

طرح سرویس تبدیل نوشتار به گفتار



شرکت فناوری اطلاعات و ارتباطات پاسارگاد آریان
ICT Holding of Pasargad Financial Group
www.fanap.ir

درخواست برای ارائه پیشنهاد طرح سرویس تبدیل نوشتار به گفتار



مرکز مطالعات و تحقیقات شرکت فناپ

تاریخ تهیه سند: ۱۴۰۰/۰۴/۱۵

شماره ویرایش: ۰.۴

هدف از این طرح تدوین مشخصات سرویس مورد نیاز شرکت فناوری اطلاعات و ارتباطات پاسارگاد آریان (فناپ) در حوزه تبدیل نوشتار به گفتار^۱ بوده که می‌بایست توسط مجری یا مجریان احتمالی، طراحی و پیاده‌سازی گردد. در این راستا و در ادامه، خروجی طرح^۲، شرح نیازمندی‌ها و محدودیت زمانی اجرای طرح مشخص شده است. همچنین به مواردی که می‌بایست در تخمین بودجه مورد نیاز متقاضیان اجرای طرح لحاظ گردد، نیز اشاره شده است.

۱- کلیات

تبدیل متن به گفتار (یا به اختصار TTS) یکی از تکنولوژی‌هایی است که در سال‌های اخیر استفاده‌های زیادی از آن شده است و در بسیاری از اپلیکیشن‌ها، دستیارهای هوشمند مجازی، سرویس‌های آموزشی و ... از آن استفاده می‌شود. تکنولوژی تبدیل متن به گفتار در حقیقت نوعی فناوری کمکی است که می‌تواند یک متن دیجیتال را برای کاربر بلند خوانی کند. فناوری تبدیل متن به گفتار تنها با یک کلیک می‌تواند یک متن بلند را به صوتی تبدیل کند که حاوی محتوای آن متن است. فایل متنی می‌تواند یک فایل Word، یک صفحه وب و ... باشد. این تکنولوژی همچنین می‌تواند خواندن اطلاعات نمایشگر رایانه را برای افرادی که در خواندن به هر دلیلی مشکل دارند، امکان پذیر کند. تبدیل نوشتار به گفتار به کمک برنامه‌های تشخیص و شناسایی صدا انجام می‌شود.

۲- مشروح مسئله تحقیقاتی

تبدیل متن به گفتار سیستمی است کامپیوتری که قابلیت خواندن^۳ هرگونه نوشته‌ای را دارد. تبدیل متن به گفتار در حقیقت ارتباطی است دو طرفه میان کامپیوتر و انسان، که توانایی تشخیص کلمات و لحن متن نوشته شده و بیان آن در کامپیوتر را دارد. تشخیص نوشتار و بیان آن به صورت گفتاری مهم‌ترین وظیفه‌ی کامپیوتر در این فرآیند است. تولید آوا از روی یک متن، یکی از جنبه‌های مهم و کاربردی در پردازش سیگنال گفتار است.

به تبدیل متن به گفتار، سنتز گفتار^۴ نیز اطلاق می‌شود. سنتز گفتار، تولید مثل مصنوعی از بیان طبیعی است که توسط کامپیوتر پردازش و انجام می‌شود. متون نوشته شده به جای اینکه از حنجره‌ی یک انسان تولید شود، توسط رایانه تلفظ می‌شود و هر میزان که این موضوع از جنبه ماشینی بودن کمتری برخوردار بوده و تلفظ جملات به صورت نرم‌تر و روان‌تری بیان شود، از ارزشمندی بالاتری برخوردار خواهد بود.

تبدیل متن به گفتار می‌تواند راه‌حل مناسبی برای افرادی باشد که خواندن یک کتاب به صورت متنی برایشان کسل‌کننده است؛ اما اگر همین کتاب به صورت صوتی برایشان قرائت شود، حس بهتری دارند و تأثیرگذاری مطالب آن کتاب بیشتر خواهد بود. علاوه‌براین، تبدیل متن به گفتار، می‌تواند علتی برای کاهش سوانح رانندگی یا راهی برای یادگیری زبان توسط زبان آموزان باشد. سیستم‌های تبدیل متن به گفتار برای اولین بار برای کمک به افراد مبتلا به نقص (مانند نابینایان)، با ارائه‌ی صدایی که توسط یک رایانه گفته شده بود ایجاد شد.

