



معرفی سیستم تبدیل متن فارسی به زبان اشاره فارسی برای ناشنوایان با استفاده از آنتولوژی

محمد لطافت ابریکوه^۱، عبدالرضا رسولی کناری^۲، مرضیه دیوانی^۳

۱- دانشجوی کارشناسی ارشد کامپیوتر- گرایش نرم افزار- دانشگاه صنعتی قم

۲- استادیار دانشکده برق و کامپیوتر - دانشگاه صنعتی قم

۳- دانشجوی کارشناسی ارشد کامپیوتر- گرایش نرم افزار- دانشگاه صنعتی قم

چکیده

مهم ترین عامل ارتباط بین انسان ها، توانمندی در برقراری ارتباط از روش های مختلف مانند گفتار، نوشتار و زبان اشاره می باشد. ناتوانی افراد کم شنوا و ناشنوا در برقراری ارتباط صمیمانه، هم چنین نبود آموزش مرسوم با توجه به نیازهای جامعه، باعث کنارگذاری این افراد شده است. این پژوهش به دنبال معرفی سیستمی می باشد که متن فارسی را دریافت کرده و بعد از انجام پردازش زبان طبیعی لازم، آن را به زبان اشاره فارسی تبدیل کند. برای انجام این مهم با چالش های متعددی چون کمبود حرکات اشاره به نسبت تعداد کلمات موجود در زبان فارسی و هم چنین مشکلات زبان اشاره فارسی مواجه هستیم، برای همین منظور معماری سیستم های مشابه قابل استفاده نمی باشد. بنابراین معماری پیشنهاد شده است که با استفاده از ماژول تبدیل متن ورودی به دنیای ناشنوایان با استفاده از آنتولوژی طراحی شده برای آن ها، نرخ کارایی سیستم مترجم متن فارسی به زبان اشاره فارسی را بهبود بخشیده و باعث افزایش دقت و کارایی در سیستم مترجم شده است.

واژگان کلیدی: زبان اشاره فارسی، آنتولوژی، تبدیل، پردازش زبان طبیعی، ناشنوایان

مقدمه

جریان اصلی جامعه به طور مرسوم جامعه کم شنوایی و ناشنوایی را کنار گذاشته است. ناتوانی افراد کم شنوا در برقراری ارتباط صمیمانه، همچنین نبود آموزش مرسوم به اندازه کافی با توجه به نیازهای جامعه در این کنارگذاری تأثیر به سزایی دارند. بنابراین آن‌ها خود را در موقعیت تقریباً نامساعدی در تمامی حوزه‌های زندگی یافته‌اند. مطالعات نشان می‌دهد که ۹۰٪ افراد کم شنوا و ناشنوا در جهان هرگز به مدرسه نرفته‌اند، بنابراین از سواد کمتری برخوردار هستند. از این رو آن‌ها نیاز به توسعه ابزاری دارند که نه تنها به آن‌ها در برقراری ارتباط با جریان اصلی جامعه کمک کند، بلکه آن‌ها را قادر به یادگیری سازد. زبان مهم‌ترین وسیله ارتباطی در جوامع اجتماعی است و شاید بتوان گفت بدون زبان یا وسیله‌ای که به رد و بدل کردن پیام کمک کند جوامع انسانی شکل نمی‌گرفت. در واقع برای در کنار هم ماندن نیاز است تا ارتباط برقرار شود و خواسته‌ها، احساسات، ترس‌ها و شاید آرزوها به دیگری منتقل شود. پس با این نگاه می‌توان ارتباطی ناگسستنی بین زبان (به معنای وسیله برقراری ارتباط) و شکل‌گیری جامعه قائل شد (طرقی، ۱۳۸۹). زبان عبارت است از تمام نمادها، علائم و وسائلی از هر نوع، که به وسیله حواس مختلف، بین افراد جامعه رابطه برقرار می‌کند. با توجه به این تعریف غیر از زبان گفتاری، زبان اشاره و... نیز برای تفهیم و ایجاد رابطه بین افراد جامعه وجود دارد (روح‌المینی، ۱۳۸۲).

زبان اشاره

زبان اشاره نوعی از زبان است که در آن ناشنویان به جای استفاده از سیستم صوتی و یا به عبارتی گفتار، از دست‌ها برای سخن گفتن بهره می‌برند. در واقع این دست‌ها هستند که به جای زبان، سخن می‌گویند (Stoke, 1976). امروزه مکتب‌های آموزشی مختلفی برای ناشنویان و کم شنویان وجود دارد. گروهی بر این باورند که تضعیف زبان اشاره و تأکید بر روش‌های دیگر، مخالفت با نظام طبیعی ذهنی کودکان و تأکید بر روش‌های مصنوعی است و گروه دیگر معتقدند تأکید بر زبان اشاره انگیزه کودکان را برای یادگیری زبان گفتاری کاهش می‌دهد. برخی دیگر نیز بر این باورند که در آموزش به ناشنویان نقش دست‌ها را نمی‌توان نادیده گرفت. از این رو زبان اشاره مهم‌ترین ابزار آموزشی ناشنویان محسوب می‌شود (سیاوشی، ۱۳۸۴). زبان اشاره، یک زبان جهانی نیست، و از هر کشوری به کشور دیگر متفاوت است. زبان اشاره فارسی از گویشی به گویش دیگر متفاوت است. زبان اشاره فارسی علی‌رغم افزایش تعداد ناشنویان و استفاده از آن، پیشرفت آن‌چنانی نداشته و دایره کلمات موجود در آن بسیار محدود است (سیاوشی، ۱۳۸۵).

زبان ناشنویان ایرانی

به طور کلی زبان ناشنویان ایرانی را می‌توان به سه گروه اصلی تقسیم کرد:

- اشاره- اشاره: که در این شیوه تنها از حرکات دست و اشاره مطلق استفاده می‌شود. در این شیوه لب‌خوانی صورت نمی‌گیرد.

- شفاهی - شفاهی: در این شیوه به هیچ عنوان از اشاره و علامت‌های دستی استفاده نمی‌شود و تنها لب‌خوانی صورت می‌گیرد. که در این شیوه تمرکز زیادی نیاز است که ناشنوا باید تمام مدت به دهان گوینده نگاه کند تا از روند کلام دور نیفتد. اگر این مکالمه طولانی باشد برای ناشنوایان بسیار خسته‌کننده خواهد شد. در این شیوه معمولاً انتقال پیام از سوی ناشنوا هم با کندی و سختی صورت می‌گیرد.
- مکتب آموزشی باغچه‌بان: که در این شیوه بین دو شیوه‌ی بالا، یعنی لب‌خوانی و اشاره تعادلی صورت می‌گیرد که در آن سرعت کلام تا حدی افزایش می‌یابد و علاوه بر اینکه منظور منتقل می‌شود از افتادن ناشنوایان در دام اشاره‌ی مطلق که در واقع گفتار را در آن‌ها به شدت تضعیف می‌کند هم در امان می‌ماند (طرقی، ۱۳۸۹).

با توجه به تکنولوژی‌های اطلاعاتی و ارتباطی جدید توجه به نیازهای ناشنوایان برای برقراری ارتباط با سایر افراد جامعه امر بسیار مهمی است که لازم است بدان توجه شود. سیستم مترجم یک زبان به زبان دیگر که کار ترجمه را به صورت خودکار انجام می‌دهد می‌تواند بسیار مفید باشد. یک سیستم مترجم ماشینی، نیازمند در نظر گرفتن و تطبیق شدن با هر زبانی است. در این سیستم، نیاز است تا جزئیات مربوط به هر دو زبان مبدا و مقصد در نظر گرفته شود تا بتوانیم بهترین رویکرد ممکن را در خصوص ترجمه‌ی ماشینی و همچنین تکنیک‌های پردازش زبان طبیعی انتخاب کنیم.

