



## بررسی مدل‌های حمایت حاکمیتی از فناوری‌های خط و زبان در فضای مجازی در کشورهای منتخب

حبیب الله اصغری<sup>۱</sup>، سالار محتاج<sup>۲</sup>، سید امید فاطمی<sup>۳</sup>

<sup>۱</sup> پژوهشکده فناوری اطلاعات و ارتباطات جهاد دانشگاهی

habib.asghari@ictrc.ac.ir

<sup>۲</sup> پژوهشکده فناوری اطلاعات و ارتباطات جهاد دانشگاهی

salar.mohtaj@ictrc.ac.ir

<sup>۳</sup> دانشکده مهندسی برق و کامپیوتر، پردیس دانشکده‌های فنی، دانشگاه تهران

omid@fatemi.net

### چکیده

نقش دولت‌ها در حمایت از زبان ملی، ضرورتی انکار ناپذیر می‌باشد. کشورهای مختلف برنامه‌های متعددی با هدف صیانت از خط و زبان ملی و محلی خود تدوین نموده‌اند. این برنامه‌های حمایتی با هدف نهایی توسعه ابزارهای کاربردی در زمینه‌های مختلف مانند گردشگری و سلامت یا دیگر حوزه‌های عمومی تبیین شده‌اند. بررسی نقشه راه کشورهای مختلف در حمایت از خط و زبان می‌تواند نقشی اساسی در تبیین یک سند راهبردی و ملی در حمایت از خط و زبان فارسی ایفا نماید. از جمله حمایت‌های حاکمیتی می‌توان به ترجمه ماشینی به عنوان پدیده مرتبط با امنیت ملی اشاره نمود. تسهیل ترجمه اسناد در زبان‌های خارجی به زبان مرجع در برخی کشورها تا تسهیل استفاده از محتوای انگلیسی با کمک ترجمه به زبان‌های محلی در برخی کشورهای دیگر از مسائل مهم مرتبط با این اسناد راهبردی در کشورهاست. براین اساس، کشورهای مختلف برنامه‌های متفاوتی به منظور حمایت و تسهیل ایجاد پیکره‌های زبانی به منظور یادگیری عبارات معادل در ترجمه و همچنین توسعه ابزارهای پردازش زبانی تبیین نموده‌اند.

در این مقاله نحوه حمایت‌های حاکمیتی از مقوله پردازش زبان طبیعی و ترجمه ماشینی به منظور توسعه ابزارهای زیرساختی و محصولات کاربردی در چهار کشور هند، ایالات متحده آمریکا، ولز و اسپانیا و همچنین اتحادیه اروپا مورد بررسی قرار گرفته است.

### کلمات کلیدی

توسعه خط و زبان فارسی، مدل‌های حمایت از خط و زبان، پردازش زبان طبیعی، فناوری‌های زبانی

### ۱- مقدمه

طی فرآیند حمایت از پروژه‌های مرتبط با تکنولوژی زبان در TDIL، مجموعه‌ای از ابزارهای کاربردی با مشارکت موسسات تحقیقاتی و شرکت‌های تجاری فعال در این برنامه توسعه داده شده‌اند. این ابزارها شامل طیف وسیعی از برنامه‌های پایه‌ای تا کاربردی مرتبط با زبان مانند افزونه‌های برنامه MS Word تا ابزارهای غلط‌گیر املا، فونت‌ها و غیره هستند [۲]. برنامه TDIL به صورت سالانه اطلاعات مرتبط با محصولات، ابزارها، خدمات، فعالیت‌ها و دستاوردهای مرتبط با فناوری زبان هندی را به صورت پیوسته از سال ۲۰۰۰ از طریق انتشار یک نشریه در اختیار پژوهشگران و افراد مرتبط با این حوزه تحقیقاتی قرار می‌دهد.

#### ۲-۱- اهداف مدل حمایت هند

هدف اصلی این پروژه "توسعه ابزارها و فناوری‌های مرتبط با پردازش اطلاعات به منظور توسعه امکان تعامل انسان و ماشین بدون موانع زبانی، ایجاد دسترسی به منابع چندزبانه دانش، و ادغام آن‌ها برای توسعه محصولات و سرویس‌های نوآورانه" بیان شده است. همچنین ارتقای استانداردهای مرتبط با فناوری زبان از طریق ارتباط مؤثر با مجامع بین‌المللی و ملی مرتبط با استانداردهای مانند ایزو، یونیکد و کنسرسیوم جهانی وب<sup>۲</sup> در این برنامه پیش‌بینی شده است. همچنین مقرر گشته است محصولات (خروجی‌های) مرتبط با این برنامه بلند مدت شامل فونت‌ها، پیکره‌ها و دیکشنری‌های توسعه داده شده به صورت رایگان از طریق درگاه‌های مرتبط با پروژه در اختیار عموم قرار گیرند [۲].

#### ۲-۲- برنامه‌های مدل حمایت هند

برنامه جامع وزارت فناوری اطلاعات و ارتباطات هند شامل هفت بخش اصلی به شرح ذیل برای توسعه ابزارهای مرتبط با فناوری زبانی است:

الف - توسعه سیستم‌های ترجمه ماشینی:

زبان و خط هر کشور از جمله دارایی‌های فرهنگی و نیز میراث ارزشمند بشری به شمار می‌روند. در سال‌های اخیر، با رشد فناوری اطلاعات و همچنین با توجه به گسترش اینترنت، مرزهای تاثیرگذاری خط و زبان از مرزهای جغرافیایی فراتر رفته است. لذا دولت‌ها توجه ویژه‌ای به حمایت از خط و زبان در فضای سایبر معطوف داشته‌اند. توجه به برنامه و رویکردهای دولت‌های مختلف در حمایت از این میراث فرهنگی، می‌تواند برای تبیین راهبردهای حاکمیتی در کشور ما در حمایت و ارتقاء خط و زبان فارسی در محیط رایانه راهگشا باشد. فناوری زبانی در کشورهای مختلف به شکل‌های مختلفی مورد حمایت حاکمیت قرار گرفته‌اند. علاوه بر این در کنار حمایت از تولید ابزارها و زیرساخت‌های زبان‌شناسی رایانشی، راهبردهای ترویجی و حقوقی مرتبط با این موضوع نیز باید مورد بررسی قرار گیرد. در این مقاله مجموعه‌ای از مدل‌های حمایت از خط و زبان در محیط رایانه در کشورهای مختلف دنیا مورد بررسی قرار می‌گیرد. همچنین حوزه‌هایی را که می‌توان به عنوان درس‌های آموخته در تدوین راهبردهای حمایت از خط و زبان فارسی مورد استفاده قرار گیرد به بحث گزارده خواهند شد.