تبدیل متن به گفتار مدت زمان زیادی است که در انواع نرم‌افزارها به کار برده می‌شود. برای مثال پس از نصب یک نرم‌افزار بر روی موبایل، کد تأییدیه برای کاربر ارسال می‌گردد. در صورتی که کاربر این کد را به صورت پیامک دریافت نکند، سرور

¹ Text to Speech

² Deliverables

³ Read

⁴ Speech Synthesis

نرم افزار، از طریق تماس صوتی، آن را برای کاربر می خواند. در حقیقت، کد تأییدیه تولید شده توسط سرور پردازش می شود و یک بار به صورت پیامکی ارسال می شود. در صورتی که پیامک به هر دلیلی به گوشی مخاطب نرسد، دستور تبدیل نوشتار (کد تأییدیه) به صوت به سرور داده می شود و با کاربر تماس گرفته شده و کد خوانده می شود.

تبدیل متن به گفتار را نباید با سیستم های پاسخ صوتی^۵ اشتباه گرفت. سیستم های پاسخ صوتی، جملاتی را که در یک پایگاه داده در نظر گرفته شده است و مملو از کلمات است، با یکدیگر تلفیق می کند. این گونه سیستم ها کاربرد بیشتری نسبت به سیستم های TTS دارند و برای مقاصد مختلفی استفاده می شوند.

۳- مسئله اصلی تحقیق و ضرورت آن

تبدیل متن به گفتار، جملات و عبارات را براساس حروف الفبا و واج های هر زبان تشکیل می دهد. سیستم تبدیل متن به گفتار، از لحاظ فوری قادر به خواندن هر رشته ای از کاراکترهای متن است تا جملات اولیه را تشکیل دهد. برای تولید گفتار در این نرم افزار باید از الگوریتم های مختلفی استفاده شود؛ به این دلیل که ذخیره سازی تمامی کلمات یک زبان غیرممکن است. تغییرات فرکانس صوت (زیر و بمی کلمات)، آهنگ یا دیرش^۶، شدت و ضعف بیان کلمات و درنگ^۷ عواملی نوایی هستند که در بیان هجا، واژه یا جملات اثر خود را نشان می دهند و می بایست در «سیستم تبدیل متن به گفتار» در نظر گرفته شوند تا تولید گفتار به نحوی صحیح صورت گیرد.

برخی از کاربردهای سرویس تبدیل نوشتار به گفتار شامل موارد ذیل می باشد:

- اطلاع رسانی سررسیدها (مانند تاریخ سررسید بیمه، اقساط، ساعت تاریخ پرواز و...)
- اطلاع رسانی اخبار و وقایع حساس
- ارسال پیامک به نابینایان و افراد مسن
- ارسال پیام صوتی به صورت انبوه
- راه اندازی انواع ماژول های خاص (کدخوان، منشی پیامک و غیره) به صورت پیام صوتی
- ارسال نوبت تعیین شده برای مقاصد مختلف (مطب، آرایشگاه و...) به صورت پیام صوتی
- ارسال پیام صوتی با ذکر نام افراد

همچنین می توان ماژول تبدیل متن به گفتار را در تمامی ماژول های موجود در سامانه پیام کوتاه استفاده کرد.

با توجه به اینکه ماژول تبدیل متن به گفتار قادر است یک وب سرویس کامل در اختیار توسعه دهندگان وب و موبایل قرار دهد، می تواند کاربردهای وسیع دیگری نیز داشته باشد. وب سرویس تبدیل نوشتار به گفتار به صورت RESTful API قابل استفاده بوده و دارای کاربردهای زیر می باشد:

- تولید گفتار طبیعی با کیفیت مطلوب و با سرعت بالا
- امکان تبدیل متن های مختلف به صوت بدون استفاده از نرم افزارهای ضبط صدا
- کاهش هزینه های تولید نرم افزارهای آموزشی و سامانه های اطلاع رسانی با حذف ضبط صدا
- خواندن نامه ها و متون الکترونیکی، اخبار و اطلاعیه های مختلف با استفاده از وب سرویس
- سیستم های تلفن گویا