پردازش زبان طبیعی

پردازش زبان طبیعی عبارتست از: استفاده از کامپیوتر به منظور پردازش زبان گفتاری و نوشتاری که با استفاده از آن می‌توان به ترجمه زبان‌ها پرداخت؛ از صفحات وب و بانک‌های اطلاعاتی نوشتاری جهت پاسخ دادن به پرسش‌ها استفاده کرد و با دستگاه‌ها و ربات‌های ساخته شده ارتباط برقرار کرد (Shamsfard, 1388). هدف اصلی از پردازش زبان طبیعی ایجاد تئوری محاسباتی از زبان با استفاده از الگوریتم‌ها و ساختارهای داده‌ی موجود در کامپیوتر است. بدیهی است که در راستای تحقق این هدف، نیاز به دانشی وسیعی از زبان است و علاوه بر محققان علوم کامپیوتر، نیاز به دانش زبان‌شناسان نیز در این حوزه می‌باشد (AleAhmad, 1385). برای درک زبان طبیعی به اشکال مختلفی از دانش نیاز است که در ذیل به آن‌ها اشاره می‌شود (بذرافکن، ۱۳۹۳):

- دانش صوت‌شناسی: ارتباط لغات با اصوات را نشان می‌دهد.
- دانش مورفولوژی: روش ساخت عبارات متنی مختلف از روی ریشه کلمات را بیان می‌کند. مورفولوژی در لغت به معنی ریخت‌شناسی است که شاخه‌ای از علم زیست‌شناسی می‌باشد، که به مطالعه شکل ظاهری و ویژگی‌های ساختاری خاص می‌پردازد. تجزیه و تحلیل مورفولوژیک یک روش برای شناسایی و تحقیق بر روی مجموعه‌ای از روابط یا ترکیب‌های ممکن برای یک مسئله پیچیده می‌باشد که از این طریق انجام یک سری فرآیند تکراری آنالیز و ترکیب داده‌های مسئله فرآیند حل مسئله را انجام می‌دهد.

- دانش نحوی: نحوه ساخت جملات را با استفاده از ترکیب کلمات مختلف نشان می‌دهد. دانش ساختاری هر کلمه موجود در جمله را تعیین می‌نماید.
- دانش معنایی: معنی هر لغت چه می‌باشد و چطور این معانی ترکیب می‌شوند تا معنی هر جمله تشکیل شود. این دانش، مطالعه معنی جملات را بدون توجه به متن شرح می‌دهد.
- دانش عملی: جملات چگونه در شرایط مختلف استفاده می‌شوند و هر نوع استفاده از جمله چه تاثیری در تفسیر جمله دارد.
- دانش سخن: این دانش نشان می‌دهد که معنی و وجود جمله قبلی در تفسیر جملات بعدی چه تاثیری دارد. این اطلاعات خصوصا برای رفع ابهام، تصحیح ضمائر و ارجاعات بسیار مفید است.
- دانش جهانی: شامل دانش عمومی در مورد ساختارهای اجتماعی، تاثیر عوامل مختلف و موثر بر یکدیگر و دنیایی که کاربران زبان باید از آن برای تفسیر جملات و متون مستقل به زبان استفاده کنند می‌باشد.
- درک شرایط: معنی جمله مبتنی بر شرایط زمان تولید جمله است. اما معنی یک کلمه به دانشی که در آن زبان، از کلمه وجود دارد برمی‌گردد.

آنتولوژی

آنتولوژی ریشه در فلسفه دارد و مبدا آن را ارسطو می‌دانند. در فلسفه، آنتولوژی شاخه‌ای از علم است که به بررسی موجودات و روابط میان آن‌ها می‌پردازد. مفهوم آنتولوژی در وب معنایی کمی متفاوت از فلسفه است. آنتولوژی از دو واژه *Onto* به معنی هستی و *Logia* به معنی مطالعه به وجود آمده است و در کل معنی هستی‌شناسی دارد. آنتولوژی در وب معنایی واژه‌ها و ارتباطات بین آن‌ها در دامنه‌ای که استفاده می‌شود را نشان می‌دهد. عناصر اصلی تشکیل دهنده آنتولوژی عبارتند از: مفاهیم، ارتباطات بین آن‌ها و خصوصیات آن‌ها (AleAhmad, 1385). آنتولوژی را خیلی کوتاه می‌توان به این صورت تعریف کرد: آنتولوژی، مشخص کردن و تعریف یک مفهوم‌سازی است. به عبارت دیگر آنتولوژی، ارتباطات بین مفاهیم در اسناد و دنیای واقعی را مشخص می‌کند که با این کار اسناد مربوطه توسط ماشین قابل پردازش و فهم می‌شود و اشتراک‌گذاری اطلاعات بین عامل‌ها را تسهیل می‌کند. در واقع می‌توان گفت (امیری و همکاران، ۱۳۸۸):

- Vocabulary + Structure = Taxonomy
- Taxonomy + Relationship, Constraints and Rules = Ontology
- Ontology + Instance = Knowledgebase

عناصر مختلف آنتولوژی شامل موارد زیر است (دوخانی، ۱۳۹۲):

- نمونه‌ها: اشیای ابتدای یا نمونه‌های اولیه که مبنای ساخت آنتولوژی می‌باشد.
- کلاس‌ها: مجموعه‌ها، مفاهیم، انواع اشیای یا انواع چیزهای ابتدای تشکیل دهنده نمونه.
- خاصیت‌ها: جنبه‌ها، ویژگی‌ها، خصوصیات یا پارامترهایی که آن اشیای (کلاس مربوط به اشیای) می‌توانند داشته باشند.



- ارتباطات: روش‌هایی که در آن کلاس‌ها و نمونه‌هایی می‌توانند با یکدیگر ارتباط داشته باشند.
- جملات تابعی: ساختارهای پیچیده‌ای که از یک ارتباط مشخص شکل می‌گیرند، که می‌توانند به جای یک کلمه یا کلمه‌ی خاص یک مورد استفاده قرار گیرند.
- قیدها: توضیحاتی که به صورت رسمی بیان می‌شوند تا مشخص کنند که چه چیزی باید صحیح باشد تا این که یک حکم به عنوان ورودی مورد پذیرش قرار گیرد.
- قوانین: جملاتی که به صورت اگر-آنگاه استنتاجات منطقی که می‌توانند از یک حکم به صورت خاصی به دست آیند را توصیف می‌کنند.
- قواعد کلی: احکام و قوانینی که در یک شکل منطقی با هم یک تئوری را که آنتولوژی در دامنه‌ی یک کاربرد شرح می‌دهد، را تشکیل می‌دهند.
- رخدادها: تغییر ارتباطات یا خاصیت‌ها

زبان‌های بیان آنتولوژی

دو نوع زبان برای ایجاد آنتولوژی وجود دارد (امیری و همکاران، ۱۳۸۸):

- زبان‌هایی که براساس گراف عمل می‌کنند. مانند:

RDF, Semantic network, UML, Topic map

- زبان‌هایی که براساس منطق گزاره‌ها عمل می‌کنند. مانند:

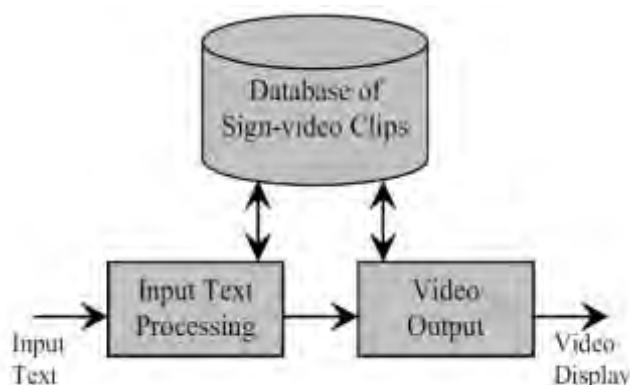
Rules(RuleML,LP/Prolog), Description Logic(OIL, DAML + OIL , OWL), First Order Logic

پیشینه پژوهش

در سال ۲۰۰۹، عده‌ای از دانشجویان بنگلادشی یک سیستم مترجم برای تبدیل متن بنگلادشی به زبان اشاره بنگال ایجاد کردند. آن‌ها برای کاهش موانع ارتباطی بین افرادی که دارای اختلال شنوایی هستند با افراد عادی جامعه، نرم‌افزار مترجمی را توسعه دادند که متن ورودی را به زبان بنگال دریافت کرده و به طور خودکار به زبان اشاره بنگالی (BASL) که توسط انجمن بنگالی در سراسر بنگال غربی ایجاد شده است تبدیل می‌کند.

