شده به دست آورده می‌شود.
۱۵. **وضعیت سیستم:** با استفاده از ترکیب دستورهای `free`، `df`، `top`، `grep` و `acpi` در ترمینال لینوکس، وضعیت سیستم نمایش داده می‌شود.
۱۶. **ترجمه:** برای ترجمه جمله مورد نظر کاربر یک برنامه پایتون نوشته شده است که در آن با استفاده از پکیج `google_trans_new` این کار انجام می‌شود.
۱۷. **تغییر صدا:** افزایش، کاهش، تنظیم، قطع و وصل صدا سیستم با استفاده از دستور `amixer` در

^۱ Crawl^۲ Download

ترمینال لینوکس انجام می‌شود.

۱۸. **تغییر پس‌زمینه:** برای این کار یک برنامه پایتون نوشته شده است که در آن با استفاده از پکیج‌های

gi.repository، glob و os به صورت تصادفی یک پس‌زمینه انتخاب می‌شود.

۱۹. **آب و هوا:** نشان دادن وضعیت آب و هوا با استفاده از اکستنشن‌ها^۱ درون AIML انجام می‌شود.

۲۰. **برنامه‌ریزی:** عملیات برنامه‌ریزی نیز با استفاده از اکستنشن‌ها درون AIML انجام می‌شود.

۲۱. **گوگل:** ابتدا آدرس URL گوگل عنوان مورد نظر ایجاد می‌شود سپس با استفاده از دستور xdg-

open در ترمینال لینوکس، این آدرس درون مرورگر باز می‌شود.

۲۲. **وای‌فای:** روشن و خاموش کردن وای‌فای با استفاده از دستور nmcli در ترمینال لینوکس انجام می-

شود.

۲۳. **باز و بسته کردن:** باز و بسته کردن پنجره‌ها با استفاده از دستور xdotool در ترمینال لینوکس انجام

می‌شود.

برای هر کدام از قابلیت‌ها، فایل‌های AIML برای زبان فارسی و انگلیسی ساخته شده است که دارای الگوهای زبانی مرتبط با آن قابلیت می‌باشد. برای مثال شکل ۴-۱ قسمتی از محتوای یک فایل AIML را نشان داده است.

در نهایت زمانی که کاربر دستوری را وارد می‌کند، در میان الگوهای فایل‌های AIML ساخته شده با توجه به دستور وارد شده اگر یکی از الگوهای تطابق داده شود دستور مرتبط با آن الگو اجرا می‌شود و اگر الگویی تطابق داده نشود به کاربر اطلاع می‌دهد که جوابی برای این دستور ندارد.

۴-۲- تبدیل گفتار به متن

در فصل سوم مدل‌های تبدیل گفتار به متن فارسی معرفی شد. بعد از آزمایش کردن آن‌ها به صورت عملی، این نتیجه به دست آمد که مدل Vosk-Model-Fa از دیگر مدل‌ها دقت بیشتری دارد و همچنین از نظر زمان پاسخگویی نیز عملکرد بهتری دارد. در نتیجه برای تبدیل گفتار به متن فارسی از این مدل استفاده شده است. برای تبدیل گفتار به متن انگلیسی به دلیل اینکه این مدل‌ها حجم بالایی دارند و برای تبدیل گفتار به متن فارسی از مدل استفاده شده است، برای کاهش استفاده از حافظه رم هنگام اجرای نرم‌افزار از API تشخیص

^۱ Extensions

```

<category>
  <pattern>سرچ در گوگل</pattern>
  <template>
    <think><set name="language">persian</set></think>
    چه چیزی را می‌خواهی در گوگل جستجو کنی
  </template>
</category>

<category>
  <pattern>جستجو در گوگل</pattern>
  <template>
    <think><set name="language">persian</set></think>
    چه چیزی را می‌خواهی در گوگل جستجو کنی
  </template>
</category>

<category>
  <pattern>_</pattern>
  <that>چه چیزی را می‌خواهی در گوگل جستجو کنی</that>
  <template>
    <think><set name="language">persian</set></think>
    <think><system>bash -c "xdg-open http://www.google.com/search?q='<star/>'&amp;> /dev/null &amp;"</system></think>
    به قرین
  </template>
</category>

