

بسم الله الرحمن الرحيم



وزارت علوم، تحقیقات و فناوری
پژوهشگاه علوم و فناوری اطلاعات ایران

طرح پژوهشی

طراحی هستی‌شناسی علوم پایه و پایگاه اطلاعات مفهوم – بنیاد آن

گزارش نهایی

مجری

ملوک‌السادات حسینی بهشتی

اسفند ۱۳۹۱

حقوق: پژوهشگاه علوم و فناوری اطلاعات ایران

طرح پژوهشی «طراحی هستی‌شناسی علوم پایه و پایگاه اطلاعات مفهوم - بنیاد آن» پیرو قرارداد شماره ۳۳/۲۷۷۵۵ مورخ ۸۹/۱۲/۱ میان پژوهشگاه علوم و فناوری اطلاعات ایران (کارفرما) و خانم ملوک السادات حسینی بهشتی (مجری) اجرا شده است. گزارش حاضر جلد نهایی از مستندات این طرح است.

این گزارش و تمامی حقوق مادی آن بر اساس «قانون حمایت حقوق مؤلفان و مصنفان و هنرمندان، مصوب سال ۱۳۴۸ و اصلاحیه‌های بعدی آن و همچنین آیین‌نامه‌های اجرایی این قانون متعلق به پژوهشگاه علوم و فناوری اطلاعات ایران است و هر گونه استفاده از تمامی یا پاره‌ای از آن، شامل: نقل قول، تکثیر، انتشار، کاربرد نتایج، تکمیل و مانند آنها به صورت چاپی، الکترونیکی یا وسایل دیگر فقط با اجازه کتبی پژوهشگاه امکان‌پذیر است. نقل قول در حد هزار واژه در انتشارات علمی مانند کتاب و مقاله با درج اطلاعات کامل کتاب‌شناختی، نیازی به مجوز پژوهشگاه ندارد.

صحت مندرجات گزارش بر عهده مجری طرح پژوهشی است.

طراحی هستی شناسی علوم پایه و پایگاه اطلاعات مفهوم - بنیاد آن

گزارش نهایی

مدیر طرح پژوهشی: ملوک السادات حسینی بهشتی

نشانی: تهران، خیابان انقلاب، چهارراه فلسطین، شماره ۱۰۹۰، طبقه ۳، اتاق ۳۱۹

صندوق پستی: ۱۳۱۵۷۷۳۳۱۴ تلفن: ۰۲۱-۶۶۴۹۴۹۸۰ ، دورنگار: ۰۲۱-۶۶۴۶۲۲۵۴

وبگاه: <http://www.irandoc.ac.ir>

رایانامه: [Email:Beheshti@irandoc.ac.ir](mailto:Beheshti@irandoc.ac.ir)

چکیده

تعامل میان هستی‌شناسی‌ها و زبانهای طبیعی مقدم بر فرآیند پردازش زبان طبیعی و مهندسی دانش است. هستی‌شناسی‌ها که عمدتاً از الگوهای معناشناختی بهره می‌برند وجوه مشترکی با واژگان دارند. یک مجموعه واژگان اصطلاحات را براساس مفاهیم مرتبط سازماندهی می‌کند، در حالی که هستی‌شناسی‌ها مفهوم سازی کرده و روابط مفاهیم را تعیین می‌نمایند.

یک واژگان مشترک پیش زمینه و ابزاری است در به اشتراک گذاشتن دانش جمعی، و یک هستی‌شناسی مشترک علاوه بر این ابزاری است در به اشتراک گذاشتن دانش از طریق فناوری اطلاعات. در ساخت الگوهای مفهومی و ساختار زبانی، این حوزه زبان‌شناسی رایانه‌ای است که در ترسیم روابط میان مفاهیم و واژگان بمنظور کاربرد در مهندسی دانش و سازماندهی اطلاعات نقش ایفا می‌کند. به منظور شکل‌گیری هستی‌شناسی مشترک به تلفیق و همسان‌سازی اصطلاحنامه‌های موجود در پژوهشگاه علوم و فناوری اطلاعات پرداخته، مغایرت‌ها و همپوشانی‌های واژگان علوم پایه را برطرف و شبکه مفاهیم این گرایش را با زیرمجموعه‌های آن به شکل هستی‌شناسی واژگانی علوم پایه طراحی نمودیم تا پس از تبدیل روابط اصطلاحنامه‌ای به مفهوم-بنیاد به زبان OWL2 و براساس استاندارد ۲۵۹۶۴ در پایگاه اطلاعات مفهوم-بنیاد به عنوان ابزاری قوی جهت ذخیره و بازیابی اطلاعات بکار گرفته شود.

کلیدواژه‌ها: هستی‌شناسی، واژگان، پردازش زبان طبیعی، مهندسی هستی‌شناسی، فناوری اطلاعات.

فهرست مطالب

عنوان	شماره صفحه
چکیده.....	ب.
فهرست مطالب.....	ت.
فهرست شکلها.....	ج.
۱-مقدمه.....	۳.
۲-بیان مسئله.....	۴.
۳-پیشینه پژوهش.....	۵.
۴- سابقه پژوهش در جهان.....	۷.
۵- اشتراک منابع اطلاعاتی و اکتشاف دانش.....	۱۰.
۶- تعامل میان هستی شناسی و واژگان.....	۱۶.
۷- روابط معنایی.....	۲۷.
۸- هستی شناسی و واژگان.....	۳۶.
۹- مهندسی هستی شناسی.....	۴۰.
۱۰- شیوه طراحی و ساخت هستی شناسی.....	۵۷.
۱۱- مقولات هستی شناسی.....	۵۸.
۱۲- تدوین ساختار درختی.....	۶۶.
۱۳- همسانسازی اصطلاحنامه.....	۷۴.
۱۴- روش استفاده از نرم افزار همسانساز.....	۷۵.
۱۵- مراحل تلفیق اصطلاحنامه ها.....	۸۲.

۱۶- مراحل تبدیل اطلاعات واژگانی به مفهوم- بنیاد.....	۸۳
۱۷- نتیجه گیری.....	۸۵
۱۸- منابع.....	۱۰۳
ABSTRACT.....	۱۰۶

فهرست شکلها

عنوان	شماره صفحه
نمودار ۱. انواع ابزار رده‌بندی معناشناختی (DACONTA ET. AL. 2003)	۱۲.....
نمودار ۲. مفهوم سازی در هستی‌شناسی (GUARION, 1999)	۱۷.....
نمودار ۳. تاکسونومی منابع آنتولکس (PREVOT, L. ET. AL. 2010)	۲۴.....
نمودار ۴ : مثلث معنی‌شناختی (DACONTA ET. AL)	۴۷.....
نمودار ۵: اصطلاحنامه در مقابل هستی‌شناسی (DACONTA EL. AL. 2003)	۴۸.....
نمودار ۶. رده‌بندی جغرافیایی سیستم چات-۸۰ (SOWA, 2000)	۶۰.....
نمودار ۷. کاربرد رده‌بندی سطوح بالا در سیستم CYC (SOWA, 2000)	۶۱.....
نمودار ۸. شبکه مفاهیم براساس تمایزات سه گانه (SOWA, 2000)	۶۲.....
نمودار ۹. سلسله مراتب نظریه‌ها	۶۴.....
نمودار ۱۰. نمونه‌ای از «اصطلاح‌نامه زیست‌شناسی» تدوین: اسماعیل اکبری	۶۷.....
نمودار ۱۱. نمونه‌ای از «اصطلاح‌نامه شیمی» تدوین: تقی رجبی	۶۹.....
نمودار ۱۲. نمونه‌ای از «اصطلاح‌نامه شیمی» تدوین: تقی رجبی	۷۰.....
نمودار ۱۳. نمونه‌ای از «اصطلاح‌نامه شیمی» تدوین: تقی رجبی	۷۱.....
نمودار ۱۴. نمونه‌ای از «اصطلاح‌نامه شیمی» تدوین: تقی رجبی	۷۲.....
نمودار ۱۵. نمونه‌ای از «اصطلاح‌نامه شیمی» تدوین: تقی رجبی	۷۳.....

گزارش نهایی

**طراحی هستی‌شناسی علوم پایه و پایگاه اطلاعات
مفهوم – بنیاد آن**

۱- مقدمه

این طرح براساس برنامه استراتژیک پژوهشگاه علوم و فناوری اطلاعات در راستای طرح‌های پژوهشی «استانداردسازی شیوه‌های ذخیره و بازیابی اطلاعات علمی-فنی» و بر مبنای تدوین اصطلاح‌نامه‌ها و گسترش آنها، یکدستی و استاندارد سازی «هستی‌شناسی» های تخصصی پشتوانه پایگاه‌های اطلاعات علمی-فنی و مبتنی بر وب» می‌باشد. ساماندهی دانش فنی و سازماندهی مفاهیم علمی یکی از مهم‌ترین مباحث علم اصطلاح‌شناسی محسوب می‌گردد. «مفهوم» به عنوان واحد دانش فنی تعریف شده، از مختصات معنایی یعنی اجزای دانش تشکیل شده است. بر این اساس بعد محتوایی و بیانی هر یک از مفاهیم به عنوان دانش اصطلاح‌شناختی یک مفهوم در نظر گرفته می‌شود. به منظور ترسیم نظام مفاهیم دانش بنیان هر یک از رشته‌های تخصصی بایستی مفاهیم زیر رشته‌های مرتبط با آن را در یک ساختار کلی و روابط معنایی میان آنها ایجاد نمود. به منظور درک و دریافت ساختار دانش فنی یک حوزه تخصصی بایستی موضوعات را در یک ساختار نظام کلان جمع‌بندی و تلفیق نمود که چنین ساختار مفهومی را هستی‌شناسی می‌نامند. برای بازیابی اطلاعات مدارک علمی-فنی رده‌بندی مفاهیم علمی را به شکل ارائه نظام‌های منطقی در قالب هستی‌شناسی‌های تخصصی سازماندهی می‌کنند. آنتولوژی به معنی هستی‌شناسی است که به موضوع موجودیت، مطالعه هستی و ساختار واقعیت می‌پردازد. به عبارتی دیگر: هستی‌شناسی یک گرایش علمی در فلسفه است که به مطالعه آنچه موجود است و آنچه باید فرض شود که وجود دارد؛ و به منظور دستیابی به یک توصیف متقاعد کننده از واقعیت می‌پردازد.

تعریف «سوا»^۱ از هستی‌شناسی به این شرح است: «هستی‌شناسی دسته‌بندی از موضوعات با عناصر موجود در یک حوزه خاص را بررسی می‌کند و براساس آن بررسی، فهرستی از موضوعات را ارائه می‌دهد که به تفصیل انواع موضوعات و روابط منطقی میان آنها را بیان می‌کند و از این رو هستی‌شناسی نامیده می‌شود. از هستی‌شناسی در حوزه‌های مختلفی از جمله هوش مصنوعی، مهندسی نرم‌افزار و اطلاع‌رسانی استفاده شده است» (Sowa 1999).

«گروبر»^۲ از پیشگامان معرفی هستی‌شناسی است که براساس نظر وی هستی‌شناسی: «ذکر مختصات صریح و روشن از مفاهیم به اشتراک گذاشته شده است» (Gruber, 1993). از هستی‌شناس‌ها بر روی وب نیز استفاده می‌شود که به بازیابی اطلاعات متن بنیاد مبتنی بر وب به کار بران کمک می‌کند.