⁵ Voice Response Systems

⁶ Duration

⁷ Pause

- سیستم‌های اطلاع رسانی مانند کیوسک‌ها، نوبت دهی بانک‌ها و ...
- قابل استفاده برای طراحان صفحات وب
- خواندن پیام‌های کوتاه (SMS) در گوشی‌های موبایل
- ارسال کد فعالسازی برای اپلیکیشن‌های موبایل به صورت پیام صوتی
- ارسال رمز یکبار مصرف از طریق پیام صوتی
- استفاده در انواع نرم افزارهای حسابداری، اتوماسیون اداری، کارتابل الکترونیک و...
- استفاده در سیستم‌های فروشگاه‌های آنلاین و تحت موبایل
- سایت‌های محتوا محور
- محتواهای سنگین از لحاظ نوشتاری
- تجهیز سیستم‌ها برای استفاده‌ی نابینایان

۴- تشریح نیازمندی‌های طرح

در این بخش ابتدا مشخصات اصلی طرح مطرح شده و سپس کاربردها، ورودی‌ها و خروجی‌های مورد انتظار و نحوه اعتبارسنجی طرح مشخص می‌گردد. در انتها نیز، قابلیت‌های فنی مورد نیاز در طرح تشریح شده است.

۴-۱- مشخصات اصلی

فناوری تبدیل متن به گفتار دارای صدایی بسیار هوشمند و طبیعی است. این فناوری می‌تواند صدایی بسیار نزدیک به صدای گوینده‌ی مورد نظر را تقلید کرده و با آن صدا صحبت کند. این تکنولوژی، قادر به گویا کردن تمامی اپلیکیشن‌ها و متون می‌باشد. علاوه بر این به دلیل هوشمندی، قادر است به جای اپراتور انسانی، به صورت مجازی پاسخگوی کاربران در هر سامانه‌ای باشد. سرویس تبدیل متن به گفتار بایستی دارای مشخصات زیر باشد:

- امکان تبدیل متن به گفتار فارسی و انگلیسی با صداهای طبیعی
- امکان شبیه سازی صداهای متنوع
- امکان تولید فرمت‌های مختلف صوتی (WAV, MP3, WMA, SWF, OGG, A-law, μ -law)
- پوشش Sample rate های 8000, 11025, 16000, 22050, 44100, 48000
- سیستم عامل Windows, Windows Server, Linux, Android, iOS
- ابزارهای Standard DLL, .NET DLL, Web Service, Windows Service, Application Service
- دارای SDK (لندروید و iOS) یا API برای فراخوانی آنلاین در هر برنامه‌ای که نیاز به خواندن متن با صداهای طبیعی است.
- امکان تعریف واژگان کاربری (Libraries, Documentation, Examples (C, C++, Delphi)
- پشتیبانی از فرمت‌های مختلف صوتی با انواع Sample Rate و bit Rate بصورت Multi-thread

۴-۲- کاربردهای سرویس

سرویس نوشتار به گفتار فارسی جهت استفاده برنامه‌نویسان و شرکت‌های ارائه‌دهنده سرویس‌های اطلاع‌رسانی مانند تلفن‌های گویا، سایت‌های خبری و تولیدکنندگان کتاب‌های الکترونیکی ارائه می‌شود. این محصول به صورت سرویس دهنده

API یا REST Service ارائه می‌گردد. با استفاده از API متن خوان روی یک سرور ویندوز، می‌توان متن‌های متفاوت را به سرویس فرستاده و به سرعت فایل صوتی را دریافت کرده و آن را ذخیره یا پخش نمود.

سرویس تبدیل نوشتار به گفتارهای کاربردهای متنوعی در فناوری اطلاعات و ارتباطات، سیستم‌های مخابراتی، وب‌سرویس‌ها، برنامه‌های کاربردی، سیستم‌های بانکی (ATM)، تلفن بانک و موبایل بانک و وب کیوسک، سیستم‌های امنیتی (اخطار دهنده‌های گویا و مانیتورینگ)، سیستم‌های مسیریابی (GPS)، نوبت دهی در بانک و آزمایشگاه‌ها و اعلام اطلاعات در فرودگاه‌ها، سیستم‌های خانگی (لوازم خانگی سخنگو و اخطار دهنده)، سیستم‌های اداری (پیگیری نامه‌ها)، سیستم‌های آموزش از راه دور و E-Learning، آموزش فارسی برای فارسی آموزان و کودکان، سی دی‌های مولتی مدیا، بازی‌های رایانه ای، گوش دادن به پیامک‌ها در تلفن همراه به هنگام رانندگی، گوش دادن به متون کتاب‌ها از طریق تلفن همراه و ... دارد. تبدیل متن به گفتار یکی از کاربردی ترین امکانات سامانه پیامک صوتی نیز هست. با استفاده از این قابلیت دیگر نیازی به ضبط صدا نیست؛ چرا که در کمترین زمان می‌توان فایل صوتی یک متن نوشته شده را دریافت نمود.