فرهنگ لغت، که مبنای کار مترجم است، در حدود هزار کلمه را شامل می‌شود که این کلمات، عمدتاً ساده هستند و از کتاب‌های درسی کودکان جمع‌آوری شده‌اند. روش کار این مترجم اینگونه است که کلمه یا جمله ساده در بخش مربوطه وارد شده و پردازش‌ها و جابجایی‌های لازم در ساختار کلمه یا جمله صورت می‌گیرد و در نهایت اشاره معادل آن کلمه یا جمله نمایش داده می‌شود.

معماری این مترجم در شکل ۱ آورده شده است



شکل ۱ معماری تبدیل متن به زبان اشاره بنگلادشی

معماری این مترجم شامل سه ماژول اصلی است که عبارتند از:

۱- ماژول پردازش متن ورودی

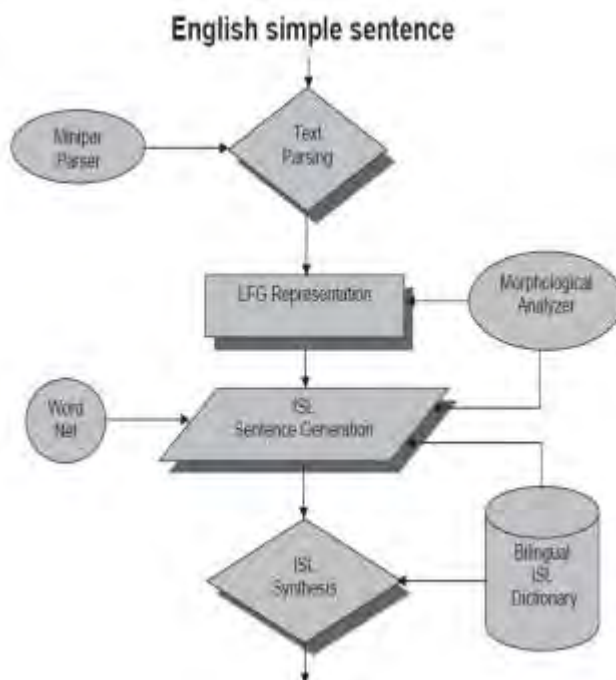
۲- ماژول پایگاه داده‌هایی از کلیپ‌های تصویری

۳- ماژول خروجی ویدئویی

این سیستم برای نمایش خروجی برنامه که زبان اشاره بنگال می‌باشد از ویدئوهای از قبل تهیه شده استفاده می‌کند. این مترجم به دنبال استانداردسازی BASL در تمام مناطق بنگلادش می‌باشد (Sarkar and Datta, 2009).

در سال ۲۰۱۰، محققان دانشگاه IIT هندوستان، نمونه اولیه سیستم تبدیل متن انگلیسی به زبان اشاره هندی را معرفی کردند. این سیستم می‌توانست با بهبود و شکل‌گیری برای مردم ناشنای هند کمک شایانی باشد. این سیستم جمله انگلیسی را در قالب متن دریافت کرده و پس از تجزیه و تحلیل نحوی و ایجاد ساختار مناسب به زبان اشاره هندی مربوط به متن ورودی تبدیل می‌کرد. از آن جایی که زبان اشاره هندی هیچ شکل نوشتاری ندارد خروجی این برنامه در قالب ویدئوهای از قبل آماده شده (ضبط شده) نمایش داده می‌شود. سیستم جاری برای انتقال متن به زبان اشاره هندی از سیستم انتقال مبتنی بر معماری بهره می‌برد.

همان گونه که بیان شده این سیستم نمونه اولیه از یک سیستم بزرگ می باشد و تنها جملات ساده انگلیسی را به عنوان ورودی دریافت کرده و زبان اشاره هندی آن را نمایش می دهد. این سیستم از معماری خاصی برای انجام این کار استفاده می کند که در شکل ۲ آن را مشاهده می کنید.



شکل ۲ معماری سیستم تبدیل متن هندی به زبان اشاره هندی

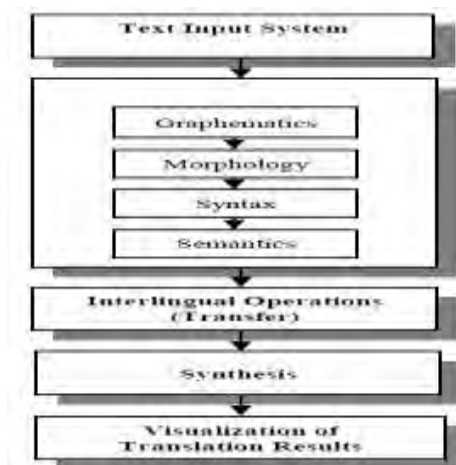
این معماری دارای چهار ماژول اصلی و ضروری می باشد:

- ماژول تجزیه و پیش پردازش متن ورودی
- ماژول نمایش LFG (نمایش گرامر واژگان)
- ماژول تولید جمله در دنیای ناشنویان
- ماژول نمایش جمله به زبان اشاره

اولین ماژول وظیفه تجزیه جمله ساده ورودی به اجزای تشکیل دهنده آن را دارا می باشد، لازم به ذکر است که تمام ورودی های این مترجم جملات ساده می باشد که تنها حاوی یک فعل می باشند. سپس با استفاده از اجزای تشکیل دهنده ساختار درختی جمله ورودی را نمایش می دهد. دومین ماژول، وظیفه بررسی روابط گرامری جمله ورودی را دارد، تا ساختار داخلی جمله از نظر گرامری بررسی شود. این امر برای شناسایی ساختار جمله می باشد تا در ماژول بعدی بتوان روی آن عملیات خاصی انجام داد. سومین ماژول، وظیفه تبدیل ساختار جمله ورودی به زبان اشاره هندی را دارد. آخرین ماژول، وظیفه نمایش خروجی مرحله قبل که تبدیل شده جمله ورودی به جمله در زبان اشاره هندی می باشد را داراست.

با استناد به ارزیابی این سیستم توسط کارشناسان زبان اشاره هندی، کارایی این سیستم قابل قبول ارزیابی شده است (A.Basu et al, 2010).

در سال ۲۰۱۳، جمعی از دانشجویان دانشگاه Novosibirsk در روسیه سیستم تبدیل و ترجمه زبان روسی به زبان اشاره روسی را برای کسانی که دارای اختلالات شدید شنوایی هستند را معرفی نمود. در آن زمان گام‌های اولیه این سیستم بررسی شد و در مرحله بررسی قرار گرفت. ساختار کلی سیستم مورد نظر آن‌ها که ترجمه کامپیوتری زبان روسی را به



زبان اشاره روسی (RSA) انجام می‌داد به صورت شکل ۳ تعریف شده بود:

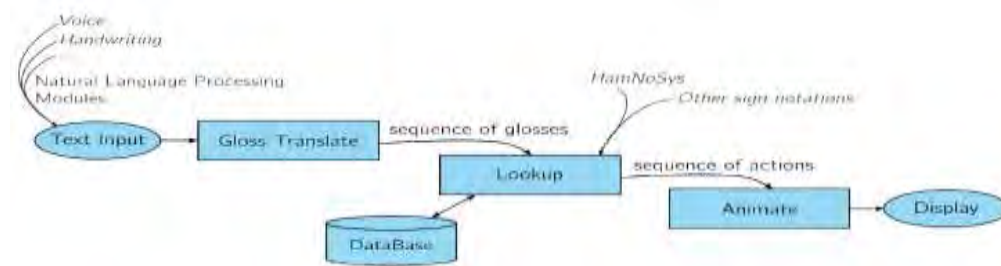
شکل ۳ معماری سیستم تبدیل زبان روسی به زبان اشاره روسی

ماژول‌های اصلی این ساختار عبارتند از:

- تفسیر متن منبع
- انتقال به زبان مقصد
- ترکیب زبان اشاره
- مشاهده نتایج با استفاده از زبان اشاره