^۱ - Sowa

^۲ - Gruber

«سورگل»^۳ با توجه به این نکته هستی‌شناسی را یک ابزار سازماندهی منابع وب و بازیابی دانش با کمک زبان طبیعی، معرفی کرده است. محیط وب دستیابی گسترده به اطلاعات را فراهم کرده است. این محیط اطلاعاتی جدید و تغییر توانایی‌های کاربران و نیازهای اطلاعاتی متفاوت، نیاز به ابزارهای یاری دهنده در امر بازیابی و سازماندهی اطلاعات رو به رشد را ضرورت بخشیده و یکی از این ابزارها هستی‌شناسی‌ها هستند (Soergel, 2004).

در پژوهشگاه علوم و فناوری اطلاعات نیز تاکنون از اصطلاحنامه‌های علوم پایه جهت سازماندهی و بازیابی اطلاعات استفاده می‌شده است و در آینده براساس گسترش کارایی آنها بایستی به شکل آنالوژی علوم پایه در آید تا بتوان از این فناوری جدید در راستای جامعیت و مانعیت بیشتر در جستجو و بازیابی اطلاعات علمی - فنی و کسب نتایج کامل‌تر و دقیق‌تر استفاده نمود.

۲- بیان مسئله

در تدوین اصطلاحنامه، ایجاد یک ساختار سلسله‌مراتبی از مفاهیم، اولین ضرورت و ایجاد و تعیین نوع روابط بین مفاهیم دومین ضرورت است که در رویکرد ساختار مبنای، فرا روی تدوین کنندگان اصطلاحنامه قرار دارد. اما پیچیدگی‌های شرایط نمایه‌سازی و نیز دیدگاههای متفاوت کاربران نظامهای بازیابی اطلاعات، ضرورت سومی را با رویکرد کار بر مبنای به عرصه سازماندهی و بازیابی اطلاعات وارد ساخته است. نمایه‌سازان در بسیاری از موارد هنگام استفاده از اصطلاحنامه در سازماندهی اطلاعات، با توصیفگرهای متعدد از چند مقوله موضوعی متفاوت مواجه می‌شوند که برای نمایاندن یک موضوع اطلاعاتی، مناسب به نظر می‌رسد. ابهام در استفاده از توصیفگرهای مناسب‌تر هنگام سازماندهی از یک سو و دیدگاههای متفاوت کاربران نهایی نظام‌های بازیابی اطلاعات از سوی دیگر، گسترش روابط معنایی در اصطلاحنامه را ضروری می‌سازد. شناخت حدود و ثغور مفاهیم و مصداق‌های آنها با ابهام کمتری روبرو می‌گردد. در چنین شرایطی اختصاص توصیفگرها با قطعیت بیشتر هنگام سازماندهی و امکان تعیین درخواست جستجوی اطلاعات با استفاده از مختصات معنایی، سطح اثربخشی نظام بازیابی اطلاعات را از دید کاربران افزایش خواهد داد. همچنین وارد ساختن مختصات معنایی به ساختار مفاهیم در اصطلاحنامه، آنرا به سطح مدل مفهومی سلسله‌مراتب هستی‌شناسی‌های تخصصی نزدیک می‌کند. سازماندهی دانش با هدف دسته‌بندی مفاهیم مشابه به جهان‌های ممکن نیز می‌نگرد. هر پدیده‌ای به منزله یک مفهوم - در هستی، حائز ارتباطی معنایی با سایر مفاهیم است، که مجموعه این پدیده‌ها به همراه ارتباط معنایی شان با یکدیگر، شبکه مفهومی از هستی را، در

^۳ - Soergel

جهانی ممکن به تصویر می کشد. جدای از آنکه دسته‌بندی و مدل‌سازی مفهومی را حاصل کشفیات علمی یا ساخت انسانی بدانیم، تعیین مفاهیم و روابط معنایی در بافت زبانی تعیین شده صورت می‌پذیرد. هستی‌شناسی در واقع ابزار بازنمون دانش در حوزه سازماندهی و هوش مصنوعی است و تأکید بر کاربرد آن در وب معنایی در دهه اخیر مورد توجه دانشمندان این رشته است. هدف از کاربرد هستی‌شناسی، کاهش ابهام مفهومی در حوزه‌ای خاص است. کاهش ابهام، نتیجه مستقیم وجود روابط معنایی روشن و دقیق میان مفاهیم یک علم است. روابط معنایی مجاز در هستی‌شناسی‌ها براساس نرم‌افزارهای گوناگون و بویژه بر زبان هستی‌شناسی RDF و OWL، جدیدترین استاندارد زبانی کنسرسیوم جهانی ISO/DIS 25964 در بازنمون دانش امکان پذیر می‌باشد. هستی‌شناسی‌ها در حوزه هوش مصنوعی و مهندسی دانش رشد یافته و از پویایی بیشتری نسبت به اصطلاحنامه‌ها و فرا اصطلاحنامه‌ها برخوردارند و می‌توانند در سازماندهی و مدیریت دانش مبنای قرار گیرند. بر این اساس ارائه الگویی از هستی‌شناسی علوم پایه می‌تواند در آینده مبنای ساخت هستی‌شناسی‌های دیگر قرار گرفته و اولاً برای بازیابی اطلاعات پایگاه‌های علمی - فنی و ثانیاً برای برنامه‌ریزی آموزش و پژوهش و ثالثاً برای تهیه نقشه جامع علمی مورد استفاده قرار گیرد.

۳- پیشینه پژوهش

در داخل کشور در زمینه کاربرد آنتولوژیها در ترجمه ماشینی و «وردنت فارسی» در دانشکده‌های برق و کامپیوتر طرحهای پژوهشی در دست تکمیل و اجراست، منجمله «وردنت فارسی» یا «فارس‌نت» که توسط دکتر مهرنوش شمس فرد در دانشکده علوم کامپیوتر دانشگاه شهید بهشتی و با حمایت مرکز تحقیقات مخابرات در دست اجرا می‌باشد و در واقع «شبکه واژگانی» براساس واژه‌نامه‌های متعدد منجمله فرهنگ آریان‌پور و با استفاده از واژه‌های عمومی زبان فارسی شکل یافته و در حال تکمیل است.

نمونه دیگر از این دست طرحهای پژوهشی در دانشگاه شریف و توسط دکتر ابوالحسنی و همکاران در دست اجرا می‌باشد که هدف نهایی آن ترجمه ماشینی و مهمتر از آن ایجاد شبکه معنایی^۴ است که مراحل از پروژه پشت‌سر گذاشته شده و در حال تکمیل است.

طرح دیگری که دکتر «هشام فیلی» در دست مطالعه و اجرا داشته و آن نیز بیشتر بر نحو زبان فارسی متمرکز بوده و هدف نهایی آن نیز «ترجمه ماشینی» دو زبان انگلیسی به فارسی و بالعکس می‌باشد. علاوه بر این در مورد ترجمه ماشینی زبان فارسی به انگلیسی، پروژه کلانی در آزمایشگاه تحقیقاتی

^۴ - Semantic Web

کامپیوتر، دانشگاه مکزیکوسیتی (نیومکزیک) توسط دکتر «کارین مگردومیان»^۵ و همکاران در دست اجرا و تکمیل است که از سال ۲۰۰۰ گزارش آن در دست است و بر "متن ماشین خوان" استوار است و هدف آن بیشتر ترجمه اخبار از فارسی به انگلیسی است و پشتوانه آن واژه‌نامه عمومی است شامل ۵۰ هزار واژه که به تدریج از سال ۱۹۹۷ تا ۲۰۰۰ از متون خبری فارسی استخراج و از نظر ساختوازی و نحوی^۶ در حجم کلان پیکره زبانی دو زبانه برچسب دهی^۷ شده و تحت عنوان پروژه شیراز^۸ در حال تکمیل می‌باشد.

در زمینه کاربرد آنتولوژیها در بازیابی اطلاعات علمی - فنی مقالات فراوانی منجمله از سوی داورپناه (۱۳۸۴)، اسدی و جمالی مهمویی، (۲۰۰۴)، عاطفه شریف (۱۳۸۴) نشاط (۱۳۸۵)، زندی روان (۱۳۸۶) و غیره در زمینه کاربرد آنتولوژیها در فرایند سازماندهی اطلاعات و مدیریت نگاشته شده است. در زمینه «پردازش زبان طبیعی و بازیابی اطلاعات» نیز مقالاتی از سوی دکتر جعفر مهرداد (مهراد ۱۳۸۷) تدوین یافته و طرح پژوهشی «ریشه ساز فارسی» در کتابخانه منطقه‌ای شیراز در دست اجرا می‌باشد. تفاوت میان طرحهای موجود و در دست اجرا با آنچه هم اکنون از سوی پیشنهاد دهنده طرح حاضر ارائه شده آن است که اکثر این واژه‌نامه‌ها و پیکره‌های زبانی^۹ شامل مجموعه‌ای از واژه‌های عمومی زبان فارسی می‌باشد، حال آنکه طرح حاضر اصطلاحنامه‌های تخصصی تدوین یافته در پژوهشگاه بویژه در حوزه «علوم پایه، مهندسی و کشاورزی» را «پیکره زبانی» این تحقیقات قرار داده که همه آنها مجموعه‌ای از اصطلاحات تخصصی این رشته‌ها می‌باشد و مهمتر آن که دارای روابط سلسله مراتبی منطقی و در نتیجه قابل تجزیه و تحلیل معنایی و تعیین مختصات معنایی^{۱۰} علاوه بر تجزیه و تحلیل ساختوازی (صرفی) می‌باشد. این ویژگی سبب می‌گردد که بتوان با کاهش همپوشانی این اصطلاحات در حوزه‌های تخصصی به سرعت به «شبکه واژگان علمی» دست یافت و «هستی شناسی‌های» مرتبط با این حوزه‌های تخصص که نسبت به اصطلاحنامه‌ها از انعطاف بیشتری برخوردار بوده و قابل گسترش و روزآمدسازی بیشتری است، طراحی گردد و در نهایت با ارائه «نقشه مفهومی»^{۱۱} هر یک از رشته‌ها بازیابی اطلاعات علمی - فنی به صورت خودکار و ماشینی درآید. علاوه بر این می‌توان از این نقشه مفهومی در بازنگری برنامه ریزی درسی رشته‌ای مرتبط و نیز ترسیم نقشه جامع علمی کشور بهره‌مند

^۵ - Karine Megerdomian

^۶ - Parsing

^۷ - tagged

^۸ - Shiraz Machine Translation Project

^۹ - Corpus

^{۱۰} - Semantic features

^{۱۱} - Concept Map

شد.

پروژه دیگر: «طراحی آنتولوژی فارسی بر اساس تزاروس اصفاء» در کتابخانه ملی ایران که برای اولین بار از طریق مهندسی مجدد در سال ۱۳۸۶ توسط دکتر فریبرز خسروی و همکارشان مهندس علیرضا وظیفه‌دوست به اجرا درآمده، ولی هنوز به نتیجه نهایی نرسیده است. و نیز پروژه: «بررسی عملی و نظری نظام ساده سازماندهی دانش در زبان فارسی» توسط دکتر محمد حسن‌زاده و سید مهدی حسینی بر اساس استاندارد بین‌المللی^{۱۲} اخیراً در دانشگاه تربیت مدرس انجام شده است و نمونه‌ای از کاربرد آن در اصطلاحنامه^{۱۳} اجرا شده است.

