

به نام خدا



وزارت علوم، تحقیقات، و فناوری

پژوهشگاه علوم و فناوری اطلاعات ایران
(ایرانداک)

خلاصه گزارش طرح پژوهشی سازمانی

عنوان طرح پژوهشی

ایجاد هستی‌شناسی از اصطلاح‌نامه کشاورزی ایرانداک (آگرووک)

تاریخ شروع طرح پژوهشی	۱۴۰۲/۳/۲۲	تاریخ پایان طرح پژوهشی	۱۴۰۳/۸/۱۵
نام و نام خانوادگی مجری طرح پژوهشی	ملوک‌السادات حسینی بهشتی		
نام و نام خانوادگی همکاران طرح پژوهشی	اسماعیل اکبری، امید قیاسوند، علی اصغر حجت پناه		

درباره طرح پژوهشی

تحول در حوزه نظام‌های ذخیره و بازیابی اطلاعات منجر به ظهور دستاوردهای نوینی شده که هر یک از آنها ابزارهایی را به کار می‌گیرند تا بتوانند پاسخگوی نیاز کاربران خود باشند. وب معنایی تجسم پیشرفت هوشمند وب کنونی برای جمع‌آوری، استفاده و تفسیر اطلاعات در فرمتی قابل پردازش برای ماشین، نمونه‌ای از این امکانات نو ظهور است که برای طراحی اطلاعات در قالب مدل‌های قابل پردازش برای ماشین، از ابزار "هستی‌شناسی" استفاده می‌کند. هستی‌شناسی به عنوان قلب فرایند ذخیره و بازیابی معنایی، در طول زمان پیشرفت و مقبولیتی کسب کرده است. وابستگی قابل توجه وب معنایی به هستی‌شناسی‌ها به سبب مهمترین ویژگی این ابزار یعنی بیان مفاهیم با تبیین روابط میان آنها در یک حوزه دانش بشری است که درک و پردازش ماشینی را سبب می‌شود.

هدف از ایجاد هستی‌شناسی کشاورزی و به کارگیری آن در نظام بازیابی اطلاعات، ساختارمند نمودن اطلاعات و در نهایت بازنمایی دانش به کمک ماشین با استفاده از فناوری‌ها و استانداردهایی است که به وسیله کنسرسیوم وب جهان گستر در حال تدوین و تکمیل است. به عبارت دیگر در وب معنایی ماشین‌ها با شناسایی داده‌های اسناد و مدارک موجود در فرایند تحلیل و استخراج اطلاعات مورد نظر کاربر مشارکت کرده و در این زمینه هوشمندانه‌تر عمل می‌کنند. در وب معنایی موتورهای جستجو بیش از پیش بر مفهوم داده‌ها و جداسازی آنها مسلط هستند.

از دهه ۱۹۸۰ سازمان غذا و کشاورزی وابسته به سازمان ملل (فائو) اصطلاح‌نامه چند زبانه "آگرووک" را تدوین و منتشر کرد، که تمامی موضوعات مورد نظر فائو اعم از تغذیه، کشاورزی، جنگلداری، ماهیگیری و محیط زیست و نیز واژگان مرجع ارگانسم‌ها و مفاهیم محیط زیست، روشهای کشت گیاهان و تحقیقات زیست محیطی و دیگر موضوعات مرتبط را در بیش از ۴۰ زبان گردآوری کرده است. در حال حاضر حدود ۲۵ سازمان ملی و بین‌المللی در سراسر دنیا اصطلاح‌نامه آگرووک را پشتیبانی می‌کنند.

آگرووک به صورت برخط در دسترس می‌باشد و مبتنی بر ساختار RDF و SKOS شکل گرفته است و بیش از ۴۰۰۰۰ مفهوم را در برمی‌گیرد که دارای URI های متمایز کننده هستند. اما در نهایت در برابر هر مفهوم در هر زبان یک اصطلاح مرجع قرار می‌گیرد.

در ادامه تبدیل اصطلاح‌نامه‌های ایرانداک به آنتولوژی، طرح ایجاد هستی‌شناسی علوم کشاورزی با استفاده از اصطلاح‌نامه علوم کشاورزی ایرانداک و با انطباق با Agrontology که در سایت FAO در دسترس است به توسعه مفاهیم و گسترش روابط معنایی آن می‌پردازیم. بدین منظور نرم‌افزار پروتژه که دسترس‌پذیر و قابل استفاده برای زبان فارسی است، به کار گرفته و در نهایت هستی‌شناسی کشاورزی حاصل خواهد شد.