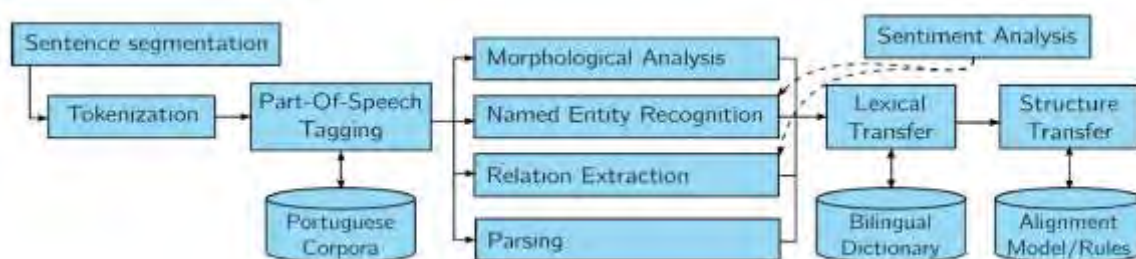
در مرحله تحلیل گرافیکی عناصر متن ورودی از هم جداسازی می‌شود. در مرحله آنالیز مورفولوژیکی از روش‌های مبتنی بر دیکشنری گرامری زبان روسی که حاوی کلمات کلیدی این زبان است بهره می‌برد. در تحلیل نحوی، ارتباط بین کلمات تعریف و بررسی می‌شود که در این تحلیل از تحلیل معنایی نیز استفاده می‌شود. روش اصلی تجزیه و تحلیل نحوی و معنایی آن‌ها، درخت وابستگی متن ورودی می‌باشد. در ماژول بعدی با استفاده از درخت وابستگی مرحله قبل، متن ورودی به زبان اشاره روسی (RSA) تبدیل می‌شود. در ماژول بعدی ساختار جمله در زبان اشاره روسی مورد بررسی قرار می‌گیرد. در نهایت توسط ماژول تصویرسازی به خروجی زبان اشاره روسی تبدیل می‌شود (Mikail et al, 2013).

در اکتبر سال ۲۰۱۴، Ines Almeida در کشور پرتغال یک سیستم مبتنی بر آواتار از زبان پرتغالی به زبان اشاره پرتغالی را معرفی کرد و چالش‌های آن را مورد بررسی قرار داد. این سیستم اولین و شاید مهم‌ترین قدم در آن کشور در خصوص مطالعات زبانی برای زبان اشاره پرتغالی بود. هدف اصلی سیستم ترجمه زبان پرتغالی به زبان اشاره پرتغالی این بود که سیستم قادر شود کلمات و جملات کامل در متن را دریافت کرده و با استفاده از ماژول‌های پردازش آن‌ها را بررسی و در نهایت به زبان اشاره پرتغالی نمایش دهد. معماری کلی این سیستم در شکل ۴ نمایش داده شده است.



شکل ۴ معماری سیستم تبدیل زبان پرتغالی به زبان اشاره پرتغالی

ماژول‌های اصلی این معماری عبارتند از: Gloss Translate، Lookup و Animate. این سیستم ابتدا توسط ماژول Input، متن ورودی را دریافت و پس از پردازش‌های لازم بر روی متن ورودی، متن پردازش شده به ماژول Gloss Translate ارسال می‌شود. در این ماژول که خود دارای ساختار می‌باشد که این



ساختار در شکل ۵ نمایش داده شده است.

شکل ۵ ساختار ماژول Gloss Translate

این ماژول‌ها برای تجزیه و تحلیل ساختار جمله ورودی در کنار یکدیگر قرار گرفته‌اند. وظیفه این ماژول‌ها، شکستن جمله ورودی به کلمات تشکیل‌دهنده و سپس برچسب‌گذاری با استفاده از POS Tagging و در نهایت نمایش درخت تجزیه می‌باشد. درخت تجزیه، وابستگی اجزای کلام را نمایش می‌دهد. ماژول بعدی Lookup است که هدف از این ماژول، به‌دست آوردن مجموعه‌ای از اشاره‌های متوالی برای نمایش آن‌ها در انیمیشن خروجی می‌باشد. برای این منظور این ماژول از پایگاه‌داده‌ای استفاده می‌کند که حرکات اشاره در آن‌ها ذخیره شده است و در ماژول Lookup کلمات ورودی را با آن‌ها معادل‌سازی می‌نماید. در نهایت ماژول Animate اشارات را به عنوان ورودی دریافت کرده

و توسط شخصیت انیمیشنی طراحی شده به صورت متوالی نمایش می دهد (Almeida, 2014). به منظور اثبات معماری و سیستم، آن را با یک نمونه اولیه مورد ارزیابی قرار دادند که نتایج حاصل قابل قبول ارزیابی شده است.

روش تحقیق

برای ایجاد یک سیستم مترجم از زبانی به زبان دیگر دو روش وجود دارد:

- روش تبدیل مستقیم زبان مبدا به زبان مقصد: این روش نیازمند اطلاعات کامل از هر دوزبان مبدا و مقصد می باشد، به علاوه این روش زمانی استفاده می شود که کلمات در هر دو زبان مبدا و مقصد نظیر به نظیر وجود داشته باشد.
- روش تبدیل مبتنی بر معماری: در این روش از روش مستقیم نیز برای برخی کلمات استفاده می شود اما به طور کلی سیستم تبدیل دارای چارچوبی است که تمام کلمات و جملات زبان مبدا پس از طی کردن آن چارچوب به زبان مقصد تبدیل می شود.

برای آن دسته از سیستم های مترجمی که دامنه و حوزه مشخصی دارد و دارای کاربرد عمومی نیستند، و زبان مبدا و مقصد به خوبی شناسایی شده است، به عبارت دیگر هر دو زبان مبدا و مقصد استاندارد هستند روش تبدیل مستقیم مناسب می باشد. اما برای کاربرد عمومی، روش مبتنی بر معماری مناسب است.

در سیستم مترجم متن فارسی به زبان اشاره فارسی به دلیل غیر ساخت یافته بودن زبان فارسی و هم چنین کامل نبودن زبان اشاره فارسی از روش مبتنی بر معماری استفاده می کنیم، که مراحل کار و معماری تبدیل سیستم مترجم متن فارسی به زبان اشاره فارسی در شکل ۶ نمایش داده شده است. این معماری از چهار گام اصلی برای تبدیل جمله ورودی در زبان



فارسی به زبان اشاره فارسی استفاده می کند؛

شکل ۶ معماری پیشنهادی برای تبدیل متن فارسی به زبان اشاره فارسی

۱- ماژول پیش پردازش

۲- ماژول تبدیل جمله به معادل آن در دنیای ناشنویان

۳- ماژول نمایش اشارات

ماژول پیش پردازش

ماژول پیش پردازش که اولین ماژول در معماری پیشنهادی سیستم ترجمه متن فارسی به زبان اشاره فارسی است خود دارای ماژول‌های زیر می‌باشد، که عبارتند از:

- Normalize
- Tokenize
- Formal
- Stemmer
- POS Tagging
- Parser

هریک از این ماژول‌ها که در ادامه در مورد آن‌ها صحبت خواهیم کرد به صورت گام به گام انجام می‌شود، یعنی به طور مثال زمانی که ماژول Tokenize در حال انجام است ماژول قبلی آن، که Normalize است قطعاً انجام شده است و نتیجه این ماژول در حال Tokenize شدن می‌باشد.