۴- سابقه پژوهش در جهان

ایجاد «پایگاه دانش واژگانی» مانند جمع‌آوری مجموعه‌ای از واژه‌نامه‌ها، اصطلاح‌نامه‌ها و شبکه واژگانی^{۱۴} می‌تواند به عنوان ابزار کلیدی در شکل‌گیری «هستی‌شناسی‌ها» محتوای یک متن علمی، بویژه مرتبط با پایگاه اطلاعات علمی و فنی باشد.

«نقشه مفهومی» براساس شکل‌گیری سلسله مراتبی واژگان، تدوین شده است. هستی‌شناسی اغلب مفاهیمی را در بر می‌گیرد که برای شناخت انسان از دنیای پیرامون وی، بنیادی و اساسی می‌باشند و دانش «پیش‌نمونه»^{۱۵} را بازنمایی می‌کنند. این که ساختار هستی‌شناسی براساس اعم و اخص یا سطوح بالا و پایین طراحی شوند، بیشتر به اهداف و حوزه مورد نظر بستگی دارد تا حل مشکلات مدیریت واژگان، زیرا تفات معنادار و واقعی میان این دو نوع هستی‌شناسی وجود دارد. هستی‌شناسی براساس سلسله مراتب و مختصات معنایی، سعی دارد حوزه معنایی خاصی، مبتنی بر مفاهیم دقیقاً تعریف شده، را به دست آورد. به عنوان مثال در علوم پایه (ریاضی، فیزیک، شیمی، زیست‌شناسی و زمین‌شناسی) دانش مربوطه به راحتی فرمول‌بندی می‌شود، زیرا در همه این حوزه‌ها نظام مفاهیم حاکم است که می‌توان آنرا به واسطه اصطلاحات علمی، با تعریف دقیق بیان نمود. در چنین مواردی سعی می‌شود مفاهیم مرتبط با اشیاء در عالم مقال و نیز به طور محض و انتزاعی و با الگوهای آرمانی و ساده‌شده‌ای از پدیده‌های واقعی در جهان ارائه شوند (گومز - پرز ۲۰۰۴)

«پایگاه‌های اطلاعات واژگانی همانند شبکه‌واژگان (شامل واژگان، اصطلاح‌نامه‌ها و فرهنگ‌های لغت) است که به قاعده‌مند نمودن معنی‌شناسی واژگانی (یعنی معنی واژه‌های یک یا چند زبان طبیعی)

^{۱۲} - SKOS

^{۱۳} - UNESCO

^{۱۴} - Word Net

^{۱۵} - Prototype

می‌پردازد. این معانی براساس شبکه مفاهیم و مبتنی بر فهم مشترک اکثریت جامعه علمی، و دارای تعاریف خاص، و در برگیرنده معانی مختلف متداول و کاربردهای خاص و ویژه و ترکیبات گوناگون واژه‌های مبتنی بر «پیش نمونه‌ها» و نیز مفاهیم پایه علوم است.

«وردنت» توسط میلر^{۱۶} و همکاران وی در سالهای ۱۹۹۰ تا ۱۹۹۵ به عنوان نمونه یک پایگاه وسیع اطلاعات واژگانی در زبان انگلیسی در دانشگاه پرینستون و براساس نظریه‌های روان‌شناسی طراحی و پیاده سازی شد. روانشناسی زبان حوزه میان رشته‌ای تحقیقاتی است که متمرکز بر علوم شناختی و توانش زبان شناختی است (فلبوم و میلر ۱۹۹۰)^{۱۷}. وردنت یک نوع سازماندهی اطلاعات واژگانی و روابط منطقی میان مفاهیم است، که از مشهورترین و پرکاربردترین فرهنگ‌های معنا بنیاد رایانه‌ای به‌شمار می‌رود. جورج میلر، روان‌شناس و محقق در زمینه واژگان که خود بنیان گذار وردنت است، آغاز شکل‌گیری وردنت را سال ۱۹۸۵ میلادی می‌داند (میلر ۱۹۹۸).

وردنت همانند یک فرهنگ معنا-بنیاد واژه‌ها را در قالب شبکه‌ای از روابط مفهومی و براساس واژگان ذهنی گویشوران یک زبان سازماندهی شده، ارائه می‌دهد (فلبوم ۱۹۹۸). روابط هم‌معنایی، شمول معنایی، رابطه جزء و کل و یا جزء واژگی، و تضاد یا تقابل معنایی در این شبکه از اصلی‌ترین روابط هستند. وردنت مبتنی بر معنی‌شناسی مولفه‌ای است که واژه‌ها را به مؤلفه‌های کوچکتر و زیر مقوله‌ای تجزیه می‌کند و روابط میان آن‌ها را ایجاد می‌نماید. واژه‌ها به مقولات دستوری اسم، فعل، صفت و قید تقسیم و با معیار معنای مشترک به مجموعه‌هایی^{۱۸} دسته بندی شده اند و هر واژه براساس معنای واحد در یکی از مجموعه‌ها جای می‌گیرد و چنانچه معانی متفاوتی داشته باشد در مجموعه‌های واژگانی جداگانه قرار می‌گیرد. در نتیجه روابط متفاوتی با دیگر واژه‌های آن مجموعه ایجاد می‌نماید. این مجموعه‌ها از واژه‌های «مترادف»^{۱۹} یا به عبارتی هم معنا تشکیل می‌شوند که در بافت زبانی، قابلیت جانشینی دارند. هم‌معنایی به دلیل ایجاد مجموعه‌ها، رابطه معنایی اصلی در وردنت می‌باشد. رابطه معنایی دیگری که در این شبکه اهمیت ویژه‌ای دارد، رابطه شمول معنایی است که سبب ایجاد ساختاری سلسله مراتبی یا درختی در مجموعه‌ها و به ویژه در میان اسامی است. علاوه بر این، دو رابطه معنایی جزء و کل یا جزء واژگی^{۲۰} در ایجاد روابط میان اسامی و رابطه تضاد یا تقابل معنایی در ایجاد شبکه واژگانی به کار گرفته شده است. هم اکنون زبان‌های زیادی علاوه بر انگلیسی

^{۱۶} - Miller

^{۱۷} - Fellbaum & Miller

^{۱۸} - Synsets

^{۱۹} - Synonyms

^{۲۰} - Meronymy

مانند فرانسه، آلمانی، ایتالیایی، اسپانیایی و چینی دارای وردنت هستند. همانگونه که گفته شد اولین «وردنت زبان فارسی» نیز در دانشکده برق و کامپیوتر، دانشگاه شهید بهشتی توسط دکتر مهرنوش شمس فرد در دست تهیه و اجرا می‌باشد.

هستی‌شناسی‌ها به طور گسترده در مهندسی دانش، هوش مصنوعی و علوم رایانه، مدیریت دانش، پردازش زبان طبیعی^{۲۱}، تجارت الکترونیک، بازیابی اطلاعات، طراحی پایگاههای اطلاعاتی، آموزش و برنامه‌ریزی درسی، و نیز در حوزه جدید «شبکه معنایی»^{۲۲} کاربرد دارند. ابداع شبکه معنایی مرحله‌ای دیگری در تکامل هستی‌شناسی‌ها را بوجود آورده است. بنابر نظریه برنرز-لی (Berners-Lee) (۲۰۰۱) شبکه معنایی، توسعه شبکه جهانی اطلاعات کنونی است که در آن، اطلاعات بر مبنای عناصر معنایی تعریف شده و با همکاری انسان و رایانه، ذخیره و اشاعه اطلاعات انجام می‌شود. در این همکاری، هستی‌شناسی‌ها به عنوان ابزار کلیدی توسعه شبکه معنایی در بازنمایی یک دامنه موضوعی و به عنوان الگوهای فرایندی استدلالی تبادل اطلاعات نقش مؤثری ایفا می‌کنند. دیویس^{۲۳} و همکاران (۲۰۰۷) نگاه جدیدی به آینده «شبکه جهانی اطلاعات»^{۲۴} دارند که در آن، تمامی دانش موجود به صورت قاعده مند رمزگذاری می‌شود تا عوامل هوشمند بتوانند به طریقی خود سامان به همه انسان‌ها خدمات اطلاعاتی ارائه دهند. این کار مستلزم تعریف رسمی و قاعده‌مند هستی‌شناسی‌ها مبتنی بر «وب» است تا بیان‌کننده دانش قابل فهم انسان و نیز عوامل هوشمند باشد و ارزش افزوده بر دانش و توسعه آن به صورت «فراداده‌ها»^{۲۵} خواهد بود، که در حال حاضر تعریف شده و قابل کاربرد در مجموعه‌های محدودی از تقاضاهاست. پیش‌بینی می‌شود که در دهه‌های آتی، ساختار دانش مورد درخواست بیشتر به صورت رسمی و به شکل نظام‌های هستی‌شناسی مبتنی بر «وب» و ارائه آنها به زبان‌های گوناگون، زمینه فعالی برای تحقیقات و مطالعات بیشتر باشد که حاصل آن تعداد قابل توجهی طرح‌های تحقیقاتی و تصویب معیارها و استانداردهاست که در قالب «آر.دی.اف»^{۲۶} و «اکس ام ال/آر.دی.اف»^{۲۷} و جدیدترین آن «او، دابلو، ال»^{۲۸} به عنوان زبانهای ارائه دانش به کار می‌روند. شبکه معنایی ارتباطات مفاهیم گوناگون را مقدور می‌سازد و برای عملکرد دقیق‌تر رایانه، نیاز به

^{۲۱} - NLP

^{۲۲} - Semantic Web

^{۲۳} - Davis

^{۲۴} - World Wide Web

^{۲۵} - Metadata

^{۲۶} - RDF

^{۲۷} - XML/RDF

^{۲۸} - OWL

ساختارمند نمودن اطلاعات و نیز اعمال قواعد معنایی دارد. بدین منظور بایستی از الگوهای یکسان معنایی استفاده شود و به این ترتیب توانایی تجزیه و تحلیل معنایی این شبکه، امکان دسترس پذیری به حجم بالایی از اطلاعات علمی- فنی را فراهم می‌سازد.

۵- اشتراک منابع اطلاعاتی و اکتشاف دانش

در عصر اطلاعات ثروت اصلی دانش است و صاحبان دانش دارای قدرت واقعی هستند. در این دوران به اشتراک گذاشتن منابع اطلاعاتی در مدیریت دانش اصلی کلیدی محسوب می‌گردد. از دیرباز در مطالعات دانشمندان بویژه حوزه زیست‌شناسی رده‌بندی نظام‌مندی^{۲۹} بویژه برای طبقه‌بندی اسامی و نامهای جانوران و گیاهان (لینه، قرن ۱۸) با توجه به روابط سلسله مراتبی آنها متداول بوده است. این شیوه رده‌بندی^{۳۰} در فناوری اطلاعات نیز بکار گرفته شد و طبق این تعریف: تاکسونومی (رده‌بندی) شامل طبقه‌بندی مدخل‌های اطلاعاتی است که به شکل سلسله مراتب^{۳۱} و براساس روابط اعم و اخص در جهان واقعی ارائه می‌شود، که در آن زیر طبقات از گره اصلی انشعاب یافته و از بالا به پائین بصورت روابط پدر و فرزندی تقسیم‌بندی می‌شوند. این نوع نظام طبقه‌بندی به تاکسونومی اعم و اخص معروف هست. این شیوه تقسیم‌بندی سلسله مراتبی تاکسونومی در واقع اسکلت معنایی را در آنتولوژیها (هستی‌شناسی‌ها) تشکیل می‌دهند و علاوه بر آن در هستی‌شناسی‌ها روابط ریزتر و دقیق‌تر معنایی به شکل مختصات معنایی و ارزشگذاری معنایی این مختصات به آنها اضافه می‌شوند (Daconta, et. al, 2003).