همچنین پیاده‌سازی آنتولوژیها تدوین شده در سامانه‌ای جداگانه و قرار گرفتن در سایت پژوهشگاه با همکاری متخصصین مرفا انجام خواهد پذیرفت.

روش پژوهش

پژوهش حاضر از نظر ماهیت کاربردی است و با توجه به این که از دو رویکرد کیفی و کمی استفاده شده است، روش ترکیبی (آمیخته) است. برای ساختن ابزار اولیه پژوهش با استفاده از منابع کتابخانه‌ای و اسنادی، مقوله‌های موضوعات اصلی و فرعی شناسایی می‌شود. روش این پژوهش، گردآوری داده‌ها در بخش کیفی و کمی دارای اهمیت است. در بخش کیفی از روش تحلیل موضوع به منظور ساخت ابزار پژوهش استفاده شده است. البته برای تعیین تعداد و فراوانی کمی مفاهیم اصلی و فرعی از روش کمی نیز استفاده شده است. روش کتابخانه‌ای و اسنادی برای گردآوری اطلاعات و شناسایی مقوله‌ها و تعیین نمایه‌های موضوعی به کار گرفته می‌شود. در مرحله بعد تحلیل محتوا، کار با داده‌ها و سازماندهی آنها، تجزیه داده‌ها به واحدهای قابل کنترل، ترکیب آنها و جست و جوی الگوها خواهد بود. این روش قابلیت این را دارد که مفاهیم به دست آمده را بسط ساختاری، دهیم و طبقه‌بندی نموده و در قالب نقشه مفهومی و مدل ارائه دهیم.

جامعه آماری این پژوهش تعداد ۱۶۰۰۰ مدخل واژگانی مستخرج از اسناد موجود حوزه کشاورزی است که به صورت سلسله مراتبی و به شکل اصطلاح‌نامه در پایگاه اصطلاح‌نامه‌های ایرانداک موجود می‌باشد. بنابراین از سویی برای استخراج مفاهیم و مقوله‌بندی آنها و برقراری روابط اصطلاح‌نامه‌ای از داده‌های اسنادی و از سوی دیگر برای تعیین میزان فراوانی مفاهیم اصلی و فرعی، از این روش استفاده می‌شود. و در نهایت با مقوله‌بندی پر بسامدترین مقوله‌های نمایه‌های موضوعی به تدوین شبکه معنایی مفاهیم رشته کشاورزی پرداخته می‌شود. به طور کلی فرایند مقوله‌بندی در تحلیل محتوا وابسته به اهداف پژوهش می‌تواند قیاسی یا استقرایی صورت پذیرد. از آنجا که پژوهش حاضر رویکرد تلفیقی دارد، سعی می‌کنیم با بررسی ادبیات نظری و منابع مرجع حوزه پژوهشی درک مناسبی برای شناخت مفاهیم فراهم گردد. (روش قیاسی) و از طرف دیگر با بررسی داده‌ها به شناسایی مفاهیم اصلی و روابط معنایی میان آنها پردازیم (روش استقرایی). در این پژوهش برای ساخت آنتولوژی از روش "مت آنتولوژی" ابداع گومز-پرز و همکاران (۲۰۰۴) استفاده می‌شود. و نهایتاً با نرم‌افزار پروتژه شبکه‌های مفهومی و ساختار آنتولوژی مورد نظر را تعیین می‌نماییم.

گام‌های فعالیت مدل‌سازی مفهومی طبق روش مت‌آنتولوژی (گومز-پرز ۲۰۰۴):

۱- ساخت واژه‌نامه اصطلاحات

۲- ساخت رده‌بندی مفاهیم

۳- ساخت نمودار روابط دودویی

۴- ساخت فرهنگ‌نامه مفاهیم

۵- توصیف روابط دودویی

۶- توصیف مشخصه‌های نمونه‌ها

۷- توصیف مشخصه‌های کلاس

۸- توصیف ثابت‌ها

۹- توصیف اصول صوری

۱۰- توصیف قواعد

۱۱- توصیف نمونه‌ها

.