سیستم تبدیل متن به گفتار برای سیستم‌های مخابراتی از جمله سیستم‌های تلفن گویا (IVR&CTI)، پاسخگویی خودکار، مرکز تماس (Call center)، مدیریت ارتباط با مشتری (CRM)، تبلیغات تلفنی، خواندن ایمیل (Voice Email) یا خواندن پیامک (Voice SMS) مورد استفاده قرار می‌گیرد. از دیگر کاربردهای این سیستم می‌توان به سامانه‌های اطلاع‌رسانی مانند سامانه‌های شهرداری، راهنمایی و رانندگی، اطلاعات راه‌ها، پست، ۱۱۸ و غیره اشاره کرد. سرویس تبدیل متن به گفتار همچنین در ایجاد Website Player نقش دارد که برای تمامی کسب‌وکارها و دارندگان وب سایت‌ها کاربرد دارد. این سرویس بر روی وبسایت نصب می‌شود و کاربر می‌تواند تمامی متون وبسایت را با صدای طبیعی بشنود. با استفاده از قابلیت تبدیل متن به گفتار امکاناتی فراهم می‌شود که شامل موارد زیر می‌باشد:

• رابط کاربری جدید

برنامه‌ها و یا اپلیکیشن‌هایی که از تبدیل متن به گفتار استفاده می‌کنند، به نوعی یک رابط کاربری جدید برای کاربران خود ایجاد می‌کنند که راهی جایگزین برای به دست آوردن اطلاعات و دسترسی به آنچه است که می‌خواهند. این راه جایگزین برای تمامی کاربران، چه افرادی که سواد دارند و قادر به خواندن هستند و چه کسانی که سواد کمی دارند و یا بی سواد هستند، کاربردی است و کمک می‌کند تا همه‌ی کاربران بتوانند به راحتی از سرویس‌های آن‌ها استفاده کنند.

• یادگیری پیشرفته

کسب و کارهایی که دارای مطالب آموزشی هستند و یا اپلیکیشن‌هایی که محتوای آموزشی ارائه می‌دهند، با فعال کردن هم زمان دو قابلیت نمایش بصری و شنیداری به کمک تکنولوژی تبدیل متن به گفتار می‌توانند به بهبود درک و به یاد ماندن مطالب در کاربران کمک کنند و میزان تاثیرگذاری و یادگیری محتوای آموزشی‌شان را بالا ببرند.

• پویایی و آزادی

استفاده از تکنولوژی تبدیل متن به گفتار موجب می‌شود تا هر محتوای دیجیتالی به یک تجربه چند رسانه‌ای تبدیل شود. به همین دلیل افراد می‌توانند هم زمان در هنگام استفاده از این سرویس‌ها چندین کار را با هم انجام دهند. برای مثال یک فرد می‌تواند هم زمان به اخبار، مقالات یک وبلاگ و یا یک فایل پی دی اف گوش دهد و با این حال بتواند از خدمات یک کسب‌وکار نیز بهره‌مند شود.

سرویس تبدیل نوشتار به گفتار کاربردهای متعددی دارد که شامل موارد ذیل می‌شود:

- تبدیل آرشیه‌های متنی به صوتی
- به کارگیری متن به گفتار در تلفن گویا
- ماژول متن به گفتار (Server-side SDK): این امکان را به توسعه‌دهندگان و برنامه‌نویسان نرم‌افزارها می‌دهد تا به سادگی، قابلیت تبدیل متن به صوت طبیعی را به سامانه‌های خود اضافه نمایند.
- افزودن قابلیت نوشتار به گفتار در نرم‌افزارهای کامپیوتر یا موبایل: با استفاده از سرویس نوشتار به گفتار توسعه‌دهندگان نرم‌افزارها در تمامی پلتفرم‌ها و سیستم‌عامل‌ها می‌توانند قابلیت متن خوان را به محصولات خود اضافه نمایند. بنابراین برنامه‌نویسان نرم‌افزارهای ویندوز (Windows)، لینوکس (Linux)، مک (Mac, iOS) یا اندروید (Android) به راحتی می‌توانند از تولید گفتار استفاده کنند.
- نرم‌افزار متن به صدا در در نرم‌افزارهای کتاب‌خوان، پیامک‌خوان صوتی، مترجم صوتی و ...: این نرم‌افزار برای خواندن تمام متون گوشه‌ای از جمله پیامک (SMS) در انواع پیام‌رسان‌ها و پخش نام مخاطب هنگام تماس ورودی بصورت آنلاین با صدای طبیعی کاربرد دارد. با کمک سرویس نوشتار به گفتار می‌توان هر نوع متن فارسی را در تلفن همراه به صورت آنلاین به صدای طبیعی تبدیل نمود. علاوه بر این مبدل متن به صوت قادر است پیامک‌ها را به محض دریافت بخواند. از امکانات این نرم‌افزار می‌توان به سرویس تماس‌خوان، سرویس پیامک‌خوان صوتی و ویجت پیامک‌خوان صوتی اشاره نمود.
- تمامی نرم‌افزارهایی که نیاز به خواندن متن دارند مانند ایجاد پادکست‌ها (Podcast)، ویرایشگرها و مترجم‌های صوتی نیز می‌توانند به تکنولوژی سنتز گفتار مجهز گردند.
- امروزه کاربران تلفن همراه از برنامه‌های مختلف تبدیل متن به گفتار برای گوش دادن به مطالب موجود در کتاب‌های مورد علاقه خود به جای خواندن آن استفاده می‌کنند. از آن جایی که ممکن است خواندن کتاب یا خیره شدن به صفحه لپ‌تاپ، کامپیوتر یا گوشی‌های همراه، برای خواندن کتاب‌ها وقت‌گیر باشد، با استفاده از برنامه‌های تبدیل متن به گفتار می‌توان مطلب مورد علاقه خود را در هر زمان و هر مکانی با استفاده از هندزفری گوش داد. در این صورت می‌توان در وقت نیز صرفه جویی نمود.
- ربات متن به گفتار در شبکه‌های اجتماعی: کاربران توسط این ربات می‌توانند به سادگی هرچه تمام تر متون مورد نظر خود را به گفتار فارسی تبدیل نمایند و فایل صوتی آن را دریافت نمایند. طریقه کار ربات به این شکل است که ربات در ابتدا متن را دریافت کرده، آن را پردازش می‌کند، سپس با پردازش ویژگی‌ها و ایجاد صدایی شبیه به صدای انسان (نوع و جنسیت صدای قابل انتخاب توسط کاربر) آن را برای کاربر ارسال می‌کند.
- استفاده در سامانه‌های تلفنی، پرسش و پاسخ و خبری: بسیاری از سامانه‌های تلفن گویا یا IVR از سیستم‌های متن باز استریسک (Asterisk)، الاستیکس (Elastix)، فری سویچ (FreeSwitch) و نظایر این استفاده می‌کنند. افزودن قابلیت متن به گفتار به این سیستم‌های تلفنی می‌تواند کاربردها و کارکردهای جدیدی را معرفی نماید. پیاده‌سازی سامانه‌های اطلاع‌رسانی مانند هواشناسی، اخبار، پیگیری پرونده، اطلاعات بیمه، بورس و نظایر این با استفاده از متن خوان به راحتی امکان‌پذیر است.

تمامی سامانه‌های تلفنی (تلفن گویا، پیگیری شکایات، ساعت گویا، اطلاعات سهام و ...) مبتنی بر ویندوز یا لینوکس می‌توانند از سرویس تبدیل متن به صدای طبیعی برای خواندن متون متغیر، تبدیل اطلاعات متنی پایگاه داده به صوت و پخش آن به تماس‌گیرنده استفاده نمایند.