بدین ترتیب تاکسونومی‌ها ساختار زیربنایی سیستم اطلاعاتی را فراهم می‌سازند و آنتولوژیها به جزئیات این عناصر و روابط معنایی می‌پردازند. تاکسونومی همانند تزاروس و آنتولوژی، راهی است برای ساختارمند نمودن داده‌های اطلاعاتی و ارائه روابط معنایی میان مدخل‌های آن. در واقع جهان نمی‌تواند بدون رده‌بندی و طبقه‌بندی مفاهیم باشد، زیرا ذاتاً تفکر انسان براساس این طبقه‌بندی پدیده‌ها شکل گرفته است.

از اینروست که می‌توان تاکسونومی را چنین تعریف نمود:

«راهی برای طبقه‌بندی و رده‌بندی مجموعه‌ای از پدیده‌ها و روابط معنایی میان آنها که به شکل سلسله مراتبی انجام می‌پذیرد.» «سلسله مراتب» ساختاری درختی دارد و همانند یک درخت دارای ریشه‌ها و

^{۲۹} - Systematic Classification

^{۳۰} - Taxonomy

^{۳۱} - Hierarchy

شاخه‌هاست و هر کدام از نقاط اتصال شاخه‌ها به عنوان گره^{۳۲} نام‌گذاری شده است. در محیط وب، تاکسونومی‌ها می‌توانند بویژه در جستجوی و بازیابی اطلاعات ویژه و مورد علاقه کاربران^{۳۳} کمک نمایند. تاکسونومی‌ها به دارندگان پایگاه‌های اطلاعاتی نیز کمک می‌کنند تا اطلاعات مورد علاقه کاربران را نیاز سنجی نمایند و اطلاعات ویژه و مورد نظر آنها را طبقه‌بندی نموده و اطلاعات گزینشی^{۳۴} به آنها ارائه دهند. از آنجا که تاکسونومی‌ها مبتنی بر معنی‌شناسی و مفهوم‌شناسی است، ماشین‌های جستجو^{۳۵} و دیگر ابزارهای جستجو را قادر می‌سازند تا با بکارگیری این سیستم رده‌بندی اطلاعات مورد نیاز کاربران را سریع‌تر و دقیق‌تر و با کیفیت بالاتر دسته‌بندی و ارائه نمایند. اگر بخواهیم میان تاکسونومی‌ها، تزاروس‌ها و هستی‌شناسی‌ها تفکیک قائل شویم، می‌توانیم بگوئیم که تاکسونومی‌ها دارای تقسیم‌بندی سلسله‌مراتبی و ساختار پدر و فرزندی مفاهیم هستند و تمایز معنایی کمتری ایجاد می‌کنند. در حالی که تزاروس‌ها علاوه بر آن اصطلاحات و مترادف‌ها و روابط اعم و اخص و وابسته و ارجاعات را نیز نشان می‌دهند و از اینرو نسبت به تاکسونومی‌ها تمایز معنایی دقیق‌تری میان مفاهیم قائل می‌شوند. آنتولوژیها در یک مرحله بالاتر با ارائه الگوی مفاهیم^{۳۶} تقسیم‌بندی معناشناختی دقیق‌تری میان مفاهیم قائل هستند که براساس یک نظریه منطقی^{۳۷} استوار است و از دانش غنی‌تر و پیچیده‌تر با تمایزات معنایی بیشتر برخوردار است (Daconta et.al. 2003).

^{۳۲} - Node

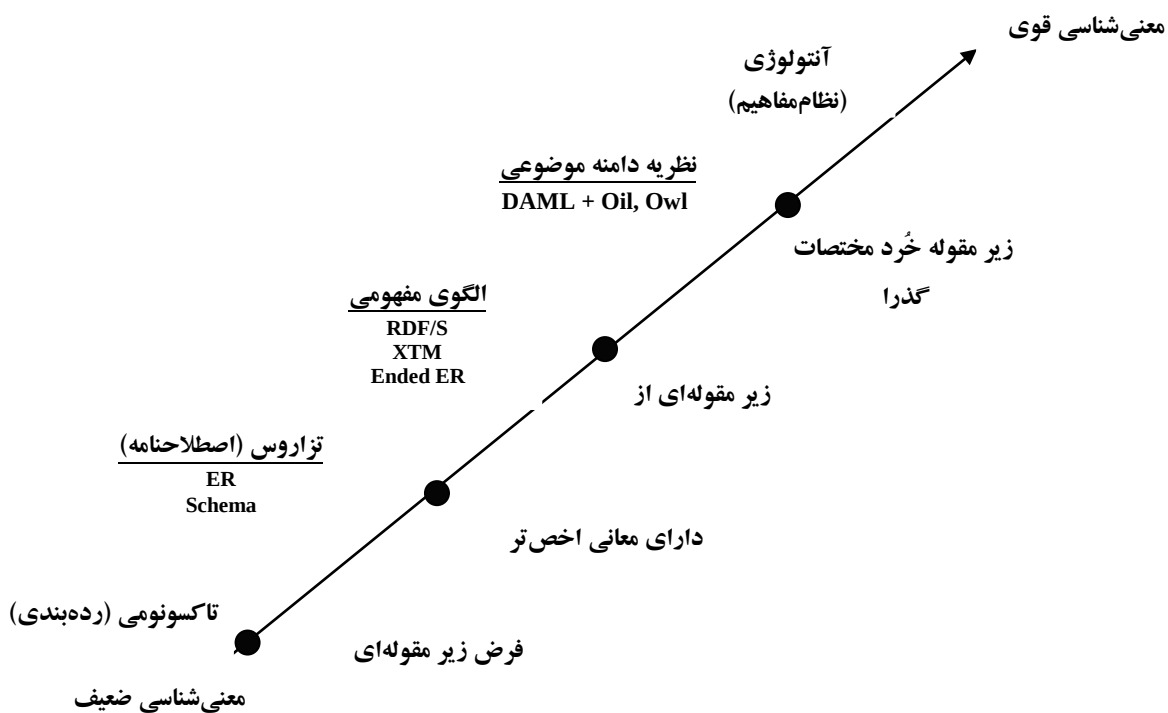
^{۳۳} - Users

^{۳۴} - SDI (Selected Dessimination Information)

^{۳۵} - Search Engines

^{۳۶} - Conceptual Model

^{۳۷} - Logical Theory



نمودار ۱. انواع ابزار رده‌بندی معناشناختی (Daconta et. al. 2003)

در نمودار (۱) تاکسونومی‌ها با داشتن ساختار تقسیم‌بندی پدر و فرزندی مفاهیم از تمایزات معنایی نسبتاً ضعیف‌تری برخوردارند. در برخی موارد روابط معنایی براساس طبقات و زیر طبقات و در مواردی دیگری براساس کل و جزء تعریف می‌شوند. بطور کل می‌توان در مورد روابط معنایی میان اجزاء معنایی تشکیل دهنده تاکسونومی و طبقه‌بندی آن گفت که نسبت به تزاروس‌ها و آنتولوژی‌ها از تمایزات معنایی ضعیف‌تری برخوردارند. اما تاکسونومی‌هایی که قوی‌تر و دقیق‌تر هستند، این طبقه‌بندی را با استفاده از ابزار منطقی تمایزات معنایی دقیق‌تری میان مفاهیم ایجاد کرده و از این‌روست که می‌توانند به عنوان اسکلت معنایی آنتولوژی‌ها مورد استفاده قرار گیرند.

در این نمودار گره بعدی به تزاروس‌ها اختصاص یافته است. در تزاروس‌ها براساس روابط معنایی میان اصطلاحات (واژه‌ها و عبارات) ساختاری همانند تاکسونومی شکل گرفته است. بدین ترتیب تزاروس‌ها نوعی رده‌بندی سلسله‌مراتبی و نیز روابط اعم و اخص و وابسته میان اصطلاحات را ارائه می‌دهند.

براساس استاندارد اصطلاحنامه تک‌زبان^{۳۸} تعریف اصطلاحنامه عبارت است از: «واژگان کنترل شده که براساس نظامی شناخته شده دسته‌بندی شده است و ساختار آن براساس روابط معنایی هم‌ارز، سلسله‌مراتبی، اعم و اخص و روابط وابسته مفاهیم شکل گرفته است. اهداف اولیه تزاروس، تسهیل بازیابی اسناد و مدارک و دستیابی به انسجام در نمایه‌سازی منابع مکتوب و غیر مکتوب و اسناد و مدارک علمی است. از اینروست که تزاروس با داشتن واژگانی کنترل شده و استاندارد پشتیبان بازیابی اطلاعات قرار می‌گیرد تا «جستجوگر اطلاعات» را در گزینش اصطلاحات درست و منطبق با مفاهیم و دستیابی به داده‌های مورد نیازش هدایت کند.

کاربردیک اصطلاحنامه تضمین می‌کند که:

- مفاهیم به شیوه‌ای منسجم توصیف شوند.
- کاربران با تجربه بتوانند به آسانی جستجوهای خود را طوری اصلاح نمایند که به نیازهای اطلاعاتی خود سریعتر دسترسی یابند.
- کاربران نیازی به آشنایی کامل به اصطلاحات فنی و تخصصی و ریز مسائل آن نداشته باشند.

هرچند اکثر محققین با اولین تزاروس روزه (Roget's Thesaurus, 1911) آشنایی دارند و پس از آخرین چاپ در سال ۱۹۱۱، تاکنون در سطح وب نیز در دسترس است. اما مهندسين اطلاعات به منبع قابل اطمینانی مانند «وردنت» دسترسی و آشنایی بیشتر دارند (Fellbaum, 1998; Miller et.al, 1993).

وردنت براساس نظریه‌های روانشناسی زبان^{۳۹} و واژگان ذهنی انسان طراحی شده و در دانشگاه پرینستون آمریکا راه‌اندازی شد. وردنت به شکل برخط^{۴۰} در دسترس است و از سوی برنامه‌نویسان نرم‌افزاری دانش تزاروسی آن در برنامه‌های کاربردی مورد استفاده قرار گرفته است. در وردنت برای یک واژه تعریف یا تعاریفی گوناگون ارائه می‌شود.

تعاریف متمایز یک واژه در زبان‌شناسی به عنوان معانی متفاوت آن واژه تلقی می‌شود و اصطلاحاً به آن چند معنایی^{۴۱} گویند.

یک واژه (یا یک عبارت) در وردنت نیز اطلاعات معنایی خاص مربوط به خود را دارد که با استاندارد ایزو ۲۷۸۸^{۴۲} سنجیده می‌شود:

^{۳۸} - ANSI/NISO

^{۳۹} - Psycholinguistics

^{۴۰} - Online

^{۴۱} - Polysemy

^{۴۲} - ISO 2788

مترادف‌ها: ^{۴۳} گره‌هایی که در تاکسونومی روابط یکسانی دارند وزیر یک «اسم» می‌آیند.