Normalize

از آن جا که متون مختلف ممکن است بسیار به هم شبیه باشند اما به دلیل تفاوت‌های ساده ظاهری از نظر ماشین متفاوت بیایند؛ به همین دلیل سعی شده است این تفاوت‌های ساده‌ی ظاهری برطرف گردد. برای رسیدن به این هدف، قبل از مقایسه متون، پیش پردازش‌هایی روی آن‌ها انجام می‌شود. طبیعتاً هر چه این پیش پردازش‌ها قوی‌تر باشد، نتایج حاصل از مقایسه متون قابل اطمینان‌تر خواهد بود (Graepel and Herbich, 2010). لازم به ذکر است که از آن جایی که زبان فارسی جزو زبان‌های غیر ساخت یافته است با مشکلات بسیار بیشتری نسبت به سایر زبان‌ها مواجه خواهیم شد. متون غیر ساخت یافته، متونی هستند که پیش فرض خاصی در مورد قالب آن‌ها نداریم و آن‌ها را به صورت مجموعه‌ای مرتب از جملات در نظر می‌گیریم. در پردازش رسم الخط زبان فارسی، با توجه به قرابتی که با رسم الخط عربی دارد، همواره

در تعدادی از حرف‌ها مشکل وجود دارد که از جمله آن‌ها می‌توان به حروف “ک”، “ی”، و... اشاره نمود. در اولین گام باید مشکلات مربوط به این حروف را برطرف ساخت. برخی از این اصلاحات در ذیل آورده شده است (استیری، ۱۳۹۱):

- بررسی همزه و انواع مختلف املاهای موجود و اصلاح هر کدام (به عنوان مثال تبدیل و به و، ی به ی، ا به ا، ا به ا و ...)
 - حذف تشدید از واژه‌ها
 - تبدیل ارقام عربی و انگلیسی به معادل فارسی
 - اصلاح اعراب و حذف فتحه، کسره و ضمه و همچنین تنوین‌ها
 - حذف نیم‌فاصله‌های تکراری
 - حذف نویسه‌ی «ل» که برای کشش نویسه‌های چسبان مورد استفاده قرار می‌گیرد. مانند تبدیل «بـ» به «بر»
 - چسباندن پسوندهای «تر»، «ترین» و... به آخر واژه‌ها
 - تبدیل «ة» به «هی»
 - حذف فاصله‌ها و نیم‌فاصله‌های اضافه بکار رفته در متن
- ماژول نرمال‌سازی در زبان فارسی با حذف نشانه‌های گرامری که از زبان عربی به زبان فارسی منتقل شده است کار خود را آغاز می‌کند. در ادامه ماژول نرمال‌سازی متن ورودی، به اصلاح حروف “ک”، “ی”، که در رسم‌الخط زبان فارسی با توجه به قرابتی که با رسم‌الخط عربی دارد دارای مشکل می‌باشد، اصلاح می‌شود. هم‌چنین در نرمال‌سازی متن ورودی باید اعداد با معادل فارسی خود جایگزین شود. هم‌چنین کش نویسه‌ها و فاصله‌ها باید اصلاح شود.

Tokenize

Token کردن ابزاری برای شکستن متن ورودی براساس واحدهای با معنی مانند کلمه است (Graepel and Herbich, 2010). نمادهای معناداری مانند Tab، Space و... لازمه ایجاد این ابزار در زبان فارسی می‌باشد؛ این کار به منظور شناسایی واحدهای مستقل معنایی می‌باشد. سپس براساس انتخاب هر کدام از این واحدها متن بر اساس آن‌ها شکسته خواهد شد. برای پیاده‌سازی و اعمال این ماژول در متن ورودی ابتدا بر روی متن ورودی ماژول Normalize اعمال شده و متن به شکل استاندارد تبدیل شده است و سپس متن ورودی به کلمات تشکیل‌دهنده خود شکسته شده است

Formal



ساختار زبان فارسی به گونه‌ای است که شکل نوشتاری کلمات و شکل گفتاری آن‌ها با یکدیگر متفاوت است. تمام پردازش‌ها بر روی متون بر روی شکل نوشتاری کلمات صورت می‌گیرد. بنابراین در سیستم مترجم متن فارسی به زبان اشاره فارسی نیاز داریم که اگر متن ورودی به شکل گفتاری نوشته شده باشد آن را به معادل خود در شکل نوشتاری آن کلمه تبدیل نماییم. در ماژول Formal تمام پردازش‌های پیشین بر روی متن ورودی انجام شده است. این ماژول تمام کلمات و اصطلاحات عامیانه که در گفتار فارسی از آنان استفاده می‌شود را با شکل نوشتاری خود در زبان فارسی جایگزین می‌کند.

Stemmer

این ماژول ابزاری برای ریشه‌یابی لغات و تشخیص نوع کلمه ساخته شده از آن (اسم مکان، اسم زمان، حالت فاعلی، حالت مفعولی و غیره) می‌باشد. در این ماژول ما به تک‌واژه‌ها نگاه می‌کنیم و به کلمات اطراف آن لغت کاری نداریم در فارسی به این کار ریشه‌یابی لغوی می‌گویند.

POS Tagging

این ماژول ابزاری برای مشخص کردن نوع کلمات از قبیل: اسم، صفت، قید، فعل و... می‌باشد. به عمل برچسب‌گذاری واژگانی به کلمات و نشانه‌های تشکیل‌دهنده یک متن POS tagging می‌گویند، به صورتی که این برچسب‌ها نشان‌دهنده نقش کلمات و نشانه‌ها در جمله می‌باشد. برچسب‌گذاری اجزای کلام مسئله پیچیده و دشواری می‌باشد، زیرا برخی کلمات برچسب چندگانه دارند و نقش آن‌ها در جملات مختلف ممکن است تغییر کند. به طور کلی برای برچسب‌گذاری برای هر کلمه از شایع‌ترین برچسب آن کلمه استفاده می‌شود (Graepel and Herbach, 2010).

تبدیل به دنیای ناشنویان

زبان اشاره فارسی دارای الفبای مخصوص به خود ناشنویان می‌باشد و اشارات تمام حروف الفبا برای آنان موجود است، به علاوه در زبان اشاره حرکات در نظر گرفته شده برای کلمات محدود است و تمام کلمات اشاره مستقیم ندارد. زبان اشاره کامل و بدون نقصی نیست، همچنین از گویشی به گویشی دیگر دارای تفاوت می‌باشد. در ایران دو نوع زبان اشاره وجود دارد:

• زبان اشاره استاندارد

• زبان اشاره طبیعی

زبان اشاره استاندارد که در مدارس ناشنوایان از این زبان استفاده می‌شود، هم‌چنین اخبار ناشنوایان بر اساس همین زبان بیان می‌شود دارای همان گرامر زبان فارسی است، و هر کلمه در همان مکان گرامری خود به زبان اشاره تبدیل می‌شود. با این تفاوت که تمام کلمات موجود در زبان فارسی اشاره معادل نداشته و باید از معادل اشاره‌دار کلمات استفاده شود. همچنین برخی کلمات هیچ‌گونه اشاره مستقیمی ندارد و برای نمایش آن دسته از کلمات در زبان اشاره از هجی کردن الفبای آن کلمه استفاده می‌شود. اما زبان اشاره طبیعی که قدمت طولانی‌تری نسبت به زبان اشاره استاندارد دارد دارای ساختار گرامری مخصوص به خود می‌باشد که با ساختار گرامری زبان فارسی متفاوت است. در سیستم مترجم متن فارسی به زبان اشاره فارسی از زبان اشاره استاندارد استفاده شده است. برای این که ماژول تبدیل به خوبی وظیفه خود را انجام دهد نیازمند داشتن اطلاعات از زبان و دنیای مقصد یعنی زبان اشاره و دنیای ناشنوایان دارد. زبان اشاره فارسی به دلیل عدم گسترش آن و همچنین عدم به‌روز کردن حرکات و اضافه کردن حرکات جدید به دایره حرکات اشاره به نسبت کلمات در گفتار فارسی بسیار محدود هستند، به همین دلیل ممکن است چندین کلمه در گفتار فارسی وجود داشته باشد که دارای یک اشاره معادل می‌باشد. هم‌چنین به دلیل محدودیت اشارات در ایران برخی کلمات که معادل مستقیم یا غیرمستقیم برای آن وجود ندارد به ناچار برای نمایش آن از هجی کردن استفاده می‌شود. برای برطرف کردن این مشکل نیز به سراغ استفاده از آنتولوژی می‌رویم. دامنه این آنتولوژی کلمات فارسی می‌باشند که اشاره نظیر داشته و مستقیماً برای ناشنوایان قابل درک هستند. به طور کلی دو روش خودکار و دستی برای ایجاد آنتولوژی استفاده می‌شود. روش‌های خودکار از روش‌های متفاوتی برای استخراج آنتولوژی از متون موجود استفاده می‌کنند و بیش‌تر این روش‌ها از تکنیک‌های پردازش زبان طبیعی استفاده می‌کنند. اکنون روش خودکاری که آنتولوژی قابل قبولی تولید کند، معرفی نشده است. بنابراین برای ایجاد آنتولوژی از روش‌های دستی و یا نیمه‌خودکار استفاده می‌شود. مستقل از روش ایجاد آنتولوژی می‌توان از متدولوژی خاصی برای ایجاد آنتولوژی استفاده کرد. متدولوژی استاندارد برای ایجاد آنتولوژی وجود ندارد و بسته به کاربرد آنتولوژی از متدولوژی‌های متفاوتی استفاده می‌شود. در این بخش یکی از این متدولوژی‌ها که روند آن در شکل ۷ نشان داده شده است، مطرح می‌شود.

