

به نام خدا



وزارت علوم، تحقیقات، و فناوری

پژوهشگاه علوم و فناوری اطلاعات ایران
(ایرانداک)

خلاصه گزارش طرح پژوهشی سازمانی

عنوان طرح پژوهشی :

ایجاد هستان‌زبان‌شناسی بر مبنای نقشه مفهومی

تاریخ شروع طرح پژوهشی	۱۴۰۳/۱۰/۱۵	تاریخ پایان طرح پژوهشی	۱۴۰۴/۱۱/۲۸
نام و نام خانوادگی مجری طرح پژوهشی	ملوک‌السادات حسینی بهشتی		
نام و نام خانوادگی همکاران طرح پژوهشی	دکتر مهدی سمائی، دکتر الهام علایی ابوذری، دکتر نصراله پاک‌نیت، دکتر امیدقیاسوند		

درباره طرح پژوهشی

در سال‌های اخیر، با افزایش اهمیت میان‌کنش‌پذیری معنایی در نظام‌های ذخیره و بازیابی و مدیریت دانش، زبان‌شناسان به مطالعه طراحی و پیاده‌سازی هستی‌شناسی به‌عنوان مناسب‌ترین ابزار برای بازنمون دانش تمرکز یافته‌اند. از هستی‌شناسی به‌عنوان ابزاری معناشناختی یاد می‌شود که قادر است مفاهیم و روابط میان آن‌ها را به‌صورت دقیق‌تر نمایش دهد. هستی‌شناسی به‌عنوان ابزاری در وب معنایی و مهندسی دانش نیز مطرح است و برای بیان مفاهیم و روابط معنایی در ایجاد روابط، نمونه‌ها، نمایش تصویری مفاهیم و ساختار مفهومی یک دامنه موضوعی کارکرد اساسی دارد و دارای ویژگی‌های خاصی است که سبب شناسایی مفاهیم، شبکه مفهومی، ترسیم ساختار دانش و کاربردهای دیگری مانند بازیابی اطلاعات در وب معنایی می‌گردد. هستی‌شناسی واژگان و مفاهیم مشترک مورد استفاده برای توصیف یک حوزه از دانش است و می‌تواند مبنایی برای سازماندهی و مدیریت دانش قرار گیرد.

روش پژوهش

- مطالعه و بررسی نتایج تحقیقات زبان‌شناختی هستی‌شناسی؛
- ایجاد نمونه‌ای از هستان‌زبان‌شناسی (Ontolinguistics) بر مبنای نقشه مفهومی با روش مت‌آنتولوژی طراحی و پیاده‌سازی آن با استفاده از زبان او.دبلیوال. (زبان وب معنایی) و نرم‌افزار پروتزه نسخه ۵/۵.

یافته‌های طرح پژوهشی

تاکنون در زبان فارسی هستی‌شناسی زبان‌شناسی ساخته نشده بود لذا برای تدوین آنتولوژی زبان‌شناسی در زبان فارسی از دستاوردهای طرح پژوهشی «طراحی نقشه مفهومی برای اصطلاح‌نامه زبان‌شناسی» که توسط "تقی رجبی و همکاران" در سال ۱۳۹۹ طراحی و به اتمام رسیده استفاده شد. در این پروژه نقشه مفهومی فارسی برای اصطلاح‌نامه دوزبانه فارسی-انگلیسی زبان‌شناسی طراحی شده بود. برای طراحی نقشه مفهومی یک حوزه علمی معمولاً از روش قیاسی استفاده می‌شود و حوزه‌های مختلف علمی به صورت عام به خاص به صورت سلسله‌مراتبی مرتب‌سازی می‌شوند. سطوح پیشروی واژگان در جهت عام به خاص در طول عملیات طراحی و با توجه به منابع مورد استفاده تعیین می‌شوند. در این طرح حدود هزار مدخل واژگانی در نظر گرفته شده است و تا شش سطح پیشروی کرده است. سرفصل دروس و کتب معتبر دانشگاهی در مقاطع تحصیلات تکمیلی زبان‌شناسی و منابع مرجع و فرهنگ‌های تخصصی زبان‌شناسی انگلیسی و فارسی و در گرایش‌های مختلف آواشناسی، واج‌شناسی، صرف، نحو و معنی‌شناسی مورد استفاده قرار گرفته است. همچنین رویکردهای تاریخی علم زبان‌شناسی از منظر سنتی، ساختاری و شناختی مورد توجه بوده است. در این پروژه برای دیداری‌سازی نقشه‌های طراحی شده از نرم‌افزار Visio استفاده شده است.

در پروژه «هستی‌شناسی زبان‌شناسی» در نظر گرفته شد که بر مبنای نقشه مفهومی مذکور، ابتدا اصطلاح‌نامه زبان‌شناسی که نیمه کاره مانده بود، مجدداً با کمک LLM ها (مدل‌های بزرگ زبانی) تدوین شده و سپس آنتولوژی مورد نظر طراحی و پیاده‌سازی شود، و از زبان او.دبلیو.ال. (زبان وب معنایی) و نرم‌افزار پروتژه نسخه ۵،۵ استفاده گردد

برای ساخت هستی‌شناسی ابتدا کدنویسی‌هایی که در هستی‌شناسی‌ها قابل تطبیق با نرم‌افزار پروتژه بودند، مورد مطالعه قرار گرفتند. به همین منظور از ابتدا کدهای نوشته شده در هستی‌شناسی زبان‌شناسی شروع به گسترش داده شدند. در این کدهای اولیه ویژگی‌های کلی فایل هستی‌شناسی مشخص گردید. بنابراین مجدداً ترجمه مفاهیم درون نقشه مفهومی زبان‌شناسی را بررسی کرده تا از یکدست بودن آن‌ها اطمینان حاصل شود. برای این منظور هم از روش‌های خودکار و بر مبنای هوش مصنوعی بهره گرفته شد و غنی‌سازی روابط معنایی آنتولوژی با انطباق با «مدل‌های بزرگ زبانی» (LLMs) توسعه یافت.

نتیجه‌گیری و پیشنهادها

□ آنتولوژی توسط متخصصان انسانی بازبینی و پالایش شود؛

□ نسخه RDF/OWL رسمی ایجاد شود؛

□ داده‌ها به صورت مستمر گسترش یابند؛

□ از ابزارهایی مانند Protégé برای مدیریت آنولوژی استفاده شود؛

□ لایه‌های جدید برای تحلیل گفتمان، سبک‌شناسی و واج‌شناسی افزوده شود.