#### ۲-۲- مدل حمایتی هندوستان

دپارتمان الکترونیک و فناوری اطلاعات وزارت ارتباطات و فناوری اطلاعات کشور هندوستان به عنوان متولی فناوری اطلاعات در این کشور در سال ۱۹۹۱ پروژه توسعه فناوری برای زبان‌های هندی<sup>۱</sup> (TDIL) را آغاز نموده است. با توجه به تنوع فرهنگی و زبانی در کشور هندوستان (وجود ۲۲ زبان به رسمیت شناخته شده در قانون اساسی و ۱۲ خط نوشتاری متفاوت در این کشور)، وزارت ارتباطات و فناوری اطلاعات هندوستان پروژه TDIL را با هدف افزایش نفوذ فناوری اطلاعات در بین اقشار مختلف مردم و تسهیل امکان بهره‌مندی از منافع بیشمار فناوری اطلاعات پایه‌ریزی نموده است.

مالی نموده است. همچنین معمولاً خط مشی دولت ولز حمایت مالی از شرکتها پس از انجام پروژه است [۱۸].

در سال ۲۰۱۲ کارگروه فناوری زبان ولزی و رسانه‌های دیجیتال<sup>۱۰</sup> شامل نمایندگان موسسات آموزشی، مراکز تحقیقاتی، تهیه‌کنندگان تلویزیونی و خبرنگاران حوزه محتوای دیجیتال و فناوری، یک برنامه عملیاتی برای این کشور در حوزه خط و زبان ولزی ترتیب داده‌اند. حوزه‌های تمرکز این برنامه عملیاتی شامل موارد زیر می‌باشد:

- بازاریابی و آگاهی‌دهی: شامل افزایش آگاهی در رابطه با فناوری‌ها و محتوای موجود در حال حاضر
  - ایجاد انگیزه در شرکت‌های اصلی فعال در این حوزه
  - تشویق برای توسعه برنامه‌های کاربردی و خدمات دیجیتال مرتبط با حوزه زبان
  - تحریک انگیزه به منظور ایجاد، اشتراک‌گذاری و استفاده از محتوای تولید شده به زبان ولزی
  - حمایت از اقدامات مثبت در بخش دولتی و خصوصی
- نتیجه پایش و ارزیابی پیوسته این برنامه عملیاتی توسط این کارگروه به اطلاع وزیر آموزش و پرورش ولز می‌رسد.
- در ادامه برنامه‌ها و نیازمندی‌های بخش دولتی در هریک از ۵ حوزه تمرکز این برنامه عملیاتی به اختصار بیان شده است.

### ۳-۱- بازاریابی و آگاهی‌دهی

تحقیقات دولت ولز نشان می‌دهد علیرغم تلاش‌های دولت به منظور افزایش نفوذ خط و زبان ولزی در محیط دیجیتال مانند برنامه‌های کاربردی و محیط وب، مردم این کشور حدود یک درصد از زمان خود در محیط وب را به گشتن در سایت‌های ولزی زبان اختصاص می‌دهند. همچنین این آمار برای برنامه‌های تلویزیونی و رادیویی به ترتیب ۶ و ۱۵ درصد می‌باشد. با توجه به این آمار و ارقام، دولت یک هدف اساسی خود در این برنامه عملیاتی را افزایش آگاهی مردم از فناوری‌ها و محتوای در دسترس آن‌ها به زبان ولزی اختصاص داده است. بنابراین نقاط هدف دولت در این بخش شامل موارد زیر می‌باشد:

- سرمایه‌گذاری در پروژه‌های مرتبط با افزایش نرخ آگاهی مردم از نرم‌افزارها، سرویس‌های آنلاین، ابزارهای تولید محتوا به زبان ولزی
- توسعه یک استراتژی بازاریابی برای حمایت از نرم‌افزارها و سرویس‌های موجود به زبان ولزی
- همکاری با ذینفعانی مانند BBC و ... برای ترویج کاربری زبان ولزی
- تأکید بر ایجاد سهولت در انتخاب زبان ولزی در نرم‌افزارهای مربوطه

### ۳-۲- ایجاد انگیزه در شرکت‌های پیشرو

هدف دولت ولز تسهیل امکان مشارکت ولزی زبانان در فضای دیجیتال از طریق زبان ولزی است. دولت برای تحقق این امر، پشتیبانی هر چه بیشتر شرکت‌های سرآمد در فضای دیجیتال از زبان ولزی را ضروری می‌داند. به همین منظور دولت، شرکت‌های بزرگ از قبیل اپل، ماکروسافت، گوگل، آمازون را تشویق به ارائه واسط کاربری به زبان ولزی در سیستم‌های عامل، نرم‌افزارها و سرویس‌هایشان، و ایجاد امکان انتخاب ساده این زبان خواهد نمود. علاوه بر این دولت این کشور قصد تسهیل ایجاد محتواهایی از قبیل کتاب‌های الکترونیک به زبان ولزی با همکاری شرکت‌های مرتبط مانند Adobe به منظور حل مشکل پشتیبانی از الفبای این زبان در محیط‌های مربوطه می‌باشد. بنابراین نقاط هدف دولت در این بخش شامل موارد زیر می‌باشد:

- تشویق شرکت‌های مطرح فناوری به ایجاد واسط کاربری ولزی و ایجاد امکان انتخاب این زبان به صورت ساده و شفاف
- تشویق شرکت‌های مطرح فناوری در ایجاد تضمین در سهولت ایجاد محتوا به زبان ولزی در محیط سیستم و نرم‌افزارهای آن‌ها
- تشویق شرکت‌های مطرح فناوری در ایجاد اطمینان از اینکه کاربر ساده‌تر قادر به یافتن محتوای ولزی خواهد بود
- ارائه راهنمایی و مشاوره به شرکت‌های مطرح فناوری به منظور حمایت از زبان ولزی در محیط نرم‌افزارهای خود
- همکاری با دیگر جوامع دارای زبان اقلیت در اروپا در ارتباط با تشویق شرکت‌های بزرگ به پشتیبانی از زبان‌های بیشتر

### ۳-۳- تشویق توسعه برنامه‌های کاربردی و خدمات مرتبط

علاوه بر برنامه دولت جهت همکاری با شرکت‌های بزرگ به منظور پشتیبانی از زبان ولزی، دولت همچنین به توسعه و حمایت از برنامه‌های کاربردی کوچک‌تر برای توسعه زبان ولزی در

بخش توسعه ترجمه ماشینی در برنامه TDIL شامل توسعه فناوری‌های مرتبط در سه زیر بخش زیر می‌باشند:

- توسعه سیستم‌های ترجمه ماشینی انگلیسی به زبان‌های هندی: این بخش شامل توسعه سیستم ترجمه ماشینی برای ترجمه متون انگلیسی به شش زبان محلی هندی شامل Hindi, Marathi, Bangla, Oriya, Tamil, Urdu می‌باشد. با توجه به هدف اولیه این برنامه که فراهم سازی امکان استفاده مردم هند از محتوای فضای فناوری اطلاعات به زبان خود است، این بخش به منظور ترجمه محتوای انگلیسی به زبان‌های محلی با تشکیل یک کنسرسیوم ایجاد شده است. برای این منظور چهار تکنولوژی ترجمه ماشینی مختلف با یکدیگر ترکیب شده و سیستم جامعی تحت نام Anuvadak را تشکیل می‌دهند. در فاز اول، توسعه سیستم در حوزه خاص گردشگری مد نظر قرار گرفته شده است.
- توسعه سیستم‌های ترجمه ماشینی انگلیسی به زبان‌های هندی با استفاده از فناوری Angla-Bharti: این یک متدولوژی ترجمه ماشینی می‌باشد که با مشارکت چهار موسسه مختص ترجمه انگلیسی به زبان‌های هندی طراحی شده است.
- توسعه سیستم ترجمه ماشینی برای ترجمه زبان‌های هندی به زبان‌های هندی (Sampark): در این زیربخش از بخش توسعه سیستم‌های ترجمه ماشینی، توسعه سیستم‌های ترجمه ماشینی برای ترجمه زبان‌های محلی هندی به یکدیگر با استفاده از روشهای مبتنی بر قاعده<sup>۲</sup> و دیکشنری و سیستم‌های یادگیری ماشین آماری پیش‌بینی شده است.