اصطلاحات اعم: ^{۴۴} گره‌هایی که در جایگاه عام‌تری در تاکسونومی قرار می‌گیرند و روابط معنایی ^{۴۵} را نشان می‌دهند.

اصطلاحات اخص: ^{۴۶} گره‌هایی که در جایگاه فرزندی یا اخص‌تر ^{۴۷} و زیر مجموعه یک واژه قرار می‌گیرند.

در وردنت روابط متعددی میان مترادف‌ها وجود دارد که مجموعه این مترادف‌ها و روابط معنایی آنها «سین ست» ^{۴۸} نامیده می‌شود.

این بدین معناست که واژه‌ها یا عبارات تمایز یافته که اصطلاح نامیده می‌شوند و از لحاظ معنایی در سطح انتزاعی مترادف محسوب می‌گردند (یعنی روابط پدر و فرزندی ندارند، بلکه هم سطح هستند) به عنوان یک مجموعه مترادف دسته‌بندی می‌شوند. از اینرو یک مجموعه مترادف‌ها همانند یک مفهوم عمل می‌کند و برای انسان و یا برای سیستم اطلاعاتی و بازنمایی دانش یک ساختار ذهنی واحد را تشکیل می‌دهند که در پشت یک اصطلاح قرار می‌گیرد و دلالت ذهنی آنرا بازنمایی می‌کند. مفهوم نیز همانگونه که می‌دانیم نشانگر یک شیئی یا پدیده واقعی و یا بالقوه در جهان هستی است. از اینرو میان اصطلاح و مفهوم در بازنمایی معنایی دانش تمایز قائل می‌شویم.

یک اصطلاح ^{۴۹} برجسی ^{۵۰} است که برای بازنمایی معنایی و محتوایی یک مفهوم بکار میرود و در واقع آن معنی و مفهوم بوسیله اصطلاح که نمایه‌ای از مفهوم، ویژگیهای معنایی، روابط معنایی اجزاء آن و روابط معنایی با دیگر مفاهیم را نمایه می‌کند. در نمودار (۱) در مرحله بعد الگوی مفهومی ^{۵۱} که از بازنمایی معنایی بیشتری برخوردار است، قرار می‌گیرد. وقتی از تاکسونومی‌ها و اصطلاحنامه‌ها به سوی الگوی مفهومی و نظریه منطقی مفاهیم حرکت می‌کنیم در واقع به هستی‌شناسی‌ها می‌رسیم. سوال این است که: یک مدل مفهومی چیست؟ مدل مفهومی، الگویی است از مفاهیم نظام یافته یک حوزه موضوعی علوم و دانش که در برخی موارد دامنه ^{۵۲} نیز نامیده می‌شود که نشانگر مدخل‌های

^{۴۳} - Synonyms

^{۴۴} - Hypernyms

^{۴۵} - Broader Term

^{۴۶} - Hyponyms

^{۴۷} - Narrower Term

^{۴۸} - Synsets

^{۴۹} - Term

^{۵۰} - Label

^{۵۱} - Conceptual Model

^{۵۲} - Domain

اولیه (اشیاء و پدیده‌های یک دامنه) روابط ساختاری و معنایی میان این مدخل‌های واژگانی و ارزشگذاری مختصات معنایی و روابط آنها و در برخی موارد شامل قواعد حاکم بر مدخل‌ها، روابط و ممیزه‌های معنایی آنها نیز می‌شود.

در یک الگوی مفهومی حقیقتاً امکان تعریف روابط میان طبقات و زیر طبقات یعنی قواعد حاکم بر روابط پدر و فرزندی حاکم است. زبانهای برنامه‌ریزی هدفمند مانند UML و قواعد منطقی حاکم بر آنها به اندازه کافی غنی هستند تا معنی شناسی زیر طبقات و نیز روابط منطقی میان طبقات را بیان نمایند. آنچه دارای اهمیت است این که تعاریف یک طبقه، زیر طبقه و زیر زیر طبقه از لحاظ معنایی کاملاً مشخص و معین شوند و در سطح الگوی برتر^{۵۳} متمایز گردند، و این سطحی است که ما به عنوان دامنه^{۵۴} موضوعی و نظام حاکم بر آن ترسیم می‌نمائیم. در سطح الگوی برتر اغلب زبان‌ها ساختار مفهومی دارند. آنچه که در سطح زبان ساختار مفهومی بیان می‌شود ما را قادر می‌سازد تا اشیاء و پدیده‌ها را به آن زبان بیان کنیم و زبان خاص و کارکرد ویژه آن در الگوی مفهومی خودمان را تعیین نمائیم.

هستی‌شناسی‌هایی که براساس نظریه منطقی ارائه می‌شوند، مستقیماً قابل پردازش هستند و تعبیر و تفسیر معنایی آنها بوسیله نرم‌افزار صورت می‌پذیرد. بسیاری از شیوه‌های مهندسی هستی‌شناسی و بازنمایی دانش، خواهان ساخت و طراحی هستی‌شناسی‌ها براساس نظریات منطقی هستند. این نظریه‌های منطقی براساس گزاره‌های منطقی معادلات و روابط دو جانبه این قواعد استوارند، که به کمک این دو می‌توان دامنه موضوعی را با در نظر گرفتن روابط منطقی طراحی و ارائه نمود. مجموعه گزاره‌ها، قواعد دو طرفه و معادلات منطقی منجر به نظریه منطقی می‌شوند. در یک نظریه منطقی می‌توان معنی شناسی یک الگوی مفهومی را تا بالاترین سطح ممکن بیان نمود.

رابطه زیر طبقه^{۵۵} می‌تواند رابطه غنی‌تری باشد، زیرا به صورت رابطه زیر طبقه جزء به جزء و ویژگی گذرا^{۵۶} بیان می‌شود. رابطه میان طبقه و زیر طبقه^{۵۷} و نیز زیر زیر طبقه آن را می‌توان رابطه‌ای کامل و جامع دانست. مشابه آن می‌توان تمایزات معنایی ظریفی میان نوع رابطه‌های معنایی و مختصات آنها و «جنبه‌ها»^{۵۸} روابط میان فراداده‌ها را ارائه نمود. پس از ارائه تفاوت میان تاکسونومی، تزاروس و

^{۵۳} - Meta-Model

^{۵۴} - Domain

^{۵۵} - Subclass

^{۵۶} - Property of Transitivity

^{۵۷} - Superclass

^{۵۸} - Facets

آنتولوژی می‌توان تعریفی از هستی‌شناسی داد.

یک هستی‌شناسی معرف واژگان همگانی و مفاهیم (معانی) آنهاست که در توصیف و بازنمایی حوزه‌ای از دانش بکار می‌رود و از اینرو معانی اصطلاحات را معیارسازی می‌نماید. آنتولوژیها از سوی متخصصین، پایگاههای اطلاعاتی و دیگر سیستم‌های رایانه‌ای بکار گرفته می‌شوند تا اطلاعات موضوعی و تخصصی دامنه (حوزه تخصصی دانش مانند پزشکی، شیمی، فیزیک و...) را به اشتراک بگذارند. آنتولوژیها شامل تعاریف مفاهیم پایه یک دامنه موضوعی و روابط معنایی میان آنهاست که برای رایانه کاربرد داشته باشد و دانش موجود در دامنه را رمزگذاری کرده و قابل استفاده و بهره‌برداری مجدد توسط متخصصین و محققین نماید.

یک آنتولوژی (هستی‌شناسی) شامل این اجزاء می‌گردد:

۱- طبقات: (پدیده‌های عام) در دامنه‌های موضوعی مورد علاقه متخصصین

۲- شواهد:^{۵۹} یعنی پدیده‌های ویژه (پیش نمونه‌ها)

۳- روابط معنای: روابط میان اشیاء و پدیده‌ها

۴- مختصات: (ارزشگذاری مختصات) پدیده‌ها

۵- کارکرد و فرآیندهای: دربرگیرنده پدیده‌ها

۶- محدودیت‌ها و قواعد: حاکم بر پدیده‌ها (Daconta et. al. 2003)

۶- تعامل میان هستی‌شناسی و واژگان

پیوند میان «بازنمایی دانش» و «زبان طبیعی» به سرعت به صورت موضوعات محوری مباحث «مهندسی دانش»^{۶۰} (KE) و زبان‌شناسی رایانه‌ای^{۶۱} (CL) درآمد، که در این زمینه «هستی‌شناسی»^{۶۲} نقش ویژه‌ای در ارتباط میان این دو حوزه داشت و دارای مراحل مشخصی بود:

الف: کاربرد زبان طبیعی در بازنمایی دانش،^{۶۳}

ب: کاربرد نظریه‌های بازنمایی دانش و کاربرد زبان طبیعی در بازیابی اطلاعات

هرچند ضرورت دستیابی به وب معنایی شکل‌گیری هستی‌شناسی‌ها را می‌طلبد، لکن فرصتی مناسب و استثنایی برای ارائه نتایج تحقیقاتی و کاربرد واقعی اطلاعات با استفاده از مهندسی زبان را فراهم

^{۵۹} - Instances

^{۶۰} - Knowledge Engineering

^{۶۱} - Computational Linguistics

^{۶۲} - Ontology

^{۶۳} - Knowledge Representation

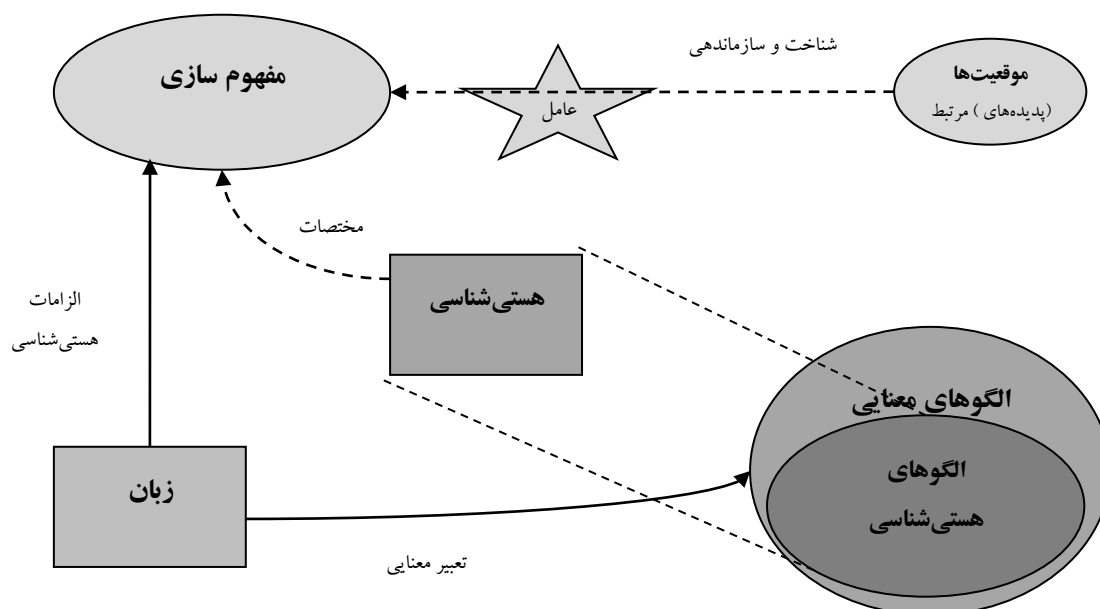
ساخت.