```

شکل ۴-۷: محتوای فایل AIML

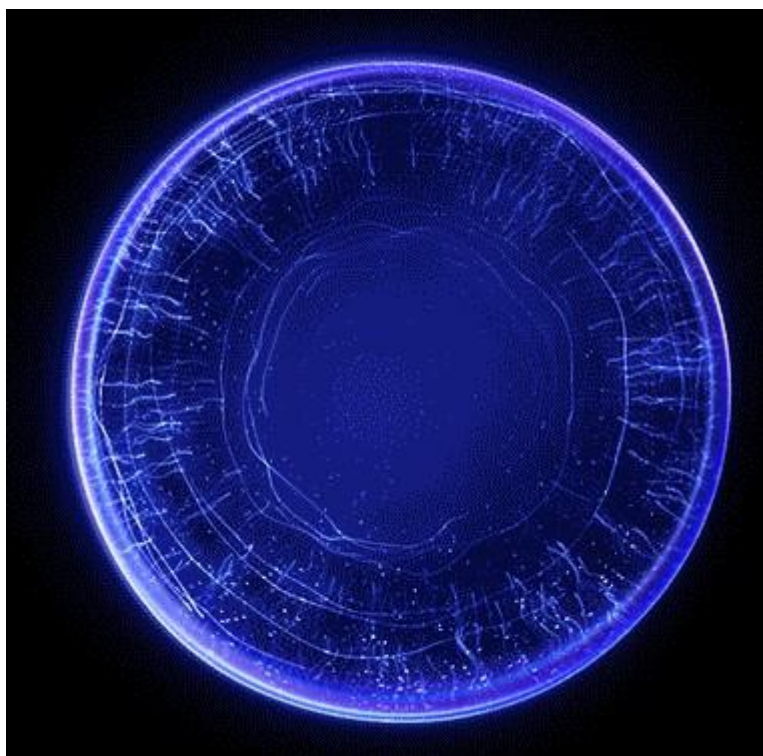
گفتار گوگل استفاده شده است. در نتیجه برای دستورات صوتی انگلیسی نیاز به اینترنت نیز می‌باشد.

۴-۳- تبدیل متن به گفتار

صدایی که با کاربر ارتباط برقرار می‌کند در کیفیت ارتباط آن‌ها تاثیر دارد. در نتیجه صدایی که جملات را می‌خواند باید واضح و با کیفیت باشد تا ارتباطی خوب با کاربر برقرار شود. برای تبدیل متن به گفتار از کتابخانه pyttsx3 در پایتون استفاده شده است. این کتابخانه به صورت آفلاین نیز کار می‌کند و از زبان‌های مختلفی مانند فارسی و انگلیسی پشتیبانی می‌کند. جملاتی که دستیار صوتی باید جواب بدهد به این کتابخانه داده می‌شود و با توجه به فارسی یا انگلیسی بودن جمله روی آن زبان باید تنظیم گردد و در نهایت به صورت صوتی برای کاربر جمله مورد نظر را می‌خواند.

۴-۴-۴ رابط کاربری گرافیکی

رابط کاربری باید کار کردن با آن برای کاربر آسان باشد و همچنین از نظر گرافیکی نیز زیبا و چشم‌نواز باشد تا برای کاربر خسته‌کننده نباشد. ابتدا با استفاده از نرم‌افزار Qt، صفحه رابط کاربری طراحی می‌شود سپس با استفاده از محتوای فایل خروجی Ui، در پایتون با استفاده از کتابخانه PyQt5 این صفحه پیاده‌سازی می‌شود. عکس دستیار صوتی که در طراحی صفحه رابط کاربری به صورت متحرک به کار رفته است را می‌توانید در شکل ۴-۲ مشاهده نمایید.



شکل ۴-۸: عکس دستیار صوتی

۴-۵- جمع‌بندی

برای کار کردن با دستیار صوتی ابتدا نیاز است با قابلیت‌هایی که دارد آشنا باشیم پس در این فصل نیازمندی‌های پروژه و قابلیت‌هایی که دستیار صوتی دارد معرفی شد. همچنین طراحی و نحوه پیاده‌سازی اجزای پروژه به صورت خلاصه توضیح داده شد تا آشنایی نسبی با پیاده‌سازی پروژه داشته باشیم.

فصل پنجم

نتایج

۵-۱- مقدمه

در این فصل ابتدا برای هر قابلیت به یک نمونه دستور برای استفاده از آن قابلیت اشاره می‌شود سپس نحوه کار کردن با دستیار صوتی توضیح داده می‌شود.