ب - توسعه سیستم‌های دسترسی به اطلاعات چند زبانه<sup>۴</sup> :

این سیستم‌ها کاربران را قادر می‌سازند تا به اطلاعات موجود در زبان دیگری به جز زبان که پرس‌وجو را وارد نموده‌اند دسترسی پیدا کنند. به بیان روشن‌تر، سیستم قادر به ترجمه پرس‌وجوی کاربر و بازاریابی اطلاعات در زبان‌های دیگر به جز زبان مبدأ خواهد بود. این پروژه نیز به صورت کنسرسیوم و با مشارکت ۱۱ موسسه آغاز شده است و حوزه‌های گردشگری و سلامت و قابلیت پشتیبانی از ۹ زبان از حوزه‌های هدف آن هستند.

ج - توسعه یک سیستم بازشناسی و تحلیل اسناد:

این سیستم به منظور تبدیل اسناد اسکن شده به فایل‌های قابل ویرایش در زبان‌های هندی پیش‌بینی شده است (OCR). قابلیت کارکرد در ۱۵ زبان محلی برای این پروژه پیش‌بینی شده است.

د - توسعه سیستم تشخیص برخط دستخط<sup>۵</sup> :

در این پروژه هفت موسسه به منظور تشخیص برخط دستخط به منظور تسهیل فرآیند ورود متن و اطلاعات در هشت زبان هدف، همکاری می‌کنند.

ه - توسعه سیستم متن به گفتار برای زبان‌های هندی<sup>۶</sup> :

در این پروژه یک کنسرسیوم متشکل از چندین موسسه به منظور توسعه سیستم تبدیل متن به گفتار برای شش زبان هندی با هدف تسهیل مشارکت افراد با مشکلات بینایی تشکیل شده است.

و - توسعه سیستم تشخیص گفتار برای زبان‌های هندی<sup>۷</sup> :

ز - توسعه سیستم ترجمه ماشینی برای زبان سانسکریت:

### ۳-۳- مدل حمایتی ولز

طبق چشم‌انداز دولت ولز آن‌گونه که در برنامه استراتژیک این کشور برای سال‌های ۲۰۱۲ تا ۲۰۱۷ تحت عنوان "یک زبان زنده، یک زبان برای زندگی"<sup>۸</sup>، تبیین شده است، دولت یک مجموعه برنامه به منظور رشد این زبان و افزایش کارکرد آن در بخش‌های مختلف زندگی مردم این کشور ترتیب داده است. در واقع امکان دسترسی ولزی زبانان به فناوری‌هایی که استفاده از زبان ولزی را در جنبه‌های مختلف زندگی روزمره تسهیل نماید، جزء اهداف اولیه دولت در حمایت از تکنولوژی پردازش زبان می‌باشد [۱۸].

در راستای هدف تسهیل استفاده از زبان ولزی در امور روزمره، در سال‌های اخیر دولت این کشور تعاملاتی با شرکت‌های بین‌المللی مانند گوگل و ماکروسافت به منظور حمایت از این زبان در محیط‌هایی مانند سیستم عامل ویندوز، برنامه کاربردی MS Office و محیط Gmail انجام داده است. همچنین در همین راستا حمایت‌های مالی و فکری از پروژه‌های مرتبط با این حوزه مانند حمایت از پروژه‌های خطایاب، ترجمه ماشینی و تولید دیکشنری و تبدیل متن به گفتار به زبان ولزی توسط دولت انجام گرفته است.

در سال ۲۰۱۵ دولت این کشور از پروژه‌های مرتبط با توسعه فناوری زبان ولزی تا سقف ۵۰ هزار یورو برای پروژه‌های معمول و تا سقف ۱۰۰ هزار یورو برای پروژه‌های شاخص<sup>۹</sup> حمایت

گزارش سال ۱۹۹۳ کمیته صنعت و فناوری<sup>۱۲</sup> شورای هماهنگی علوم، فناوری و مهندسی<sup>۱۳</sup> دفتر ریاست جمهوری ایالات متحده با عنوان "فناوری ترجمه ماشینی، یک توان کلیدی در عصر اطلاعات" بر لزوم نقش‌پذیری دولت این کشور در توسعه سیستم‌های ترجمه ماشینی به عنوان عنصر شتاب دهنده در توسعه به منظور پیشی گرفتن این کشور در این صنعت از کشور ژاپن به عنوان کشور پیشگام در صنعت ترجمه در آن سال، تاکید دارد. همچنین در این سند بر اهمیت و کاربرد این فناوری در دیگر بخش‌های اقتصادی و اجتماعی مانند سلامت، علوم، فناوری و محیط زیست تاکید شده است. همچنین برنامه‌ریزی، ایجاد هماهنگی و پشتیبانی دولت از این بخش موجب تقویت موضع دفاعی این کشور در مقابل رقیبان عنوان شده است. پیشنهادات این کمیته به دولت به طور مشخص شامل: حمایت از پژوهش، حمایت از توسعه سیستم‌های موجود برای پاسخگویی به نیازمندی‌های دولت، ارزیابی عملکرد سیستم‌های موجود، حمایت از کارگاه‌های آموزشی مرتبط با ترجمه ماشینی و شناسایی نیازمندی‌های جدید بخش خصوصی و دولتی در رابطه با این موضوع می‌باشند [۹].