شیوه ارائه اطلاعات و طراحی بازیابی دانش از طریق کاربرد زبان طبیعی وجوه مشترک میان «هستی شناسی‌ها» و «منابع واژگانی» است، که از طرفی تلفیق و ادغام آنها را ضروری می‌گرداند و از طرف دیگر محدودیت‌هایی در کاربرد هستی شناسی‌ها ایجاد می‌کند که عبارتند از:

الف. کاربرد هستی‌شناسی در ساختار غیر وابسته به زبان در منابع واژگانی رایانه‌ای چند زبانه.

ب. کاربرد ابزاری قدرتمند برای ارائه گنجواژه‌ها در حوزه پردازش زبان طبیعی است (NLP)^{۶۴} که در رفع ابهامات واژگانی بسیار مؤثر واقع می‌شود.

«هستی شناسی‌ها» بطور کلی بعنوان تخصیص ویژگی‌های معنایی مشترک مفاهیم در مفهوم سازی^{۶۵} تعریف می‌شوند (Huang, et. al. editors, 2010).



نمودار ۲. مفهوم سازی در هستی‌شناسی (Guarion, 1999)

براین اساس «مفهوم سازی» شامل دانش غیرصوری^{۶۶} است. که از طریق تجربه، مشاهده و کاوش^{۶۷}.

^{۶۴} - Natural Language Processing

^{۶۵} - Conceptualization

^{۶۶} - Informal

^{۶۷} - Introspection

ویژگی آن^{۶۸}، به شکل دانش زبانی ارائه و نمایه^{۶۹} می شود (نمودار ۲).

اگر بپذیریم که واژگان شامل دانش زبان شناختی است که فرد از تجارب زبانی خود استخراج می کند، تعریف بالا شامل هستی شناسی نیز می گردد. هرچند ویژگی^{۷۰} یک مجموعه واژگان این است که مفاهیم از نظر زبانشناختی در قالب کلمات (واژگان) رمز نگاری می شوند.

بمنظور فهم و درک تفاوت های زیربنایی^{۷۱} می توان نگاهی دقیق تر به عناصر اصلی ساخت و ابداع هستی شناسی ها یعنی: «مفهوم سازی» و «تخصیص ویژگیها»^{۷۲} انداخت. آنچه که میان واژگان و هستی شناسی ها در تفاسیر و تعبیر دقیق تر هستی شناسی ها و منابع معنا شناسی واژگانی دیده می شود، این که این دو ظاهراً به قدر کافی مشابهت دارند تا بتواند در برخی موارد به یکدیگر تبدیل و یا در هم ادغام شده و بصورت منابع تلفیقی مورد استفاده قرار گیرند.

هرچند، واژگان واقعاً همان هستی شناسی ها نیستند. (Hirst, 2004) به عنوان مثال، مترادف ها و شبه مترادف ها روابط معنایی مهم در معنی شناسی واژگانی هستند و لکن برای آنها در هستی شناسی صوری، که مفاهیم بایستی بدون ابهام باشند و اصطلاحات مترادف زیر یک مفهوم دسته بندی شوند، جایگاهی وجود ندارد. از دیدگاه هستی شناختی، موضوع مترادف ها در بازنمایی هستی شناسی کاملاً شفاف است. زمانی به رفع مشکلات هستی شناسی صوری می پردازیم که مسئله مترادف ها حل شده باشد. نمونه دیگر اطلاعات مربوط به کاربرد واژه است (مانند: ثبت اطلاعات واژگانی) که در مجموعه واژگان، ارائه می شود، اما ربطی به هستی شناسی ها ندارد.

علاوه بر این منابع زبانشناختی مانند واژگان بر مبنای تعاریف زبانشناختی شکل می گیرد و نه براساس مفاهیم زیر ساخت آن، در حالی که هستی شناسی واژگانی در برگیرنده مفاهیم زیر ساخت آن نیز است. دانشی که این منابع کسب می کنند ماهیتی کاملاً متفاوت دارند، و به منظور مدیریت این نوع هستی شناسی ها نیاز به میانجی «آنتولکس» یعنی مجموعه هستی شناسی و واژگان است. جزئیات این تمایزات را همانگونه که در بخش های ذیل آمده است مورد بررسی قرار می دهیم.

در این مسئله از نظر کاربردی، مهمترین تمایز، تفاوت قائل شدن میان هستی شناسی های صوری^{۷۳} و واژگانی^{۷۴} است. هستی شناسی های صوری براساس روابط منطقی بدست آمده اند و از نظر صوری،

^{۶۸} - Specification

^{۶۹} - Representation

^{۷۰} - Crucial

^{۷۱} - Subtle

^{۷۲} - Specification

^{۷۳} - Formal Ontology

^{۷۴} - Lexical Ontology

ساختار مفهومی خوش ساخت دارند، در حالی که هستی‌شناسی‌های واژگانی بر مبنای زبان طبیعی انسان استوارند و از اینرو «موسوم به واژگانی هستند» که اغلب از لحاظ صوری ساختار مفهومی چندان دقیقی ندارند. در پشت این تفاوت صوری / واژگانی، طبقه‌بندی لایه زیرین می‌تواند با تحلیل‌های دقیق این منابع اصطلاح‌شناسی نیز مفهوم‌سازی و تخصیص معنایی یابند.

در سطح اصطلاح‌شناختی، واژگان رایانه‌ای و منابع واژگانی و وابسته از یکدیگر متمایز می‌گردند. هرچند چون مسئله «پردازش زبان طبیعی»^{۷۵} و «وب معنایی»^{۷۶} نیز در اینجا مطرح می‌شوند منابع واژگانی ما بایستی ماشین خوان بوده و از اینرو مشابه واژگان کامپیوتری محسوب می‌گردند. سرانجام، چون روابط معنایی عناصر ماهتاً واژگان کامپیوتری هستند، ما نیز واژگان وابسته را بعنوان مترادف آنها قرار می‌دهیم.

«آنتولکس» وجه مشترک و میانجی^{۷۷} میان هستی‌شناسی و واژگان نشأت گرفته از ویژگی‌های معنایی کاملاً تمایز یافته و مرتبط بین دو حوزه است. «واژگان» مجموعه‌ای از واژه‌ها و «هستی‌شناسی» مجموعه‌ای از مفاهیم است که وجه مشترک هر دو آنها «مفهوم‌سازی» است. برای کاربرد آنها در فناوری پردازش زبان طبیعی، واژگان شامل وجه مشترک میان انسان و رایانه می‌شود. برای کاربرد وب معنایی، یک هستی‌شناسی رایانه را قادر می‌سازد تا دانش را بطور مستقیم پردازش کند. در این بافت است که «آنتولکس»^{۷۸} وجه مشترک موضوع تحقیقات پیرامون ارتباط میان دانش انسانی و دانش وب معنایی قرار می‌گیرد (Berners Lee et. Al. 2001).

۱-۶- مفهوم‌سازی^{۷۹}

ماهیت مفهوم‌سازی عمدتاً وابسته به چگونگی تلفیق و یا چگونگی ابداع و اختراع آن است. مفهوم‌سازی فرایندی است که به استخراج و تعمیم اطلاعات مرتبط با تجربیات اشخاص منجر می‌شود. مفهوم‌سازی به تنهایی نیز اطلاعات نسبی مرتبط را شامل می‌گردد. مفهوم‌سازی به موقعیت خاص و زبان بازنمایی مفهوم وابسته است، تا جایی که مربوط به خود بازنمایی نباشد. براساس تحقیقات حاصله مفهوم‌سازی پس از طی مراحل تخصیص ویژگی‌های معنایی براساس مطالعات علوم شناختی و توصیف معنایی مستقیم قابل دسترس است (Shalley and Zaeffer 2007).

هر نوع «مفهوم‌سازی» وابسته به یک عامل واحد است. به عنوان مثال تولیدات ذهنی^{۸۰} نمایانگر جهان

^{۷۵} - NLP

^{۷۶} - Semantic Web

^{۷۷} - Interface

^{۷۸} - Ontalex

^{۷۹} - Conceptualization

بینی اکتسابی فرد (عامل) است و بر این اساس هستی‌شناسی‌ها، همان ویژگی‌های زبانشناختی فعالیت‌های ذهنی فرد است. در اینجا «واژگان ذهنی»^{۸۱} مطرح است که بیش از یک زبان صوری و واژگان آن است و الزاماً نباید یک زبان طبیعی باشد، اما می‌تواند به عنوان مثال زبان منطقی حاکم بر ویژگی‌ها و ارتباطات معنایی واژگان ذهنی یک نظریه باشد.

دو عامل هوش انسان و هوش مصنوعی تعیین می‌کنند که آیا مفهوم سازی حاصل می‌تواند به اشتراک گذاشته شود یا نه و آیا مفید و با ارزش و پذیرفتنی و قابل انتقال است یا خیر. زبان وسیله منحصر بفرد ارائه مفاهیم است، زیرا عامه مردم بطور معمول، از زبان طبیعی استفاده می‌کنند و جوامع دارای عاملین چندگانه شامل کامپیوترها، یا ربات‌ها و انسانها، به زبان صوری قابل فهم و درک برای ماشین (کامپیوتر) نیاز دارند.

برای مفیدتر بودن "مفهوم سازی" لازم است که این امر میان انسانها به اشتراک گذاشته شود و در این مورد توافق حاصل شود، حتی اگر این توافق فقط تلویحی باشد.

به عبارتی دیگر، مفهوم سازی که در زبان طبیعی بازنمایی می‌شود، مجموعه‌ای از فرآیندهاست و یک فرآیند واحد نیست. اطلاعات حاصله تعریف شده از سوی جمعی از گویشوران زبان است.

فیلسوفان زبان می‌پنداشتند که در مرحله اول داده‌های زبانی و تحلیل معنایی آنها می‌تواند برای مفهوم سازی و ساخت آنتولوژی زبان طبیعی بکار گرفته شود. در گذشته معنی‌شناسان واژگانی عمدتاً منابع واژگانی را بعنوان زیربنا و زمینه مفهوم سازی بکار می‌گرفتند.

متخصصین علوم شناختی دامنه آنرا به منابع اطلاعاتی دیگر نیز گسترش داده‌اند که ممکن است شامل اطلاعات شنیداری و دیداری نیز باشد. در اینجا است که تمایز میان هستی‌شناسی زبانشناختی و هستی‌شناسی مفهومی مشخص می‌شود و به منابع اطلاعات زبانشناختی محدود نمی‌گردد. این نوع هستی‌شناسی که اهمیت مفهوم سازی گویشور را مورد تاکید قرار می‌دهد به هستی‌شناسی توصیفی^{۸۲} معروف است که در مقابل "هستی‌شناسی بازنگر"^{۸۳} قرار می‌گیرد. «استراوسن» هستی‌شناسی توصیفی، زبان طبیعی و شم زبانی را بعنوان منابع مهم دانش هستی‌شناسی و تحلیل داده‌های آن می‌داند، در حالی که هستی‌شناسی بازنگر این موضع‌گیری را رد می‌کند و ترجیح می‌دهد که ماهیت روابط معنایی مفاهیم را بدون وابستگی به عامل مفهوم سازی یعنی گویشور زبان به دست آورد. در

^{۸۰} - Mental Product

^{۸۱} - Mental Lexicon

^{۸۲} - Descriptive Ontology

^{۸۳} - Revisionary Ontology

هستی‌شناسی واژگانی، مفهوم سازی مبتنی بر مقولات زبان‌شناختی است و یا دقیق‌تر بگوییم مبتنی بر اطلاعاتی است که در منابع واژگانی مانند فرهنگ‌ها و تزاروس‌ها (اصطلاحنامه‌ها) یافت می‌شوند (Masolo et. al. 2003, Guarina, 1999).