۵-۲- دستورات دستیار صوتی

برای کار کردن با دستیار صوتی و استفاده از قابلیت‌های آن نیاز به دانستن دستورات مرتبط با هر قابلیت می‌باشد در نتیجه در این بخش برای هر قابلیت یک نمونه دستور فارسی آورده شده است. در جدول ۵-۱ می‌توانید نمونه دستور فارسی مربوط به هر قابلیت را مشاهده نمایید.

جدول ۵-۱: نمونه دستور فارسی قابلیت‌ها

قابلیت دستیار صوتی	نمونه دستور مرتبط با قابلیت
بلوتوث	بلوتوث را روشن کن بلوتوث را خاموش کن
وای‌فای	وای‌فای را روشن کن وای‌فای را خاموش کن
گوگل	سرچ ^۱ در گوگل
تغییر پس‌زمینه	والپیپر را تغییر بده
ساعت و تاریخ	ساعت چند است تاریخ را نشان بده امروز چندم است

¹ Search

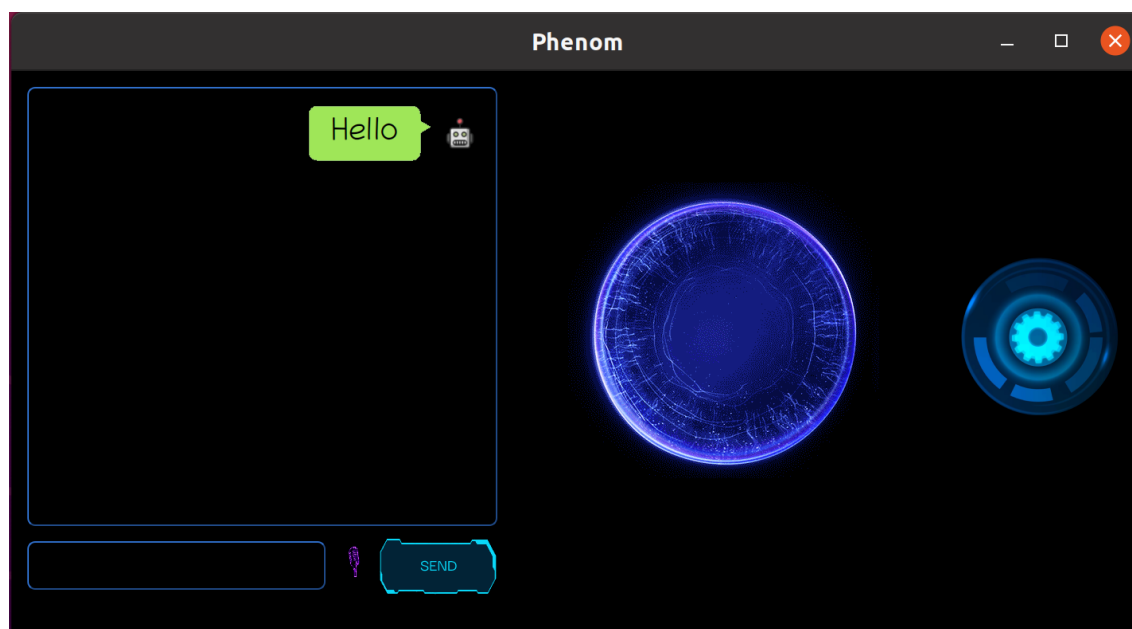
نقشه	تقشه اصفهان را نشان بده
خواندن	این متن را بخوان
خاموش و Restart	سیستم را خاموش کن سیستم را ری استارت کن
وضعیت سیستم	وضعیت سیستم را نشان بده
ترجمه	ترجمه دانشگاه اصفهان
یادداشت	یادداشت اضافه کن حذف یادداشت یادداشت نشان بده همه یادداشت ها را نشان بده
اخبار	اخبار را نشان بده اخبار برگزیده را نشان بده
آمار کرونا	آمار کرونا را نشان بده آمار کرونا ایران را نشان بده
باز و بسته کردن	این تب ^۱ را ببند یک پنجره باز کن
BMI	بی ام آی من را حساب کن
ماشین حساب	محاسبه معادله
زنگ هشدار	تنظیم زنگ هشدار لیست زنگ هشدار حذف زنگ هشدار شماره ۲
تغییر روشنایی	افزایش روشنایی ۲۰ درصد کاهش روشنایی ۲۰ درصد تنظیم روشنایی به ۶۰ درصد
تغییر صدا	افزایش صدا ۲۰ درصد کاهش صدا ۲۰ درصد تنظیم صدا به ۶۰ درصد