#### ۴-۱- پروژه GALE

یکی از مهم‌ترین و آخرین برنامه‌های دولت آمریکا در رابطه با توسعه سیستم‌های ترجمه ماشینی پروژه GALE<sup>۱۴</sup> در سال ۲۰۰۶ میلادی بوده است. این پروژه توسط سازمان پروژه‌های تحقیقاتی پیشرفته‌ی دفاعی<sup>۱۵</sup> (DARPA) تعریف شده و هدف از آن توسعه سیستمی است که بتواند به صورت خودکار شبکه‌های خبری، اسناد متنی و دیگر اشکال ارتباطی را ذخیره و ترجمه نماید و اطلاعات آن‌ها را به زبان انگلیسی در اختیار قرار دهد [۱]. تمرکز اولیه این پروژه پایش اخبار و ترجمه آن از دو زبان عربی و چینی بوده است. این پروژه دارای سه بخش اصلی به شرح زیر می‌باشد:

- سیستم تشخیص خودکار صوت برای پردازش اطلاعات گفتار
  - سیستم ترجمه ماشینی برای ترجمه اسناد غیر انگلیسی
  - سیستم واکنشی استخراج بخش‌های مهم اطلاعات مرتبط با پرس‌وجوی وارد شده
- هدف اولیه مطرح شده در این پروژه همکاری با سه گروه از محققان از SRI, IBM و BBN بوده است که طی دوره اجرای پروژه عملکرد هر گروه به صورت سالانه بررسی شده و در صورت لزوم تیم دارای کمترین کیفیت از روند پروژه حذف می‌شود. همچنین امکان توقف کلی پروژه پیش از اتمام آن در صورت عدم کسب موفقیت‌های برنامه‌ریزی شده، در طرح اولیه پیش‌بینی شده است [۵].
- هدف GALE رسیدن به دقت بین ۹۰ تا ۹۵ درصد در ترجمه چینی و عربی به انگلیسی در سال ۲۰۱۰ میلادی تبیین شده است. همچنین هدف تعریف شده برای این پروژه در سال اول رسیدن به دقت ۶۵ درصد در ترجمه صوت و دقت ۷۵ درصد در ترجمه متن برای دو زبان یاد شده عنوان شده است. تحقیقات DARPA پیش از آغاز پروژه نمایانگر دقت حدود ۵۵ درصدی سیستم‌های موجود در ترجمه اخبار بوده است. اما DARPA خواستار دقت بیان شده نه تنها در حوزه‌های رسمی مانند خبر، بلکه حوزه‌های محاوره‌ای مانند گپ‌های اینترنتی و مصاحبه‌های تلویزیونی بوده است.

#### ۵- مدل حمایتی اتحادیه اروپا

اتحادیه اروپا طی برنامه‌های ۷ ساله FP7<sup>۱۶</sup> و H2020 و CEF برنامه‌های جامعی به منظور حمایت از فناوری‌های مختلف از جمله فناوری زبان شناسی و حوزه‌های وابسته در نظر گرفته است. برنامه جامع FP7 مربوط به سال‌های ۲۰۰۷ تا ۲۰۱۳ میلادی و برنامه جامع H2020 مربوط به سال‌های ۲۰۱۴ تا ۲۰۲۰ میلادی می‌باشد. این حمایت‌ها شامل حوزه‌های مرتبط با زبان مانند ترجمه ماشینی، شناسایی صوت، تولید پیکره و منابع داده‌ای و ... می‌باشند. با توجه اهمیت از بین بردن مرز میان کشورها به عنوان هدف اولیه تشکیل اتحادیه اروپا، از میان برداشتن مرزهای زبانی میان کشورهای حاضر در این اتحادیه یکی از اهداف مهم کنونی این اتحادیه به شمار می‌رود. با توجه به این مهم، توسعه سیستم‌های ترجمه ماشینی و حمایت از تولید داده برای استفاده در سیستم‌های مرتبط با فناوری زبان از جمله مهم‌ترین اهداف و برنامه‌های این اتحادیه در حال حاضر هستند.

استراتژی تجزیه ارزش داده<sup>۱۷</sup> با هدف ایجاد ارزش افزوده تجاری و اجتماعی در داده بر اساس استفاده، مدیریت و استفاده مجدد هوشمند از منابع داده در اروپا از طریق ترکیب تحقیقات و نوآوری و پیاده‌سازی قوانین مورد نیاز شکل گرفته است. بر همین اساس از سال ۲۰۱۲ میلادی پرتالی برای مدیریت داده‌های موجود در اتحادیه اروپا در نظر گرفته شده است. داده‌های موجود در این پرتال دارای توسط دولت‌ها و بخش‌های مختلف اتحادیه تولید شده و برای استفاده در اختیار عموم قرار می‌گیرند.

آن‌ها علاقه‌مند است. بنابراین دولت در پی حمایت مالی از پروژه‌هایی که ولزی زبانان را قادر به دستیابی به اطلاعات به زبان ولزی با استفاده از آخرین فناوری‌های روز نماید، می‌باشد. توسعه دهندگان به منظور تولید و توسعه نرم افزارهای مرتبط با زبان ولزی نیاز به بهره بردن از طیف وسیعی از زیر ساخت‌ها و فناوری‌های مرتبط با این زبان را خواهند داشت. این فناوری‌ها شامل تبدیل متن به گفتار، تشخیص صوت، ترجمه ماشینی و دیگر فناوری‌های مرتبط می‌باشد. این فناوری‌ها زیرساختی لازم جهت تسهیل توسعه ابزارهای مرتبط با زبان ولزی را فراهم می‌نمایند.

همچنین در بلند مدت برای اطمینان از وجود تعداد کافی ولزی زبان قادر به توسعه نرم‌افزار برای توسعه زیرساخت فناوری زبانی، دولت قصد ایجاد انگیزه در جوانان در علوم مرتبط با کامپیوتر و برنامه نویسی را دارد. بنابراین نقاط هدف دولت در این بخش شامل موارد زیر می‌باشد:

- حمایت مالی از ابزارها و ابزارهای ولزی در پلتفرم‌های مختلف مانند رایانه، تبلت و گوشی‌های هوشمند
- حمایت مالی از پروژه‌های زیرساختی که منتج به توسعه ابزارها و نرم افزارهای ولزی می‌شوند. این زیر ساخت‌ها شامل توسعه سیستم‌های تبدیل متن به گفتار، تشخیص صوت، ترجمه ماشینی و ... می‌باشد.
- حمایت مالی از طرح‌های ایجاد مهارت برنامه‌نویسی در جوانان

#### ۳-۴- تحریک انگیزه جهت ایجاد، اشتراک گذاری و استفاده از محتوای تولید شده به زبان ولزی

دولت ولز یک عامل اساسی در رابطه با بقاء و نشاط و چالاکی زبان ولزی را میزان محتوای تولید شده و به اشتراک گذاشته شده می‌داند. از این رو دولت این کشور در پی تسهیل شرایط تولید و به اشتراک گذاری محتوا به زبان ولزی در محیط وب می‌باشد. با توجه به اهمیت نسل جوان در این مهم، دولت این کشور حمایت مالی از ابتکارات و تشویق نسل جوان در چنین فعالیت‌هایی را در برنامه دارد. در واقع دولت این کشور در پی ایجاد اعتماد و مهارت در افراد در تمامی رده‌های سنی به منظور ایجاد محتوای آنلاین به زبان ولزی است. نقاط هدف دولت در این بخش شامل موارد زیر می‌باشد:

۱. امکان سهولت تولید محتوای دیجیتال برای ولزی زبانان در وبلاگ‌ها، وبسایت‌ها و شبکه‌های اجتماعی با حمایت مالی از تولید ابزارهای مرتبط مانند وبلاگ و غیره
۲. ایجاد مهارت و اعتماد در ولزی زبانان به منظور تولید محتوای آنلاین از طریق حمایت مالی از ابتکارات مرتبط با ایجاد مهارت در تولید محتوا، تشویق و حمایت از فعالیت‌های جوانان به منظور تولید محتوا، حمایت از طرح‌های ویکیدیا در آموزش به ولزی زبانان و افزایش سطح آگاهی با اضافه نمودن مطالب به ویکی پدیا، کسب اطمینان از در دستور کار قرار گرفتن زبان ولزی در بخش دیجیتال دولت ولز
۳. تشویق و حمایت از سازمان‌ها و مجتمع‌های مرتبط با زبان ولزی به منظور افزایش نرخ استفاده از فناوری و ایجاد برنامه‌های آموزشی