در بسیاری از موارد وجوه مشترک مؤلفه‌های معنایی عمدتاً مؤلفه‌های دانش زبان‌شناختی را نیز در برمی‌گیرد و نیز در بسیاری از موارد دانش دایره‌المعارفی و یا دانش جمعی را نیز شامل می‌شود. «هستی‌شناسی‌های واژگانی»^{۸۴} از نظر جایگاه خاص در واژگان ذهنی و شناخت بشر بسیار قابل توجه‌اند. البته دو دیدگاه در مورد واژگان وجود دارد. واژگان به معنی مجموعه‌ای از اطلاعات زبان‌شناختی مفاهیم، و از نظر علوم شناختی انباشته‌ای از دانش فردی است که به راحتی می‌توان آنها را در قالب واژگان بازیابی نمود.

هستی‌شناسی‌های کاربردی در مهندسی اطلاعات که مفهوم سازی آنها براساس تجارب مشترک میان خبرگان^{۸۵} رشته‌های علمی صورت می‌گیرد به پردازش زبان مربوط می‌شوند و این مسئله که این هستی‌شناسی چگونه با هستی‌شناسی‌های عام‌تر تلفیق شوند، نیز یکی دیگر از زمینه جالب مطالعاتی می‌باشد.

و سرانجام مفهوم سازی زبان‌شناختی که مرتبط با یک زبان^{۸۶} و یا چند زبان^{۸۷} می‌شوند. که هرچند مبتنی بر زبان است، اما از تعابیر معنایی معطوف به ملاحظات میان‌زبانی به مفهوم سازی منجر می‌شود و کمتر وابسته به ویژگی‌های فردی است. مسئله مهم این است که آیا مفهوم سازی براساس زبانهای مختلف مناسب‌تر است یا نه و اگر چنین نیست چگونه باید این اطلاعات زبانی بکار گرفته شوند. مجموعه واژگان چندزبانه عمدتاً برای رسیدن به اهداف کاربردی مشخص و معین، اهمیت دارند، اما توسعه آنها نیز به ایجاد و تعیین روابط پیچیده میان زبان، فرهنگ و اندیشه کمک می‌کند.

همچنان این مسئله در زمینه علوم شناختی و در زمینه «پردازش زبان» مطرح است که آیا نیازی به تمایز میان «سطح مفهومی» که فرض می‌شود غیر وابسته به زبان است و سطح معنایی که می‌تواند عمیقاً متأثر از زبانها باشد، وجود دارند یا خیر. این نوع مسائل در مباحث علوم شناختی و نیز کاربردهای «پردازش زبان» به چالش کشیده می‌شود. فرآیند مفهوم سازی برای ساخت هستی‌شناسی بسیار سرنوشت ساز است (Huang, C. et. al., 2010).

^{۸۴} - Lexical Ontologies

^{۸۵} - Experts

^{۸۶} - Monolingual Linguistic Ontology

^{۸۷} - Multilingual Linguistic Ontology

۶-۲- مختصات^{۸۸}

مرحله دوم عملیات «تعیین مختصات معنایی» است، زیرا هستی‌شناسی ویژگیهای مفاهیم را در قالب زبان طبیعی ارائه می‌کند. فارغ از سطح پیچیدگی و صراحت (شفافیت)، هستی‌شناسی‌ها در تعیین ویژگیهای معنایی مفاهیم وابسته به معنی‌شناسی زبان (البته نه وابسته به هر زبان طبیعی، بلکه زبانی که بتواند فرمول‌بندی هستی‌شناسی را ارائه دهد) به عنوان عامل زیربنایی و اساس ارتباط و تبادل نظر است و پلی برای فهم مشترک انسانها ایجاد کرده است. ماهیت این زبان مارا به دومین منبع تمایز دهنده هستی‌شناسی‌ها هدایت می‌کند. هستی‌شناسی‌های صوری^{۸۹} در قالب یک زبان صوری (ساختگی) هستی‌شناسی‌های غیر متداول ارائه می‌کنند. به عنوان مثال در قالب زبان طبیعی و یا زبانی نیمه صوری یا تلفیق هر دو، هستی‌شناسی را ارائه می‌کند. یکی از مهمترین وجوه این تمایز، ابهام زدایی کامل در هستی‌شناسی‌های صوری است در حالی که در نیمه صوری ممکن است ابهاماتی وجود داشته باشد.

هستی‌شناسی‌ها ممکن است کاملاً^{۹۰} مجرد و غیرقابل انعطاف پذیر و محدود باشند، حتی اگر در قالب زبان طبیعی بیان شده باشند و صرف صوری بودن نمی‌تواند به تنهایی سبب عدم انعطاف پذیری و مجرد و محدود بودن آن باشد. هستی‌شناسی‌های واژگانی ویژگیهای معناشناختی واژگان را در منابع واژگانی به شکل صوری یا نیمه صوری مانند وردنت بکار می‌گیرند و مفاهیمی را که سازنده هستی‌شناسی واژگانی هستند خلق و ابداع می‌کنند. این حرکت بسیار دشوار است و اگر با دقت انجام نگیرد منجر به منابع ضعیف هستی‌شناسی می‌شوند.

در اصل هیچ چیزی مانع از آن نمی‌شود که یک هستی‌شناسی واژگانی صوری نباشد. مشکل آن است که چنین پروژه‌ای هستی‌شناسی‌های واژگانی را نیمه صوری^{۹۱} ارائه می‌کند. دقیق تر بگوئیم این تعاریف بصورت گزاره‌های منطقی^{۹۲} که شامل ابهام زدایی اصطلاحات مربوط نیز می‌شود بیش از آنکه فرآیند استاندارد سازی شیوه اجرای هستی‌شناسی‌های صوری باشد، از موضوعات تحقیقاتی است که هنوز به نتایج کامل نرسیده است (Harabagiu et. al , 1999).

۶-۳- دامنه یا گستره^{۹۳}

در تحقیقات و اجرای هستی‌شناسی‌ها سه سطح از مختصات معنایی شناسایی شده‌اند:

^{۸۸} - Specification

^{۸۹} - Formal

^{۹۰} - Rigorous

^{۹۱} - Semi-Formal

^{۹۲} - Axiomatizing

^{۹۳} - Scope

الف - سطح بالایی (بالا ترین)^{۹۴}

ب - هم ارزی^{۹۵}

ج - قلمرو (دامنه)

برخی از منابع داده بنیاد با ساختار بالا به پایین^{۹۶} هنوز در هم آمیخته اند، و هر دوی آنها بستگی به طبقه بندی های عام و خاص روابط آنها دارند و درخت دانش را از بالا به پایین شکل می دهند.

هستی شناسی های پایه همچنین از ساختار سطح بالا به پائین تمایز می یابند و براساس مقتضیات هر کدام توصیفات فراوان را در اختیار می گذارند، در حالی که هستی شناسی سطوح بالا شامل گره های ساده می شود. این دو براساس منابعی مانند واژگان تخصصی و یا هستی شناسی های دامنه^{۹۷} با سلسله مراتب ویژگی های (مختصات) کاربردی که ممکن است بسیار محدود کننده باشد، از یکدیگر متمایز می گردند. فاصله ایجاد شده میان سطوح بالا و پایین ضرورت داشتن یک سطوح میانی را ایجاب می نماید. سطح میانی، سطحی را تشکیل می دهد که مفاهیم میانجی را در بر می گیرند و رابط میان منابع زیرین و مفاهیم حاکم در دامنه است.

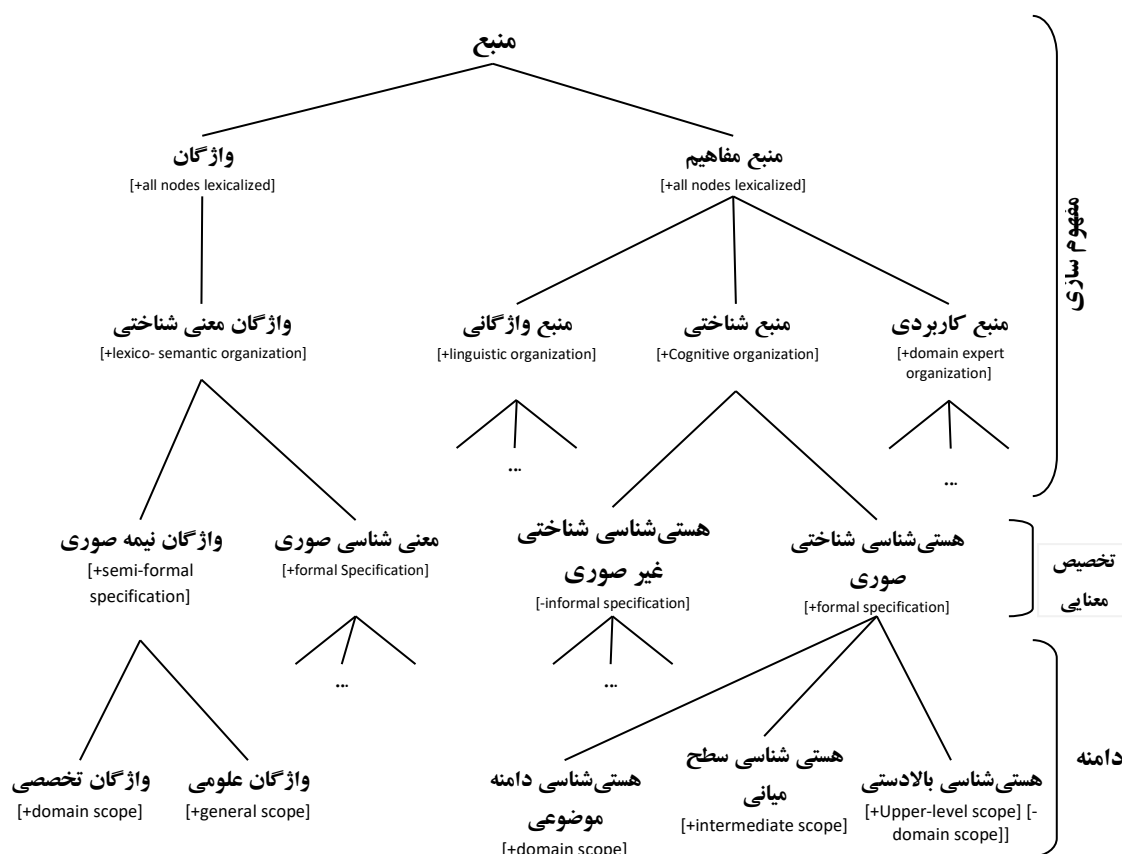
هرچند این سطوح در محتوا بسیار متنوع و متناوب، لکن به کارکرد اصلی شان وابسته هستند و برای شکل گیری سطحی میانی و ترسیم روابط مفاهیم با دو سطح دیگر تعامل می کنند. وجوه مشترک (فصل مشترک) هستی شناسی و واژگان را می توان به شکل زیر ترسیم نمود.