یافته‌های طرح پژوهشی

در گام اول: طراحی هستی‌شناسی با استفاده از روش «مت‌آنتولوژی» و بر مبنای اصطلاحنامه کشاورزی و تبدیل اصطلاحنامه به هستی‌شناسی انجام گرفت. از آنجایی که ایزو ۲۵۹۶۴ مفهوم-بنیان است، همه مفاهیم اصطلاحنامه به عنوان کلاسهای هستی‌شناسی و اصطلاحات اصطلاحنامه به عنوان برچسب کلاس‌ها در نظر گرفته شده‌اند. تعاریف، یادداشت‌های دامنه، و دیگر یادداشت‌ها و اطلاعات به توضیحات تبدیل و روابط سلسله‌مراتبی در اصطلاحنامه، بعنوان روابط نوعی در هستی‌شناسی در نظر گرفته و از طریق اصل زیرکلاسی نمایش داده می‌شود. روابط وابسته که نشان می‌دهد-شناسی در رابطه‌های هستی‌شناختی بین دو مفهوم وجود دارد، بخشی از رابطه‌ای هستی‌شناختی در نظر گرفته شده پس از این مرحله کلیه روابط اصطلاحنامه بر اساس تبدیل نحوی تعیین شده در قالب زبان ادبلیو ال پیاده‌سازی و به هستی‌شناسی تبدیل شده است. OWL.

در گام دوم: ما تلاش کردیم که بین مفاهیمی که در اصطلاحنامه به عنوان "رابطه وابسته" وجود دارد یک رابطه معنایی برقرار کنیم. روابط وابسته در اصطلاح نامه، نشانگر وابستگی بین زوج‌های مفاهیمی است که به صورت سلسله‌مراتبی با یکدیگر مرتبط نیستند اما از نظر معنایی یا مفهومی به یکدیگر وابستگی قوی دارند. این رابطه در اصطلاح‌نامه یک رابطه دوطرفه است.

برای دستیابی به این هدف، ابتدا الگوهای روابط معنایی از هستی‌شناسی سطح بالای "آگرانولوژی" استخراج می‌شود. این الگوها شامل نام رابطه، دامنه و برد و معکوس آن^۱ است. علاوه بر این، تعریف و توضیحات مربوط به هر رابطه معنایی برای استفاده^۲ متخصصان موضوعی در جدولی تنظیم و تعدادی از الگوهای روابط معنایی نشان داده شده است. هر رابطه معنایی با یک عبارت مشخص شده است که بیانگر نوع رابطه معنایی بین دو مفهوم است. این رابطه در صورتی قابل اعمال است که نوع مفهوم اول مطابق با نوع مشخص شده برای دامنه رابطه و نوع مفهوم دوم مطابق با برد رابطه باشد. معکوس یک رابطه، نشان‌دهنده رابطه^۳ ای است که مفهوم دوم با مفهوم اول خواهد داشت. دامنه و برد معکوس رابطه به ترتیب مطابق با برد و دامنه^۴ خود رابطه است. برخی از روابط معنایی می‌توانند الگوهای متفاوتی داشته باشند مانند رابطه «دارای حاصل» که می‌تواند از یک مفهوم از نوع «فرآیند» به مفاهیمی از نوع «موجودیت مادی»، «موجودیت غیرمادی» و یا «موجودیت اطلاعاتی» تعریف شود.

نتیجه‌گیری و پیشنهادها

پس از تبدیل اصطلاح‌نامه به هستی‌شناسی کشاورزی، مراحل بعدی این طرح کلان پیش‌بینی و قابل اجرا می‌باشد:

- ۱- توسعه دامنه مفاهیم و روابط مفهومی از تعداد موجود به تعداد فعلی مفاهیم در هستی‌شناسی کشاورزی انگلیسی یعنی ۴۰۰۰۰ مفهوم موجود در "آگرانولوژی" Agrontology.
- ۲- توسعه روابط پیوندی (Object Properties) در ویرایش هستی‌شناسی فارسی کشاورزی.
- ۳- توسعه دامنه مفاهیم به مفاهیم بومی کشاورزی در ایران (بومی‌سازی هستی‌شناسی کشاورزی با مفاهیم مورد تاکید و تایید در بخش کشاورزی کشور).