#### ۳-۵- حمایت از اقدامات مثبت بخش دولتی و خصوصی

این بخش از برنامه دولت برای حمایت از زبان ولزی شامل دو بخش استفاده از این زبان در محیط کار و خدمات عمومی می‌باشد. در رابطه با استفاده از زبان ولزی در محیط کار دولت در پی اجبار سازمان‌ها و شرکت‌ها برای تسهیل استفاده کارمندان از زبان ولزی از طریق ایجاد نرم‌افزارها و فناوری‌های مرتبط می‌باشد. همچنین در رابطه با خدمات عمومی مانند خدمات دولتی، دولت بخش‌های مختلف را مکلف به ارائه خدمات به زبان ولزی و تسهیل استفاده از این زبان در محیط‌های مرتبط نموده است.

#### ۴- مدل حمایتی ایالات متحده آمریکا

دولت ایالات متحده آمریکا برای بیش از چهار دهه نقش حمایت و توسعه سیستم‌های ترجمه ماشینی را ایفا نموده است. تحقیقات نشان می‌دهند پیش از نامه معروف کمیته آلیاک<sup>۱۱</sup> در سال ۱۹۶۶ دولت این کشور بیش از ۱۳ میلیون دلار صرف توسعه چنین سیستم‌هایی نموده است [۳]. اما نامه کمیته آلیاک در این سال در ارتباط با ارزیابی سیستم‌های ترجمه ماشینی و زبان شناسی رایانشی در آمریکا در آن سال‌ها تردیدهایی در رابطه با کیفیت و توسعه این سیستم‌ها وارد نمود و منجر به کاهش شدید بودجه دولتی در سال‌های پس از آن در این بخش و تقریباً قطع حمایت دولتی از توسعه این سیستم‌ها در آمریکا شد. پس از این انتشار این سند، بخش خصوصی تبدیل به بخش پیشگام در توسعه سیستم‌های ترجمه ماشینی در آمریکا شد. با توجه به جریان جنگ سرد بین ایالات متحده و اتحاد جماهیر شوروی در آن سال‌ها، تمرکز دولت آمریکا در توسعه سیستم‌های ترجمه ماشینی - روسی بوده است.

## ۵-۱- برنامه CEF

رئیس جمهور، وزارت گردشگری و وزارت بهداشت و خدمات عمومی می‌باشد. این کمیته راهبری در اولین گام از این برنامه، یک کمیته از خبرگان به منظور تنظیم یک سند اولیه به عنوان سند پشتیبان برنامه اصلی ایجاد نموده است. کمیته خبرگان شامل فرهنگستان سلطنتی زبان اسپانیایی، موسسه سروانتس، کتابخانه ملی اسپانیا، آکادمی سلطنتی مهندسی، متخصصان گروه پردازش زبان طبیعی، شرکت‌های بزرگ اسپانیایی فعال در حوزه IT، گروه‌های کسب و کار تخصصی در حوزه زبان و ترجمه، کمیته سیاست‌گذاری استفاده مجدد از اطلاعات بخش دولتی، کمیته اصلاحات نظام اداری اسپانیا و گروه همکاری بین‌المللی با آمریکای لاتین تشکیل شده است. گروه خبرگان همچنین تحلیلی جامع از نقاط ضعف، قوت، فرصت‌ها و تهدیدهای پیش روی دولت در حوزه پردازش زبان را ارائه نموده اند [۱۰].

## ۶-۱- اهداف برنامه

اهداف دولت از توسعه این برنامه جامع در سه بخش به شرح زیر تبیین شده است:

۱. توسعه زیر ساخت‌های زبان شناسی: شامل توسعه و رشد صنعت پردازش زبان برای زبان اسپانیایی با ایجاد و فراهم آوری بسترهای زیر ساختی و داده ای لازم، کاهش فاصله بین ابزارهای زیر ساختی برای زبان اسپانیایی و انگلیسی از منظر کیفیت، کمیت و دسترس پذیری، کسب اطمینان از رایگان یا ارزان بودن ابزارهای زیرساختی برای استفاده توسط بخش های دولتی و پژوهشی، جلوگیری از دوباره کاری در توسعه ابزار و ایجاد استانداردها و قوانین لازم.
۲. توسعه صنایع مرتبط با زبان شناسی در بخش IT، شامل تسهیل انتقال دانش به بخش صنعتی و افزایش اطلاعات در رابطه با ابزارهای موجود و کمک به فعالیت بین المللی بخش‌های صنعتی فعال در این حوزه در این کشور می‌باشد. این امر با ایجاد تعاملات بین‌المللی به خصوص با کشورهای آمریکای لاتین (با توجه به نزدیکی زبانی) و کمک به بازاریابی شرکت‌های فعال در این بخش انجام می‌پذیرد.
۳. مدیریت بخش عمومی به عنوان محرک بخش زبان شناسی: این هدف شامل ایجاد یک سکوی واحد در دولت در بخش پردازش زبان و ترجمه ماشینی و تسهیل فرآیند تولید داده با استفاده از سیاست‌گذاری‌های استفاده مجدد از اطلاعات بخش دولتی می‌باشد.

## ۶-۲- مراحل انجام برنامه

با توجه به اهداف تبیین شده در این برنامه و وجود محورهای مختلف در این برنامه جامع شامل دولت، پشتیبانی از توسعه زیرساخت‌های زبانی، ارتقای صنعت پردازش زبان، بخش عمومی به عنوان محرک صنعت پردازش زبان، گام‌هایی برای رسیدن به اهداف تبیین شده در هریک از محورهای تبیین شده پیش بینی شده است.

الف - محور دولت

دلیل وجود این محور ایجاد مکانیزمی به منظور هماهنگی، کنترل و پایش اجزای برنامه و افراد و سازمان‌های مرتبط با آن است. همچنین گام‌های پیش‌روی دولت از ابتدا تا انتهای برنامه به ترتیب زیر تبیین شده است.