شکل ۷ متدولوژی ایجاد آنتولوژی



در روند ایجاد آنتولوژی به صورت بازگشتی عمل می‌شود، یعنی در هر مرحله ممکن است نیاز به ایجاد اصلاحات و بازگشت به مرحله‌ی قبلی باشد. در ادامه مراحل متدولوژی شکل شرح داده می‌شود.

تشخیص محدوده

مراحل مختلف توسعه‌ی یک آنتولوژی را با تعریف دامنه و حوزه‌ی آن آغاز می‌کنیم. این کار شامل پاسخ‌گویی به سوالات پایه زیر است:

- آنتولوژی در حال توسعه، چه دامنه‌ای را پوشش می‌دهد؟
- برای چه منظوری از این آنتولوژی استفاده می‌کنیم؟
- برای چه نوع مشکلاتی، اطلاعات آنتولوژی باید قادر به پاسخ‌گویی باشد؟

باید به این نکته توجه داشت که پاسخ به این سوالات در فرآیند طراحی آنتولوژی ممکن است تغییر کنند. اما در هر حال، به محدودسازی حوزه‌ی مدل کمک خواهند کرد. به‌طور کلی مراحل تعیین و تثبیت دامنه مورد نظر برای



آنتولوژی در شکل ۸ آورده شده است.

شکل ۸ مراحل تعیین دامنه آنتولوژی

البته مراحل تعیین دامنه آنتولوژی به صورت بازگشت پذیر می باشد، یعنی در صورت تغییر پاسخ سوالات طراحی شده در راستای تعیین دامنه آنتولوژی، محدوده سازی دامنه آن نیز تغییر می کند. هم چنین هرچه پاسخ به سوالات طراحی شده برای آنتولوژی دقیق تر باشد، محدوده سازی دامنه مورد نظر با دقت بالاتری صورت می گیرد.

استفاده ی مجدد

در این مرحله مطالعه ی جامعی بر روی آنتولوژی موجود انجام و مشخص می شود که از کدام یک از آنتولوژی های موجود می توان برای ایجاد آنتولوژی مورد نظر کمک گرفت. استفاده ی مجدد این مزیت را دارد که صحت آنتولوژی های قبلی در عمل بررسی شده و از ابزارهایی که بر اساس آنها کار می کنند نیز می توان استفاده کرد.

تشخیص واژه ها

تهیه لیستی از تمامی واژه هایی که تمایل داریم برای آنها جمله بسازیم یا در یک کاربرد دربارهی آنها توضیح دهیم نیز کاری مفید محسوب می گردد. لازم به ذکر است در حال حاضر سیستم اتوماتیک و خودکاری برای تعیین واژه ها وجود ندارد و تمامی مراحل به صورت دستی انجام می شود. برای تهیه لیست واژگان پاسخ گویی به سوالات زیر می تواند مفید باشد:

- تمایل داریم در مورد چه واژگانی صحبت کنیم؟

- این واژه ها چه ویژگی هایی دارند؟

- درباره این واژه ها چه می خواهیم بگوییم؟

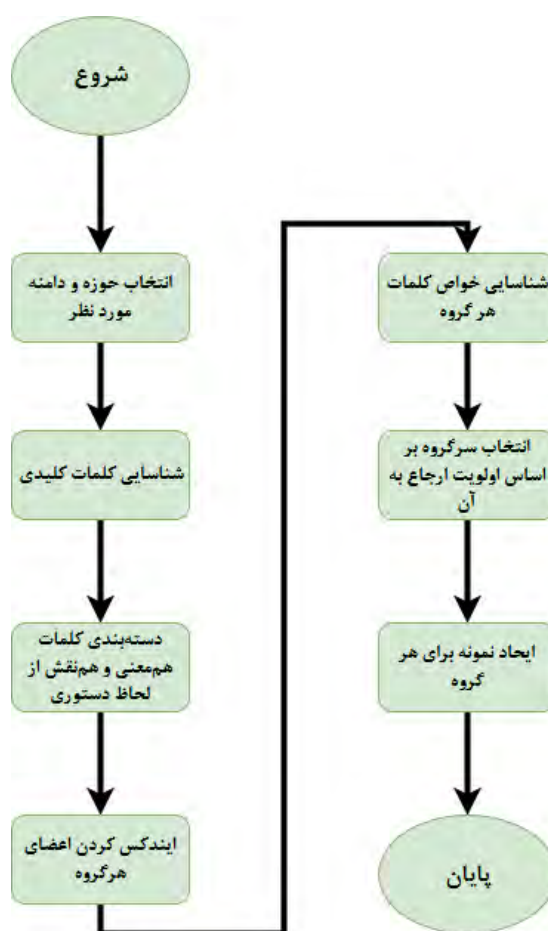
دو گام بعدی شامل توسعه سلسله مراتب کلاس ها و تعریف ویژگی های مفاهیم بسیار نزدیک به هم و در هم تنیده هستند. به گونه ای که تمایز میان آنها و قائل به تقدم و تاخر شدن میان آنها بسیار سخت است. معمولاً با ایجاد تعریف برای تعدادی محدود از مفاهیم در سلسله مراتب آغاز می کنیم و سپس با توصیف ویژگی های آنها (مفاهیم ایجاد شده) کار را ادامه می دهیم. این دو گام، مهم ترین گام ها در فرآیند طراحی آنتولوژی هستند.

تشخیص کلاس ها و ساختار آن ها

با استفاده از واژه های به دست آمده کلاس ها و زیر کلاس های آن ها را مشخص می کنیم. برای این کار می توان از روش پایین به بالا، بالا به پایین یا ترکیبی از این دو روش استفاده کرد.

- **رهیافت بالا به پایین:** فرآیند توسعه بالا به پایین با تعاریف مفاهیم عام در دامنه آغاز می کند. سپس با ایجاد زیر کلاس های خاص تر از این مفاهیم، فرآیند توسعه را ادامه می دهد.
- **رهیافت پایین به بالا:** فرآیند توسعه پایین به بالا با تعریف کلاس های بسیار خاص آغاز می کند. برگ های سلسله مراتب کلاس ها سپس با گروه بندی این کلاس ها در قالب مفاهیم عام تر، فرآیند توسعه را ادامه می دهد.
- **رهیافت ترکیبی:** فرآیند توسعه ترکیبی، ترکیبی از روش های بالا به پایین و پایین به بالا محسوب می گردد. در این روش، ابتدا مفاهیم برجسته و حائز اهمیت تعریف می شوند. سپس با عمومی سازی (ترکیب) و یا خصوصی سازی (تجزیه) این مفاهیم، فرآیند توسعه ادامه می یابد.