^۱ Tab

آب و هوا	آب و هوا در اصفهان
موسیقی	اجرا سرور پخش کننده آهنگ شروع پخش کننده آهنگ پخش لیست پخش آهنگ بعدی آهنگ قبلی آهنگ را متوقف کن ادامه آهنگ خروج از پخش کننده آهنگ

۵-۳- کار کردن با دستیار صوتی

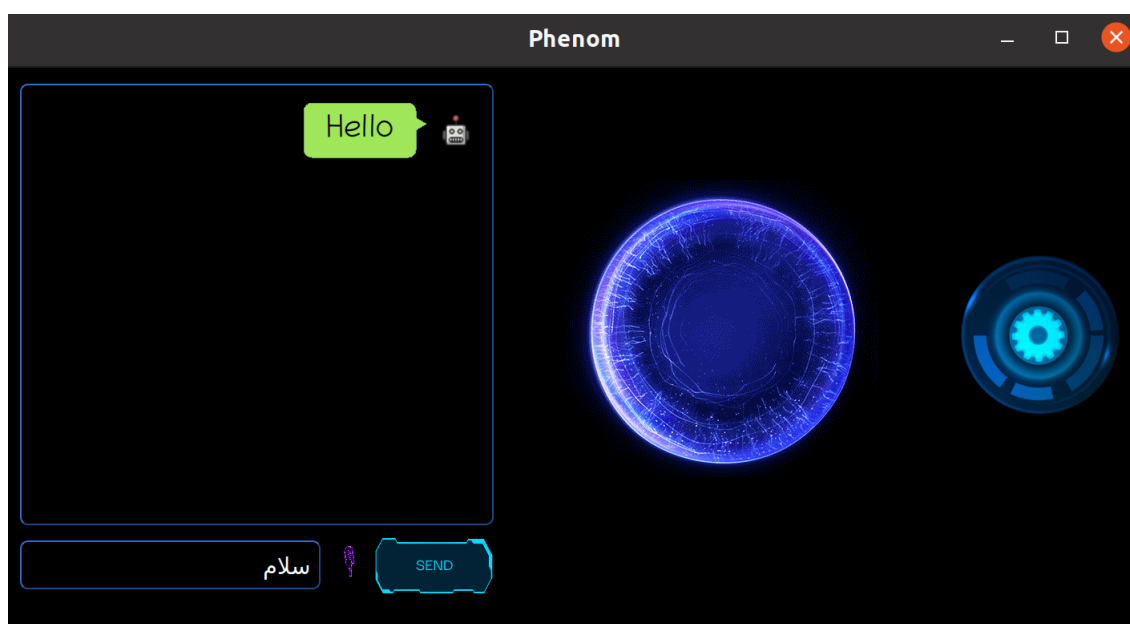
در این بخش نحوه تعامل با دستیار صوتی به همراه عکس توضیح داده می‌شود. ابتدا نرم‌افزار را اجرا می‌کنیم و بعد از چند ثانیه صفحه رابط کاربری نشان داده می‌شود. در شکل ۵-۱ صفحه رابط کاربری نشان داده شده است.