- ایجاد کمیته راهبری: به عنوان رکن اصلی تولید این برنامه
- ایجاد کمیته خبرگان که نقش مشاوره فنی برنامه را ایفا خواهد نمود.
- تشریح برنامه عملیاتی
- ارزیابی میزان اجرا و کیفیت اجرای برنامه
- هماهنگی با دیگر دولت ها با هدف تسهیل هماهنگی با دیگر برنامه‌های موجود مانند برنامه‌های اتحادیه اروپا و ایجاد امکان فعالیت بین‌المللی شرکت های فعال به خصوص در آمریکای لاتین

ب - محور توسعه ابزارهای زیر ساختی

این محور به توسعه ابزارهای زیر ساختی، منابع داده‌ای و پیکرها به منظور انجام فرآیندهای مرتبط با پردازش زبان طبیعی می‌پردازد. بنابراین دولت در طول این برنامه قصد حمایت از تولید چنین ابزارها و داده‌هایی و در اختیار قرار دادن آن‌ها برای پژوهشگران و شرکت‌های فعال در این حوزه را خواهد داشت. گام‌های رسیدن به این مهم در برنامه به شرح زیر عنوان شده‌اند:

- انتخاب استانداردهای فنی تعامل متقابل، سیاست‌های صدور مجوز و مکانیزم‌های لازم به منظور حمایت از تولید منابع داده‌ای لازم
- جذب یا توسعه ابزارهای لازم به منظور ایجاد یک زیر ساخت زبانی لازم
- اجرای طرحی به منظور توسعه زیر ساخت زبانی مورد نظر. تهیه لیست ابزارهای زیر ساختی موجود و ارزیابی کمی و کیفی ابزارهای زیر ساختی
- تسهیل دسترسی عمومی به ابزارهای زیر ساختی

ج - محور ارتقای صنعت پردازش زبان

برنامه CEF<sup>۱۸</sup> یک برنامه جامع به منظور توسعه یک شبکه کارآمد، مطمئن و مؤثر در اتحادیه اروپا در زمینه انرژی، حمل و نقل و سرویس‌های دیجیتال می‌باشد. در بخش مخابرات هدف از این پروژه تسهیل تعاملات بین مرزی برای بخش دولتی، کسب‌وکارها و شهروندان اروپایی با توسعه یک زیرساخت سرویس‌های دیجیتال<sup>۱۹</sup> و شبکه‌های پهن باند است. بیش از یک میلیارد یورو برای توسعه سرویس‌های زیرساختی دیجیتال برای سال‌های ۲۰۱۴ تا ۲۰۲۰ میلادی در نظر گرفته شده است. یکی از بخش‌های در نظر گرفته شده در این برنامه به عنوان زیرساخت‌های توسعه دیجیتال اتحادیه، توسعه سیستم‌های ترجمه ماشینی و منابع زبانی تخصصی شامل ابزارها و واسط‌های برنامه‌نویسی ضروری در راستای توسعه سرویس‌های دیجیتال در یک محیط چند زبانی در نظر گرفته شده است.

## ۵-۲- برنامه H2020

علاوه بر برنامه CEF، توسعه فناوری زبانی در برنامه جامع H2020 اتحادیه اروپا نیز دیده شده است. این برنامه بزرگ‌ترین برنامه تحقیقاتی اتحادیه اروپا با بودجه ای نزدیک به ۸۰ میلیارد یورو، برای توسعه علم و فناوری در این اتحادیه می‌باشد. این برنامه جنبه‌های مختلف علوم مانند علوم فضایی، فناوری اطلاعات، سلامت و ... را در بر می‌گیرد. شاخه فناوری اطلاعات و ارتباطات در بخش علوم نیازمند توسعه سریع‌تر برای در اختیار گرفتن بازارهای آینده جهانی در این برنامه قرار گرفته است. برنامه عملیاتی سال ۲۰۱۴ و ۲۰۱۵ از برنامه جامع H2020 شامل ۳۹ بخش در توسعه فناوری اطلاعات و ارتباطات است که حوزه‌های مرتبط با زبان در ۳ بخش از این برنامه عملیاتی مورد اشاره قرار گرفته‌اند. اولین حوزه مرتبط با توسعه فناوری زبان در برنامه عملیاتی سال ۲۰۱۴ و ۲۰۱۵ میلادی مرتبط با توسعه سیستم داده‌ی باز می‌باشد که در آن بر توسعه سیستم فعلی تأکید شده است. بخش دوم مرتبط با توسعه پردازش زبان طبیعی از برنامه عملیاتی توسعه فناوری اطلاعات در اتحادیه اروپا، مربوط به تحقیقات در حوزه کلان داده‌ها با محوریت توسعه داده‌کاوی و فهم زبانی می‌باشد. سومین و مهم‌ترین بخش مرتبط با توسعه ابزارهای زبانی از این برنامه عملیاتی تحت عنوان از بین بردن موانع زبانی در این برنامه گنجانده شده است. در این بخش مشابه برنامه جامع CEF، هدف ایجاد یک محیط یکپارچه دیجیتال با از میان برداشتن موانع زبانی است. سیستم های مد نظر در این بخش باید قابلیت مدیریت حجم بالای از داده، قابلیت پردازش انواع مختلف متون و دقت بالا در ترجمه به زبان‌های مختلف اروپایی را دارا باشند. همچنین توسعه سیستم برای ۲۱ زبان اروپایی با اولویت زبان‌های فاقد سیستم ترجمه ماشینی یا دارای سیستم‌های ضعیف در این برنامه مد نظر بوده است. در چشم انداز این برنامه تمرکز بر حوزه‌های زبانی دارای کیفیت پایین و حوزه‌هایی که توسعه سیستم‌های موجود نیازمند هزینه هنگفت یا داده‌های حجیمی است که در حال حاضر امکان فراهم نمودن آن‌ها نیست، بیان شده است. برای این منظور به طور کلی مبلغ ۱۵ میلیون یورو در سال ۲۰۱۴ میلادی در نظر گرفته شده است [۶].

همچنین برنامه عملیاتی سال ۲۰۱۶ و ۲۰۱۷ میلادی از برنامه جامع H2020 توسعه فناوری زبانی را در سه بخش مد نظر قرار داده است. توسعه فناوری زبانی در این برنامه شامل توسعه سیستم‌های بین زبانی و بین بخشی برای تجمیع داده‌ها و آزمایشات می‌باشد. در این بخش توسعه یک سیستم هوشمند و یکپارچه به منظور اشتراک‌گذاری دانش و اطلاعات در تمامی اتحادیه اروپا به عنوان هدف مد نظر قرار گرفته است. روشن است توسعه سیستم‌های پردازش زبانی از پیش نیازهای توسعه چنین سیستمی می‌باشد [۷].

## ۶- مدل حمایتی اسپانیا

دولت اسپانیا برنامه جامع ترویج فناوری زبان را در سال ۲۰۱۶ آغاز می نماید و قرار است در این برنامه طی پنج سال ۹۰ میلیون یورو برای ارتقای فناوری زبان در این کشور هزینه نماید. دولت اهداف کلی از این برنامه را افزایش کیفیت و کمیت ابزارهای زیرساختی مرتبط با پردازش زبان برای زبان اسپانیایی و دیگر زبان‌های رسمی این کشور، ارتقای صنعت زبان شناسی با تسهیل انتقال دانش از دانشگاه به صنعت، کمک به فعالیت بین‌المللی شرکت‌های فعال در این حوزه و بهبود کیفیت و ظرفیت خدمات عمومی با بهره‌گیری از فناوری‌های مرتبط با پردازش زبان و ترجمه ماشینی بیان نموده است. همچنین دولت نقش استانداردسازی و انتشار منابع داده‌ای تولید شده در طول این برنامه را برای خود در نظر گرفته است. با توجه به برنامه‌های اخیر دولت اسپانیا در جلوگیری از تعدد سازمان‌های ارائه دهنده خدمات مشابه، یک هدف اساسی دولت در این برنامه جلوگیری از دوباره کاری و افزایش هم‌افزایی از طریق ترکیب پروژه‌های موجود بیان شده است. وزارت ارتباطات اسپانیا به عنوان متولی برنامه، برای تبیین این برنامه یک کمیته راهبری با همکاری سازمان‌های دولتی این کشور ترتیب داده است. این کمیته راهبری شامل دبیرخانه بنگاه‌های کوچک و متوسط، وزارت علوم و تحقیقات، وزارت فرهنگ، وزارت آموزش، فرهنگ و ورزش، بخش آمریکای لاتین وزارت امور خارجه، معاون