هیچ یک از این سه روش، به صورت ذاتی از مابقی بهتر نیست. رهیافتی که جهت توسعه آنتولوژی اتخاذ می گردد، کاملاً به دیدگاه شخصی طراح، از دامنه وابسته است. اگر یک توسعه دهنده، دیدگاه بالا به پایین و سیستماتیک به



دامنه داشته باشد، در این صورت بهتر است از روش بالا به پایین بهره ببرد. اما به نظر می‌رسد که رهیافت ترکیبی برای بسیاری از توسعه‌دهندگان آنتولوژی‌ها راه حل ساده‌تری محسوب می‌گردد و تمایل بیشتری به استفاده از این روش وجود دارد. جدا از این که کدام یک از این رهیافت‌ها را مورد استفاده قرار می‌دهیم، معمولاً با تعریف کلاس‌ها آغاز می‌کنیم. از لیستی که در گام سوم ایجاد کرده‌ایم واژه‌هایی را انتخاب می‌کنیم که اشیاء مستقل را توصیف می‌کنند. این واژه‌ها "کلاس" را مشخص کرده و نقاط لنگرگاه در سلسله مراتب کلاس‌ها را تشکیل می‌دهند. به عبارت دیگر کلاس‌ها مجموعه‌ای از نمونه‌ها است که به توصیف دقیق موجودیت‌ها اشاره دارد. یکی از تصمیمات مهم در فرآیند ساخت آنتولوژی انتخاب ابزار برای ساخت آنتولوژی و زبان نمایش آنتولوژی می‌باشد. این دو پارامتر روی چرخه‌ی حیات آنتولوژی، کیفیت و استانداردسازی آن تاثیر مستقیم دارند. ابزارهای متنوعی برای پیاده‌سازی آنتولوژی وجود دارد. به طور کلی مراحل ساخت و ایجاد آنتولوژی در شکل ۹ آورده شده است؛

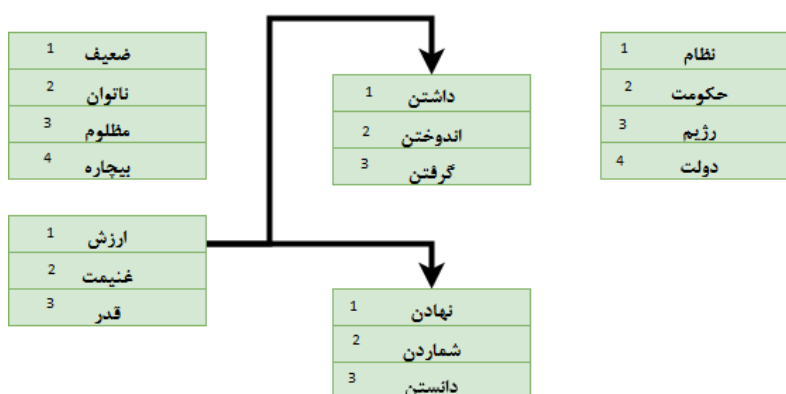
شکل ۹ مراحل ساخت آنتولوژی

الگوریتم ایجاد یک آنتولوژی به شرح شکل ۱۰ می‌باشد؛

1. Algorithm Create Ontology
2. Input (Domain)
3. Output (Ontology for Domain)
4. begin
5. For all word in Text do
6. Tokeniz Text & index Word
7. Property Detection for Words
8. Group by According to similar properties
9. Selected Header Group According to Referrals maximum
10. Create Sample for Groups
11. End

شکل ۱۰ الگوریتم ایجاد آنتولوژی

یکی از مهم ترین مراحل ایجاد یک آنتولوژی شناخت دامنه و حوزه کاری آن می باشد. با شناخت دامنه ایجاد آنتولوژی می توان واژه های محدود به آن حوزه را شناسایی کرد. پس از شناسایی واژه ها، با استفاده از یکی از رهیافت های ذکر شده، و خواص شناسایی شده برای آن ها، کلاس بندی انجام می شود. در گام بعدی اعضای هر یک از کلاس براساس میزان ارجاع به آن واژه شماره گذاری می شود. اصطلاحا در آنتولوژی به این عمل sunset بندی می گویند. سپس بین کلاس های ایجاد شده ارتباط برقرار می شود، روابط بین کلاس ها عموما is-a ، part-of ، contained-in و related to می باشد. در نهایت برای هر یک از کلاس ها نمونه تعریف می شود. حال آنتولوژی در دامنه مورد نظر ایجاد شده است



و آماده استفاده می باشد. بخشی از آنتولوژی طراحی شده را در شکل ۱۱ مشاهده می کنید.

شکل ۱۱ بخشی از آنتولوژی تبدیل

برای تبدیل کلمات از گفتار فارسی به دنیای ناشنویان مراحل کار به ترتیب شکل ۱۲ اجرا می شوند:



شکل ۱۲ مراحل کار ماژول تبدیل گفتار فارسی به دنیای ناشنوایان

برای استفاده از ماژول تبدیل متن از گفتار فارسی به همتای خود در دنیای ناشنوایان، در ابتدا متن ورودی پیش پردازش می شود. خروجی این کار تجزیه متن ورودی به کلمات تشکیل دهنده خود می باشد که ملاک تشکیل آنتولوژی می باشد. پس از تجزیه متن ورودی، بررسی می شود که هریک از کلمات در آنتولوژی مورد نظر تعریف شده است یا خیر؟ اگر کلمه مورد نظر در آنتولوژی موجود باشد، پس از یافتن کلمه در آنتولوژی به سراغ یافتن کلاس مربوط به آن کلمه می رویم، زیرا در هر کلاس تنها سرگروه آن کلاس در دنیای ناشنوایان شناخته شده است و اشاره نظیر دارد، از این رو برای تبدیل کلمه به دنیای ناشنوایان، سرگروه کلمه مورد نظر در کلاس را یافته و جایگزین کلمه مورد نظر می شود. با این کار کلمه ورودی هم در دنیای ناشنوایان دارای اشاره معادل می باشد. حال اگر کلمه مورد نظر در آنتولوژی مورد نظر تعریف نشده باشد به آن معنی است که این کلمه در دنیای ناشنوایان فارغ از اشاره معادل می باشد، به همین دلیل برای

1. Algorithm Translator
2. Input(text)
3. Output(text in deaf word)
4. Preprocessing input text
5. Tokenize text
6. Compare word in Ontology
7. If (word in ontology)
 - a. Replace word for word in deaf word
 - b. View sign Language
8. Else
 - a. Parsing word for Alphabet
 - b. View Alphabet sign word
9. End

نمایش آن در دنیای ناشنوایان از الفبای مربوط به زبان اشاره که در آنتولوژی تعریف شده است، استفاده می شود. لازم به ذکر است که برای انجام این عمل کلمه ورودی این بار به حروف تشکیل دهنده خود تجزیه می شوند. و در نهایت به صورت هجای حروف نمایش داده می شوند. مراحل بیان شده برای انجام تبدیل متن ورودی به همتای خود در دنیای ناشنوایان را در الگوریتم شکل ۱۳ نیز مشاهده می کنید.

شکل ۱۳ الگوریتم ماژول تبدیل گفتار فارسی به همتای خود در دنیای ناشنوایان

نمایش اشارات

برای نمایش زبان اشاره دو روش وجود دارد، یکی استفاده از فیلم های از قبل تهیه شده و دیگری استفاده از انیمیشن می باشد. هر دو روش مشکلات و محاسن مربوط به خود را دارد. استفاده از فیلم های از قبل تهیه شده دارای سادگی در

پایاده سازی و هم چنین به دلیل استفاده از اشخاص حقیقی که دارای تسلط کافی به زبان اشاره می باشد دارای دقت و سرعت بالایی می باشد. اما به دلیل فضای ذخیره سازی دارای سرباز حافظه است. البته برای استفاده در یک دامنه محدود و کوچک استفاده از فیلم بسیار مناسب تر و بهتر است. استفاده از انیمیشن برای نمایش زبان اشاره دارای پویایی بیشتر می باشد و دچار سربار حافظه نیز نمی باشد، اما پایاده سازی آن پیچیده و دشوار می باشد. همچنین به دلیل نمایش آن توسط کاراکترهای انیمیشنی دقت پایین تری نسبت به حالت قبل دارد. این نوع نمایش برای استفاده در دامنه بزرگ مناسب تر است. به طور کلی سیستم مترجم متن فارسی به زبان اشاره فارسی برای چندین جمله ساده و چندین جمله پیچیده و دارای جناس تهیه شده است و قابلیت توسعه آن برای یک زمینه خاص به راحتی وجود دارد.