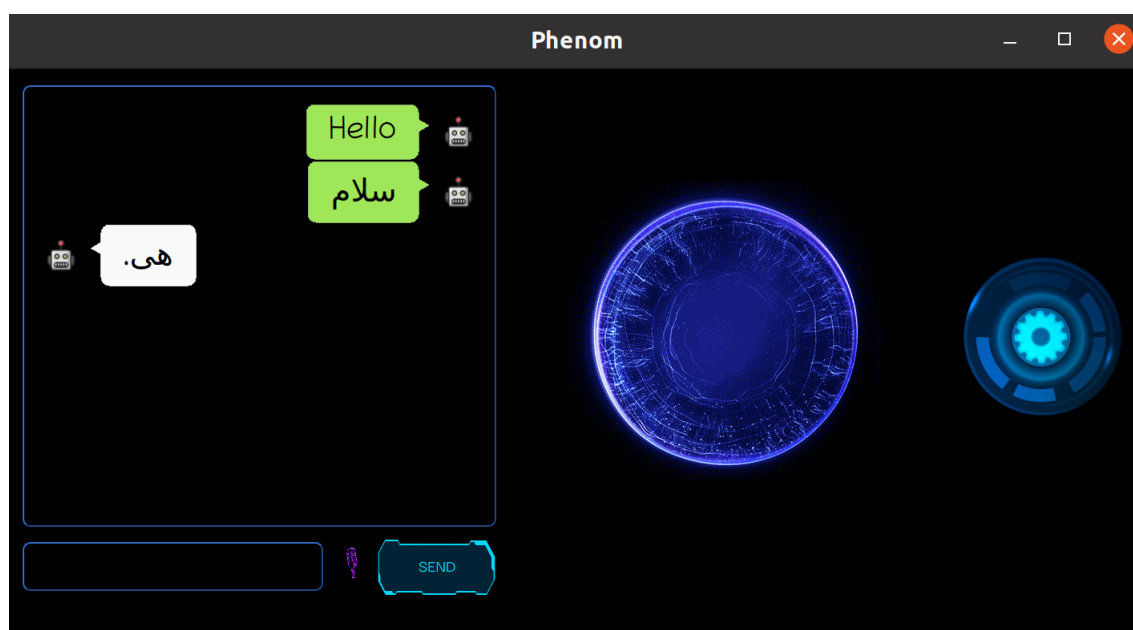
شکل ۵-۱: صفحه رابط کاربری

سپس درون باکس^۱ کوچک می‌توان متن مورد نظر را دستی وارد نمود و با استفاده از دکمه SEND متن مورد نظر را ارسال نمود. در شکل ۵-۲ و شکل ۵-۳ این عملیات نشان داده شده است.

^۱ Box



شکل ۵-۲: صفحه رابط کاربری قبل از ارسال متن



شکل ۵-۳: صفحه رابط کاربری بعد از ارسال متن

همچنین می‌توان با استفاده از آیکون^۱ میکروفون به صورت صوتی متن مورد نظر را ارسال کرد. کافی است روی میکروفون کلیک کرد و بعد از چند لحظه متن مورد نظر را به زبان آورد و در پایان سکوت کرد. در شکل ۵-۴ این عملیات نشان داده شده است.