- هم افزایی با دیگر بخش‌های این برنامه جامع به منظور ارتقای سطح پردازش زبان طبیعی و ترجمه ماشینی
- توجه ویژه برای کسب تجربه برای پروژه‌های آینده
- اولویت اولیه تعریف پروژه در بخش‌های سلامت، گردشگری و آموزش در سال ابتدایی برنامه و تعریف پروژه در زمینه‌های دیگر مانند خدمات شهروندی، امنیت و ... در سال‌های آتی

#### ۷- نتیجه‌گیری

در این تحقیق نحوه حمایت‌های حاکمیتی از پروژه‌های مرتبط با پردازش زبان طبیعی در کشورهایی از نقاط مختلف دنیا مورد بررسی قرار گرفت. مجموعه راهبردها و سیاست‌های توسعه خط و زبان در این مقاله شامل بررسی چگونگی حمایت‌های مالی از پروژه‌ها و محصولات مرتبط با فناوری زبان طبیعی در چند کشور منتخب از جمله هند، ایالات متحده آمریکا، ولز، اسپانیا و اتحادیه اروپا می‌باشد.

گرچه تمامی راهبردهای مطرح شده در اسناد فوق، قابل اعمال در سیاست‌های حاکمیتی ایران در گسترش خط و زبان در رایاسپهر نیستند، ولی برخی از اهداف و راهبردهای استخراج شده از این اسناد می‌توانند در توسعه اسناد ملی به منظور ارتقاء خط و زبان فارسی در محیط رایانه مورد استفاده قرار گیرند. از جمله این اهداف می‌توان به موارد ذیل اشاره نمود:

- فراهم سازی امکان استفاده جامعه از محتوای فضای فناوری اطلاعات به زبان فارسی
- تسهیل امکان مشارکت فارسی زبانان در فضای دیجیتال از طریق زبان فارسی
- کاهش فاصله بین ابزارهای زیر ساختی برای زبان فارسی و دیگر زبان‌ها از منظر کیفیت، کمیت و دسترس پذیری
- تلاش در جهت بقاء و نشاط و چالاکی زبان فارسی از طریق افزایش میزان محتوای تولید شده و به اشتراک گذاشته

همچنین از مجموعه راهبردهای پیشنهادی می‌تواند به موارد زیر اشاره نمود:

- ارائه رایگان محصولات زیرساختی مرتبط با فناوری‌های زبانی (مانند فونت‌ها، پیکره‌ها و ابزارهای زیرساختی) از طریق درگاه‌های مرتبط و جلوگیری از دوباره کاری در توسعه ابزار
- ایجاد یک سکوی واحد در بخش پردازش زبان در دولت
- گسترش سیاست استفاده مجدد از اطلاعات بخش دولتی به صورت دسترسی باز به منظور توسعه ابزارها و پیکره‌های زبانی و تسهیل فرآیند تولید داده‌های معیار
- تدوین استانداردهای مورد نیاز برای محصولات خط و زبان فارسی
- پیوند محصولات خط و زبان با یکی از فعالیت‌های کلیدی در کشور (از جمله اپورتورهای تلفن همراه، حوزه گردشگری، حوزه سلامت)
- حمایت از محصولات و برنامه‌های کاربردی که فارسی زبانان را قادر به دستیابی سریع‌تر به اطلاعات به زبان فارسی نموده و کار آنان را در جنبه‌های مختلف زندگی روزمره تسهیل نماید
- تعامل با شرکت‌های بین‌المللی به منظور حمایت از زبان فارسی در محیط‌هایی مانند سیستم‌های عامل، برنامه کاربردی MS Office و محیط رایانامه‌ها، در راستای هدف تسهیل استفاده از زبان فارسی در امور روزمره
- ارائه راهنمایی و مشاوره به شرکت‌های بین‌المللی مطرح فناوری به منظور حمایت از زبان فارسی در محیط نرم‌افزارهای خود
- توسعه یک استراتژی بازاریابی برای حمایت از نرم افزارها و سرویس‌های موجود به زبان فارسی
- ایجاد محتوای آموزشی برخط به زبان فارسی و همچنین بهبود دروس دانشگاهی مرتبط
- تسهیل آموزش زبان فارسی مبتنی بر فناوری
- ایجاد انجمن‌های علمی مرتبط
- برگزاری کنفرانس‌ها و نمایشگاه‌های ملی و بین‌المللی
- تبیین یک طرح حمایتی فراتر از این طرح به منظور کسب اطمینان از برقراری آن ورای طول عمر برنامه
- تشویق و حمایت از فعالیت‌های تولید، اشتراک‌گذاری و استفاده از محتوای فارسی در فضای سایر

یکی از چالش‌ها و مسائل موجود در ایران در خصوص تدوین این اسناد، تعدد مراکز سیاست‌گذاری جهت تدوین و تبیین راهبردهای حمایت از خط و زبان فارسی در محیط رایانه است. بر این اساس باید با در نظر گرفتن تمامی نقش آفرینان این حوزه به تدوین یک سند

هدف از وجود این محور لزوم شناساندن بخش پردازش زبان و ترجمه ماشینی به مقامات دولتی و بخش‌های مولد صنعت و همچنین تربیت نیروهای خبره در این زمینه به منظور کسب اطمینان از حفظ روند رو به رشد این فناوری در کشور عنوان شده است. برای نیل به این هدف، گام‌های زیر پیش‌بینی شده است:

- توسعه یک برنامه به منظور افزایش آگاهی در رابطه با این فناوری
- برنامه‌ریزی اقداماتی با هدف بهبود آگاهی و آموزش نیروهای خبره در این حوزه با مشارکت دولت‌های دیگر، به خصوص آمریکای لاتین
- اجرای آموزش از طریق ایجاد محتوای آموزشی برخط، بهبود دروس دانشگاهی مرتبط، کمک‌های مالی برای پروژه‌های دانشگاهی و غیره
- افزایش آگاهی از طریق برگزاری کنفرانس‌های پایه برای شرکت‌های متوسط و حرفه‌ای، برگزاری انجمن‌ها و نمایشگاه‌های ملی و بین‌المللی
- همچنین در این محور حمایت دولت از فعالیت بین‌المللی شرکت‌ها و ایجاد بازارهای بین‌المللی برای آنها، با توجه به نرخ بالای اسپانیایی زبانان در دنیا دیده شده است. برای رسیدن به این هدف، گام‌های زیر در این برنامه پیش‌بینی شده است:
- توسعه یک برنامه برای بین‌المللی سازی
- افزایش رایزنی با موسسات فعال در آمریکای لاتین
- تشویق به توسعه ابزارهای زیرساختی و منابع داده‌ای برای گویش‌های دیگر اسپانیایی
- شناسایی نمایشگاه‌ها و کنفرانس‌های مرتبط به منظور شناساندن هرچه بیشتر شرکت‌های اسپانیایی فعال در این حوزه

#### د - محور بخش عمومی

به عنوان محرک صنعت پردازش زبان، با توجه به بهره‌مندی بیشمار بخش عمومی از مزایای سیستم‌های پردازش زبان طبیعی و ترجمه ماشینی شامل افزایش ظرفیت، کیفیت و خودکارسازی خدمات عمومی، کاهش هزینه در انتقالات بین زبانی، افزایش کیفیت خدمات به معلولین و ... گام‌های زیر در این محور در نظر گرفته شده است.