نتایج

برای ارزیابی سیستم مترجم متن فارسی به زبان اشاره فارسی از یک مجموعه تست که حاوی ۲۱ جمله ورودی می باشد، استفاده شده است. در هریک از جملات این مجموعه چندین جمله به چشم می خورد. به طور کلی این مجموعه شامل جملات در زمینه های مختلف می باشد. بیشتر جملات به صورت خبری هستند؛ و محدود به یک دامنه و حوزه خاص نمی شود. در کل مجموعه تست شامل ۴۴۰ کلمه می باشد، ۸۶٪ جملاتی که برای تست سیستم مترجم در نظر گرفته شده اند نیازمند تبدیل جملات به دنیای ناشنویان می باشد. این جملات برای درک توسط سیستم مترجم و بالا بردن کارایی سیستم و کاهش خطا در روند ترجمه باید به همتای خود در دنیای ناشنویان تبدیل شود. به عنوان نمونه در جمله «پنج زمان را برای استعجاب دعا غنیمت شمارید.» چندین کلمه این جمله در دنیای ناشنویان ناشناخته می باشد و نیازمند تبدیل آن کلمات به معادل خود در دنیای ناشنویان می باشد. برای مثال کلمات «موقع»، «موقع» و «غنیمت شمارید» معادل ندارد، این معادل گیری توسط آنتولوژی که ایجاد شده است انجام می گیرد، کلمه «موقع» با کلمه «زمان» و کلمه «استعجاب» با کلمه «قبولی» هم چنین کلمه «غنیمت شماردن» با کلمه «ارزش نهادن» جایگزین می شود تا تمام کلمات تشکیل دهنده این جمله در دنیای ناشنویان شناخته شده باشد و توسط زبان اشاره قابل نمایش گردد. در این مجموع تست ۱۸ جمله دارای نیازمند تبدیل به دنیای ناشنویان است که برای ترجمه به زبان اشاره فارسی و درک آن توسط ناشنویان باید این کلمات به همتای خود در دنیای ناشنویان تبدیل گردد. البته ۳ جمله برای انتقال مفهوم، علاوه بر ماژول تبدیل گفتار فارسی به همتای خود در دنیای ناشنویان نیازمند ماژول دیگری نیز می باشد. بنابراین تنها ۱۵ جمله تنها با تبدیل به دنیای ناشنویان، توسط آن ها قابل درک می باشد. هم چنین ۳ جمله بدون استفاده از هیچ ماژول و پردازشی توسط ناشنویان قابل درک می باشد. به طور کلی برای ارزیابی کارایی ماژول تبدیل گفتار فارسی به دنیای ناشنویان در

جملات، هم چنین میزان تاثیر آن در دقت پردازش متن از متدهای Precision و هم چنین Recall و در نهایت متد کارایی استفاده شده است.

معيار Precision: به حاصل تقسيم «تعداد مستندات بازیابی شده مرتبط» بر «تعداد کل مستندات بازیابی شده» گفته می شود.

$$\text{Precision} = \frac{15}{18} = 0/83$$

معيار Recall: به حاصل تقسيم «تعداد مستندات بازیابی شده مرتبط» بر «تعداد کل مستندات مرتبط موجود» گفته می شود.

$$\text{Recall} = \frac{15}{21} = 0/71$$

با بهره گیری از این دو معیار به ارزیابی کارایی استفاده از ماژول تشخیص و تصحیح جناس در جملات موجود در مجموعه تست با استفاده از متد F-measure می پردازیم:

$$\text{F-measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} = \frac{2 * 0/83 * 0/71}{0/83 + 0/71} = 0/76$$

براساس این محاسبات استفاده از ماژول تبدیل گفتار فارسی به دنیای ناشنویان کارایی را به ۷۶٪ افزایش می دهد، این درحالی است که در صورت عدم استفاده از این ماژول همان گونه که در رابطه زیر مشاهده می کنید، کارایی ۲۲٪ می باشد.

$$\text{F-measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} = \frac{2 * 0/14 * 0/5}{0/14 + 0/5} = 0/22$$

بنابراین استفاده از ماژول تبدیل گفتار فارسی به همتای خود در دنیای ناشنویان کارایی سیستم مترجم متن فارسی به زبان اشاره فارسی را تقریباً ۳٫۵ برابر می کند. در نمودار شکل ۱۴ میزان کارایی سیستم مترجم با استفاده از ماژول تبدیل گفتار



فارسی به همتای خود در دنیای ناشنویان هم چنین بدون استفاده از این ماژول را مشاهده می کنید.

شکل ۴ نمودار ارزیابی کارایی استفاده از ماژول تبدیل

نتیجه گیری

زبان اشاره نه تنها برای استفاده مردم ناشنوا است، بلکه برای پدران و مادرانی که دارای کودک ناشنوا هستند و یا با افراد ناشنوا در تعامل هستند، و مربیان و معلمان کودکان ناشنوا برای برقراری ارتباط با آنها بسیار کاربرد دارد. طبق آمار فدراسیون جهانی ناشنویان (WFD) حدود ۷۰ میلیون ناشنوا در جهان وجود دارد. این درحالی است که تعداد مترجمان زبان اشاره تنها حدود ۲۵۰ نفر می باشند. به عبارتی یک مترجم به ازای هر ۷۲۰۰۰ نفر می باشد، این مسئله چالشی بزرگ برای برقراری ارتباط با ناشنویان خواهد بود. ایجاد سیستم مترجم برای هر کشور که زبان اشاره مخصوص به خود را دارد بسیار حیاتی می باشد و می تواند کار برقراری ارتباط با ناشنویان را بسیار سهولت بخشد. همان طور که در ارزیابی این سیستم مشاهده کردید استفاده از ماژول تبدیل متن فارسی به همتای خود در دنیای ناشنویان کارایی این سیستم برای انتقال مفاهیم به دنیای ناشنویان حدود ۳,۵ برابر افزایش داشته است، و دقت نمایش این سیستم مترجم برای استفاده افراد قابل قبول ارزیابی شده است.

منابع

- استیری، ا. (۱۳۹۱). ابزارهای پردازش زبان طبیعی. چهارمین کارگاه سالانه آزمایشگاه فناوری وب، دانشگاه فردوس مشهد .
- امیری و همکاران. (۱۳۸۸). طراحی و پیاده سازی آنتولوژی دروس دانشکده فنی و مهندسی دانشگاه اراک .
- بذرافکن، م. (۱۳۹۳). مروری بر مشکلات پردازش زبان فارسی با استفاده از سیستم های کامپیوتری. همایش ملی مهندسی رایانه و مدیریت فناوری اطلاعات .
- دوخانی. (۱۳۹۲). چشم اندازی به نرم افزار پروج. ماهنامه علمی الکترونیکی انجمن علمی مدیریت اطلاعات ایران .
- روح المینی، م. (۱۳۸۲). گردشگر با چراغ .
- سیاوشی، س. (۱۳۸۴). آموزش ناشنویان، دفاع از زبان اشاره در آموزش و رفع تبعیض .
- سیاوشی، س. (۱۳۸۵). زبان اشاره فارسی و ضرورت تجدیدنظر در آموزش و ارزشیابی ناشنویان ایرانی .
- طرقی، گ. خ. (۱۳۸۹). ناشنویان و زبان اشاره، محدودیت ها و امکانات .

A.Aleahmad. (1385). An Introduction to Semantic Web. *University Of Tehran*, 33.



- A.Basu, T. D. a. (2010). An English to Indian Sign Language Machine Translation System.
- Abolfazl AleAhmad, H. A., Ehsan Darrudi, Masoud Rahgozar and Farhad Oroumchian (2009). A standard Persian text collection.
- Almeida, I. (2014). Exploring Challenges in Avatar-based Translation from Eruopean Portuguese to Portuguese Sign Language.
- Biswajit Sarkar, K. D., C.D.Datta, Debranjana Sarkar, Shashanka J. Dutta, Indranil Das Roy, Amalesh Paul, Joshim Uddin Molla, Anirban Paul. (2009). A Translator for Bangla Text to Sign Language.
- Graepel, R. H. a. T. (2010). Natural Language Processing. 676.
- Mikhail G.Grif, Y. A. D., Olga O.Korolkova and Evgeniy B.Tsoy. (2013). Computer Sign Language Translation System for Hearing Impaired Users.
- Shamsfard, M. (1388). Challenges and Open Problems in Persian Text Processing