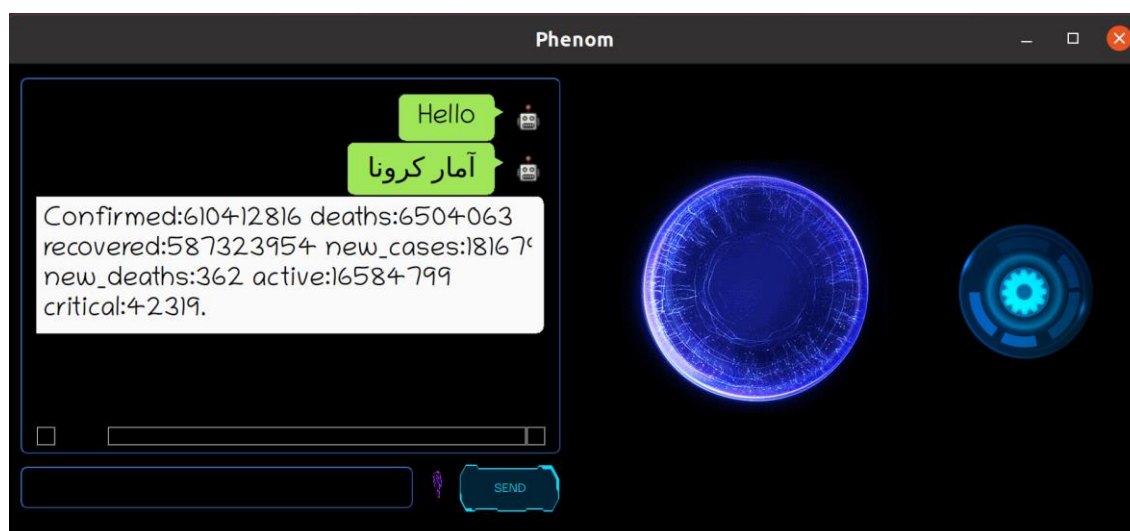
^۱ Icon



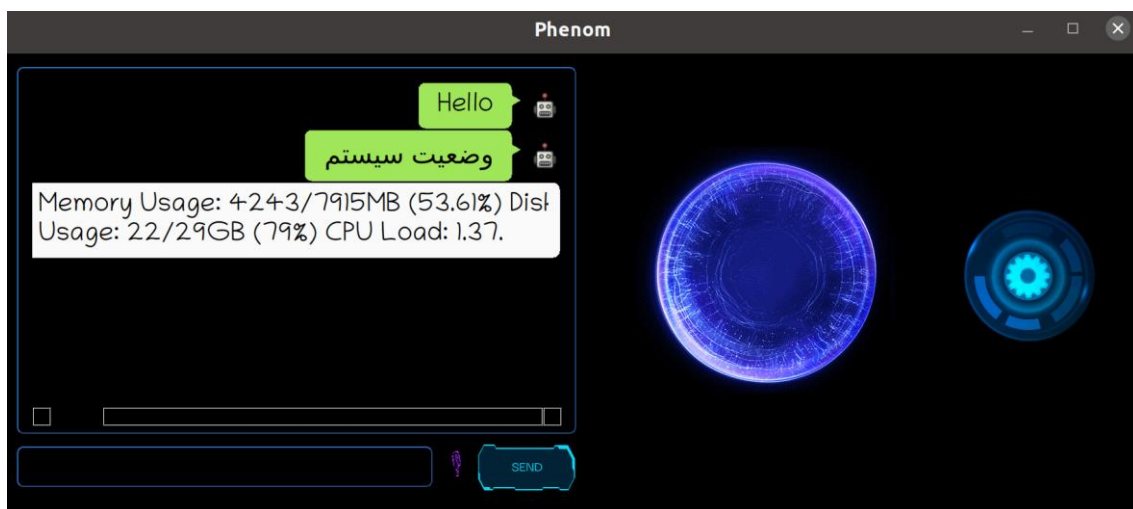
شکل ۵-۴: صفحه رابط کاربری بعد از ارسال متن به صورت صوتی

۵-۴- نتایج پیاده‌سازی قابلیت‌ها

در این بخش نتایج پیاده‌سازی قابلیت‌هایی که خروجی آن‌ها به صورت متنی می‌باشند و در عکس قابل مشاهده می‌باشند، آورده شده است. در شکل‌های ۵-۵، ۵-۶، ۵-۷ و ۵-۸ خروجی تعدادی از این قابلیت‌ها قابل مشاهده است.



شکل ۵-۵: خروجی قابلیت آمار کرونا



شکل ۵-۶: خروجی قابلیت وضعیت سیستم



شکل ۵-۷: خروجی قابلیت ترجمه



شکل ۵-۸: خروجی قابلیت اخبار

۵-۵- جمع‌بندی

در نهایت در این پروژه یک دستیار صوتی برای لینوکس پیاده‌سازی شد و نتایج پیاده‌سازی ارائه گردید. در آینده می‌توان این نرم‌افزار را توسعه نمود و قابلیت‌های جدیدی به آن اضافه کرد و همچنین مدل تبدیل گفتار به متن آموزش داد تا نیازی به مدل‌های شرکت‌ها یا اشخاص دیگر نباشد. در ابتدا تلاش شد یک مدل گفتار به متن فارسی آموزش داده شود و تحقیق‌هایی در مورد آن شد و کد مورد نظر برای آموزش مدل نیز نوشته شد ولی در آموزش نهایی به دلیل زمان‌بر بودن و همچنین کمبود منابع به سرانجام نرسید. لازم به ذکر است این پروژه ادامه یک پروژه‌ی درسی که با همکاری آقای نوید شاقوزائی پیاده‌سازی شده است می‌باشد.

منابع:

- [1] “6 best PC software that can talk to you,” Windows Report - Error-free Tech Life, Apr. 02, 2020. <https://windowsreport.com/talking-pc-software>.
- [2] senseid23@7h4r, “About Mycroft | The Open Source Artificial Intelligence Voice Assistant,” Mycroft. <https://mycroft.ai/about-mycroft>.
- [3] “AIML docs,” AIML Foundation. [Online]. Available: <http://www.aiml.foundation/doc.html>.
- [4] “M3HRDADFI/WAV2VEC2-large-xlsr-Persian-v3 · hugging face,” m3hrdadfi/wav2vec2-large-xlsr-persian-v3 · Hugging Face. [Online]. Available: <https://huggingface.co/m3hrdadfi/wav2vec2-large-xlsr-persian-v3>.
- [5] Oct4Pie, “Oct4pie/Persian-STT: A text-to-speech model developed using 🐱 STT,” GitHub. [Online]. Available: <https://github.com/Oct4Pie/persian-stt>.
- [6] Vosk offline speech recognition API. [Online]. Available: <https://alphacephei.com/vosk/>.