- توسعه یک طرح به منظور ایجاد یک سکوی استفاده از پردازش زبان طبیعی و ترجمه ماشینی در دولت
- تبیین یک طرح حمایتی فراتر از این طرح به منظور کسب اطمینان از برقراری آن ورای طول عمر این طرح
- ایجاد یک پلتفرم مشترک برای استفاده از خدمات پردازش زبان طبیعی و ترجمه ماشینی در بخش دولتی

همچنین در این محور، پتانسیل بالای بخش دولتی در تولید اطلاعات به عنوان منابع داده‌ای مورد نیاز در پردازش زبان طبیعی مد نظر قرار گرفته است. به عنوان نمونه در این سند به نام افراد، نام سازمان‌ها، نام شرکت‌ها، متون دسته‌بندی شده و موارد مشابه آن به عنوان اطلاعات موجود در بخش دولتی که کاربردهای بیشماری در پردازش زبان طبیعی نیز دارند اشاره شده است. با توجه به این موارد، گام‌های پیش رو برای رسیدن به اهداف تعیین شده در این محور به شرح زیر عنوان شده است:

- معرفی و گسترش سیاست استفاده مجدد از اطلاعات بخش دولتی در بدنه دولت
- انتخاب استانداردهای فنی و سیاست‌های صدور مجوز لازم برای حمایت از اطلاعات فردی و دولتی
- توسعه و تهیه ابزارهای لازم به منظور پردازش منابع داده دولتی
- شناسایی منابع مهم و کارای بخش دولتی که قابلیت تبدیل به پیکره و منابع داده‌ای برای پردازش زبان را دارند

#### ه - محور پروژه‌های شاخص پردازش زبان طبیعی

هدف از وجود این محور معرفی پروژه‌های شاخص عمومی با ضریب تأثیرگذاری بالا در خدمات عمومی و اجتماعی به منظور تکمیل زنجیره ارزش پروژه‌های زیر ساختی می‌باشد. به بیان روشن‌تر، این پروژه‌ها به شکل خطوط قطع‌کننده پروژه‌های زیر ساختی و افقی تعریف شده در بخش‌های قبل در نظر گرفته شده‌اند. وجود چنین برنامه‌هایی باعث بهره‌مندی بخش عمومی از پردازش زبان طبیعی و ترجمه ماشینی و موجب بهبود کیفیت خدمات این پروژه‌ها در نظر گرفته شده است. برای رسیدن به این مهم تعریف تعداد محدودی پروژه با بکارگیری پردازش زبان طبیعی و ترجمه ماشینی با میزان تأثیر اجتماعی بالا و دارای شرایط زیر مد نظر قرار گرفته است:

- کسب اطمینان از ارتباط پروژه با بدنه اصلی برنامه
- دارای تأثیر اجتماعی و اقتصادی بالا
- تولید منابع دارای قابلیت استفاده مجدد. پرهیز از تولید منابع اختصاصی و خاص که قابلیت استفاده مجدد در دیگر پروژه‌های مرتبط را ندارند.

جامع اقدام نمود. همچنین پایش و ارزیابی پیوسته برنامه مذکور نیز یکی از نکات مهمی است که باید مد نظر قرار گیرد.

<sup>۱۸</sup> Connecting Europe Facility

<sup>۱۹</sup> Digital Service Infrastructures (DSIs)

#### سیاسگذاری

این مقاله با حمایت و همکاری گروه پژوهشی مدل‌های کسب و کار پژوهشکده فناوری اطلاعات جهاد دانشگاهی تالیف شده است. نویسندگان مقاله مراتب قدردانی خود را از اعضای محترم ستاد توسعه فناوری اطلاعات معاونت علمی ریاست جمهوری جهت هدایت‌های راهبردی در تدوین این مقاله ابراز می‌دارند.

#### مراجع

- [1] Cohen J., *the GALE project: A description and an update*. In Automatic Speech Recognition & Understanding, 2007.
- [2] Kumar A., Goyal V., *TDIL Programme: A Government Initiative*, an International Journal of Engineering Sciences, 2010.
- [3] Pedtke TR., *US government support and use of machine translation: current status*. MT Summit (1997).
- [4] Vikas O., *Language Technology Development in India*. Ministry of Information Technology, New Delhi, India. 2001.
- [5] *DARPA challenge - build the ultimate speech translation machine*
- [6] *Horizon 2020 Work Programme 2014 – 2015*, European Commission Decision C (2015)2453 of 17 April 2015
- [7] *Horizon 2020 Work Programme 2016 – 2017*, European Commission Decision C (2015)6776 of 13 October 2015.
- [8] *LT2013 Status and Potential of the European Language Technology Market*, January 2013.
- [9] *Machine Translation Technology: A Potential Key to the Information Age*, Report of the FCCSET Committee on Industry and Technology, 1993.
- [10] *Plan de Impulso de las Tecnologías del Lenguaje*, October 2015.
- [11] *Welsh-language Technology and Digital Media Action Plan*, <http://www.wales.gov.uk/welshlanguage>

#### زیر نویس‌ها

- <sup>۱</sup> Technology Development for Indian Languages
- <sup>۲</sup> World-wide-Web consortium (W3C)
- <sup>۳</sup> Rule based approaches
- <sup>۴</sup> Cross-lingual Information Access (CLIA)
- <sup>۵</sup> On-line handwriting recognition system (OHWR)
- <sup>۶</sup> Text to Speech System (TTS)
- <sup>۷</sup> Automatic Speech Recognition (ASR)
- <sup>۸</sup> A living language: a language for living
- <sup>۹</sup> Ambitious Infrastructural Projects
- <sup>۱۰</sup> The Welsh-language Technology and Digital Media Group
- <sup>۱۱</sup> Automatic Language Processing Advisory Committee (ALPAC)
- <sup>۱۲</sup> Committee on Industry and Technology
- <sup>۱۳</sup> Federal Coordinating Council for Science, Engineering and Technology
- <sup>۱۴</sup> Global Autonomous Language Exploitation
- <sup>۱۵</sup> Defense Advanced Research Projects Agency (DARPA)
- <sup>۱۶</sup> 7th Framework Program (FP7)
- <sup>۱۷</sup> The Data Value Chain Strategy