- **سنتر گفتار در سامانه‌های نوبت‌دهی و کیوسک‌ها:** استفاده از تبدیل نوشتار به گفتار در سامانه‌های نوبت‌دهی و کیوسک‌های اطلاع‌رسانی موجب افزایش کارایی و سهولت کاربری برای استفاده‌کنندگان می‌شود. در برخی از سامانه‌های نوبت‌دهی مانند نوبت‌دهی مطب‌ها و بیمارستان‌ها، نام مراجعه‌کننده نیز ذخیره می‌شود که سیستم تبدیل متن به صوت به راحتی می‌تواند در هنگام اعلام نوبت، اسم شخص را نیز بیان نماید.
- **استفاده در وب سایت‌ها برای گویا کردن مطالب سایت:** گویا کردن وب سایت‌ها به ویژه وبسایت‌های خبری و تحلیلی، علاوه بر این که برای نابینایان می‌تواند مفید باشد برای بسیاری از بازدیدکنندگان نیز سودمند است. زیرا بسیاری از افراد، شنیدن یک متن طولانی را به خواندن آن ترجیح می‌دهند. برای تبدیل نوشته‌های سایت به گفتار و پخش آن، برنامه‌نویسان سایت می‌توانند از سرویس تبدیل متن به صوت بهره ببرند.

۴-۳- ورودی سرویس

ورودی این سرویس دریافت هر گونه متن یا فایل متنی می‌باشد.

۴-۴- خروجی مورد انتظار سرویس

خروجی این سرویس صوت یا یک فایل صوتی می‌باشد. امکان تولید فرمت‌های مختلف صوتی (WAV, MP3, WMA,) (SWF, OGG, A-law, μ -law) از متن دریافت شده باید وجود داشته باشد.

۴-۵- نحوه اعتبارسنجی طرح

به منظور سنجش اعتبار طرح مورد نظر، دقت فایل صوتی استخراج شده از متن دریافتی سنجیده می‌شود. حداقل دقت مورد انتظار الگوریتم بکار رفته، ۹۹ درصد صحت در سطح جمله، بر روی مجموعه دادگان ارزیابی می‌باشد. مجموعه دادگان ارزیابی بر اساس نیازمندی‌های فناپ با جمع‌آوری متون و جملات مورد نظر، تامین می‌گردد. نمونه‌های مجموعه دادگان ارزیابی به تیم پیاده‌سازی داده می‌شود.

۴-۶- قابلیت‌های مورد انتظار سرویس

سرویس تبدیل متن به گفتار بایستی دارای قابلیت‌های زیر باشد:

- تبدیل نوشتار به گفتار فارسی با صدای طبیعی
- سرعت بالا در تبدیل متن به صدا (۱۰ جمله در هر ثانیه)
- قابلیت شناسایی و خواندن اعراب، اعداد، علائم نگارشی، و کلمات انگلیسی
- قابلیت تغییر لحن یا ایجاد وقفه در صدا با توجه به علائم نگارشی از قبیل علامت سوال، ویرگول و ...
- تشخیص کلمه‌های هم‌نگاره و کسره اضافه
- توانایی ذخیره کردن درخواست‌های متن به گفتار پرتکرار (Cache)
- امکان انتخاب جنسیت صدا، بلندی صدا، زیربومی صدا و سرعت خواندن
- Web API متن به گفتار با پشتیبانی از پروتکل REST

- پشتیبانی از انواع قالب‌های صوتی WAV, MP3, WMA, SWF, OGG, A-law, μ -law
- اختصاصی سازی واژگان: در کاربردهایی که تولید متن به صدا در موضوعات تخصصی مورد نیاز باشد، پایگاه داده با واژگان و اصطلاحات حوزه مورد نظر کاربر تطبیق داده می‌شود تا مبدل نوشتار به گفتار برای نوشته‌های تخصصی نیز صدایی صحیح و طبیعی تولید نماید.
- امکان خواندن متن فارسی- انگلیسی (متن ترکیبی)

۵- محدودیت زمانی اجرای طرح

این طرح پس از توافق طرفین می‌بایست حداکثر در یک بازه‌ی زمانی شش ماهه به اتمام رسد.

۶- تخمین بودجه اجرای طرح

تخمین بودجه اجرای طرح در قالب هزینه‌های نفر/ساعت نیروی انسانی، سربرار و تجهیزات جانبی مورد نیاز می‌بایست مشخص شود.