^{۹۴} - Upper-level

^{۹۵} - Core (reference)

^{۹۶} - upper-level

^{۹۷} - Domain Ontologies



نمودار ۳. تاکسونومی منابع آنتولکس (Prevot, L. et. al. 2010)

در مباحث گذشته دریافتیم که منابع واژگانی و هستی شناسی ها را به عنوان دو حوزه دارای تشابه ها و تفاوت ها می شناسند، که با توجه به مفهوم سازی، تخصیص معنایی و دامنه موضوعی همانند نمودار (۳) ترسیم شده است. منابع هستی شناسی و واژگانی از دیدگاه فناوری جدید، به برنامه های کاربردی مختلف تعلق دارد که اخیراً بصورت یک واحد درآمده اند و هم زمان مورد توجه قرار گرفته اند.

از دیدگاه هستی شناسی، بخش های زیر ساختی آنتولوژیها، مفاهیم و روابط میان آنهاست.

شناسایی این پدیده ها و مفاهیم مرتبط و تصمیم گیری راجع به ماهیت آنها وظیفه تحلیل گران هستی شناسی، متمرکز بر اصطلاحات و روابط میان آنها براساس منابع واژگانی موجود است.

این منابع گاهی تحت عنوان واژگان مرتبط^{۹۸} خوانده می شوند (Events, 1988) زیرا شبکه ارتباطات براساس دلالت های معنایی مدخل های واژگانی ارائه می گردد. از اینرو مفاهیم (یا واژگان) و روابط میان آنها دو عنصر اصلی هستند که بایستی در کاربر روی آنتولوژیها و منابع واژگانی مورد توجه قرار گیرند. این کار بطور موازی بر روی ساختار آنها و وجه مشترک میان هستی شناسی و واژگان انجام

^{۹۸} - Relational Lexicons

می‌گیرد و آنتولکس خوانده می‌شود (Prevot, L. et al. 2010).

در ساخت و تحلیل هستی‌شناسی‌ها با مفاهیم سرو کار دارند (که می‌تواند قابل دسترس از طریق واژه‌ها باشد) و مبتنی بر استدلال‌های بازنمایی دانش (تجانس، شفافیت، ایجاز و غیره) است. از سوی دیگر هستی‌شناسی‌های واژگانی معمولاً با مجموعه واژگانی حجیم و دسترس پذیر، آغاز می‌کنند تا این اصطلاحات به صورت منطقی و مفید سازماندهی شوند. از اینرو کاربر روی آنتولکس بستگی به بهترین تلفیق این دو رویکرد دارد و هدف آن ترکیب دقیق اطلاعات مفهومی در هستی‌شناسی‌های متداول و تبیین اطلاعات واژگانی در این کارها است.

همچنین آنتولکس به عنوان میانجی یا واسطه تا حدود زیادی در طراحی منابع واژگانی چند زبانه کارآمد خواهد بود: در چارچوب نظری «وردنت اروپایی»^{۹۹} این منابع نوعاً شکل گرفته از منابع تک‌زبانه وابسته به یک زبان میانجی هستند. هرچند این زبان میانجی بطور عام غیر ساختارمند است (Vossen, 1998) و ساختارمند نمودن آن گامی مهم در توسعه هستی‌شناسی‌های دامنه موضوعی است (Have and Nirenberg, 1992).

زبان میانجی ساختارمند بایستی با سطح مفهومی که قبلاً طراحی شده است، مطابقت داشته باشد. علاوه بر کاربرد آن در مهندسی زبان، این نوع منابع چند زبانه می‌توانند تسهیل کننده پژوهش‌های مربوط به همگانی‌های جهانی واژگان در بحث‌های میان زبان‌شناسان معتقد به جهان شمولی و نسبی‌گرا باشد.

۴-۶- محتوای هستی‌شناسی (مفاهیم و اصطلاحات)

در یک هستی‌شناسی، گره‌ها از ماهیت مفهومی برخوردارند و شامل مفاهیم، انواع، طبقات و مختصات معنایی می‌شوند. گستره مصداق اغلب آنها به صورت طبقات و رده‌بندی مختصات مفاهیم و اصطلاحات بازنمایی می‌شوند و متناظر با آن دسته‌بندی مفردات و نمونه‌ها (شواهد) است. در هستی‌شناسی‌هایی که مستقیماً از منابع واژگانی استخراج شده‌اند، با مفردات (اسامی عام و مدخل‌ها) همانند دیگر مفاهیم رفتار می‌شود.

در برخی از این منابع کمتر توجه به تمایز میان طبقه (رده) و عناصر^{۱۰۰} می‌شود: هر دوی آنها گره مفهومی در نمودار درختی هستند و بازنمایی آنها از طریق یکسانی انجام می‌شود. هر دو طبقات و نمونه‌ها با یک رابطه یکسان مدخل گذاری می‌شوند.

تفاوت میان واژگان واژه - بنیاد و واژگانی مانند وردنت که معنا - بنیاد است پیچیده‌تر می‌نماید. گره‌ها در منابع واژگانی نه اصطلاحات ساده و نه مدخل‌های صرفاً مفهومی است، بلکه معانی مختلف

^{۹۹} - Euro Word Net

^{۱۰۰} - Instance

واژه‌هایی است که با کاربرد متداول یک واژه مطابقت دارد، و منطبق با نمونه‌هایی که مصداق آن در یک پیکره زبانی موجود است، می‌باشند. در وردنت، گره‌ها و خوشه‌های معنایی^{۱۱} وجود دارند، یعنی مجموعه‌ای از معانی یک واژه که مجموعه مترادف‌ها را شکل می‌دهند. از اینرو وردنت در ابتدا واژگانی است که همه مدخل‌های آن دارای تعاریف زبانشناختی است، اما ساختار معنایی آن به شکل «مجموعه مترادف‌ها» و روابط آنها به گونه‌ای است که در یک هستی‌شناسی زبانشناختی بکار برده می‌شود (Guarino and Welty, 2000).

۵-۶- رویکرد بالا به پائین در مختصات معنایی واژه‌ها

در هستی‌شناسی صوری، ابهامات در اصطلاحات بایستی حتی‌الامکان قبل از ورود به مرحله تعیین مختصات معنایی واژه‌ها برطرف شود. هدف آن است که برای تخصیص معنی هر یک از اصطلاحات و بمنظور پیشگیری از ایجاد سوء تفاهات به دقت بسیار بالایی دست یابیم. وظیفه اصلی ساخت (تدوین) هستی‌شناسی، ردیابی و برطرف شدن ابهامات از درختواره دانش^{۱۲} و تهیه و تدوین هستی‌شناسی صوری بسیار دقیق‌تر از طریق تحلیل اطلاعات معنایی است. مرحله ضروری فرآیند تحلیل هستی‌شناسانه وابسته به تعیین اسکلت معنایی تاکسونومی دارد که طبقات اصلی و معماری تاکسونومیک (رده‌ای) را سازماندهی و بصورت رابطه انواع (اعم و اخص) در می‌آورد. بالاترین سطح این اسکلت بندی تمایزات میان اشیاء (پدیده‌ها) فرآیندها و کمیت‌ها، تفاوت اشیاء مصنوعی و طبیعی را بیان می‌کند. کاربرد این ساختارهای واژگانی براساس رویکرد «بالا به پایین»^{۱۳} مختصات معنایی واژه است، زیرا بایستی کاملاً جایگاه‌شان نسبت به تاکسونومی تعیین شود. علیرغم سودمند بودن رویکرد بالا به پائین در بازنمایی دانش، این رویکرد محدودیت‌های خاص خود را دارد. در زبان طبیعی واژگان مکانیسم زایشی برای اشتقاق معانی جدید را دارا هستند و بایستی در نظر داشته باشیم که نمونه‌های بالقوه و غیر قاعده‌مند نمی‌توانند به طور کامل ضبط و ثبت شوند. (Pustejovsky, 1995; Copestake and Boiscoe 1995; Apresjan 2000)

اکنون تصور معنا و مفهوم واژه به عنوان یک واحد معناشناختی مجزا خود زیر سوال رفته است. در توسعه و رشد مباحث «پردازش زبان» طبیعی و «آنتولوژی» حل این مسئله که آیا معنا و مفهوم بدون وابستگی به واژه نیز حقیقتاً موجودیت دارد یا خیر، ضرورتی ندارد، اما ارجاعات زیاد و مکرر به معانی متعدد یک واژه از منابع اطلاعاتی مهم، اهمیت این عناصر را در حوزه اطلاعات نشان می‌دهد.

^{۱۱} - Synsets

^{۱۲} - Knowledge Domain

^{۱۳} - top-down

۷- روابط معنایی

در هستی‌شناسی‌ها، مفاهیم در یک کل منسجم با روابط معنایی نظام یافته‌اند. ماهیت و شمار این روابط معنایی، موضوع بسیاری از مطالعات این حوزه است. هستی‌شناسی‌ها از دیدگاه مفهومی تقسیم‌بندی شده‌اند که در آنها مفاهیم به عنوان مدلول محسوب می‌شوند. از سویی دیگر، منابع واژگانی بر واحدهای واژگانی متمرکز هستند. روابطی که ویژگیهای معنایی را تنها بطور غیر مستقیم متمایز می‌گردانند، ماهیت مفهومی دارند. مانند "رابطه متضاد" که در ابتدای امر رابطه میان اشکال مختلف یک واژه و سپس میان معانی و مفاهیم مختلف یک واژه است (Fellbaum, 1998).

از اینرو نتایج تحقیقات در مورد این روابط از دو دیدگاه هستی‌شناسی صوری و هستی‌شناسی واژگانی کاملاً متفاوت هستند. برای هستی‌شناسی صوری آشکار ساختن ماهیت و ویژگیهای صوری این روابط مهم است: این که به چه نوع مدخل واژگانی (طبقات یا مفردات) مربوط هستند و این که آیا «برگشت پذیر»^{۱۰۴} یا «متقارن»^{۱۰۵} یا «گذرا ناپایدار»^{۱۰۶} هستند یا خیر؟

به عنوان مثال، برای هستی‌شناسی‌های صوری مهم است که توجه خود را به رابطه is-a معطوف کنند، زیرا این رابطه‌ها بی‌شمارند. این رابطه به طور فراوان و گسترده به کار برده می‌شود. و اغلب تنها با مسامحه تعریف می‌گردد و صرفاً باشم زبانی مطابقت دارد. از سویی دیگر روابط منابع واژگانی بین معانی یک واژه (مانند شمول معنایی و چند معنایی) و یا میان واژه‌های ساده‌ای که روابط واژگانی همانند (تقابل معنایی) دارند، ایجاد می‌گردد.

در مطالعات اخیر پیرامون منابع واژگانی تمرکز بر تعاریف دقیق این روابط نمی‌باشد، بلکه بر روی روش‌های کشف خودکار (اتوماتیک) این روابط و نیز کاربرد آنها در استخراج^{۱۰۷} دانش واژگانی است. به این مسئله مهم باید توجه داشت که رده‌بندی کلی این روابط معنایی از نظر جانشینی^{۱۰۸} و همنشینی^{۱۰۹} میان رویکردهای مفهومی و واژگانی مشترک است.

۷-۱- روابط جانشینی^{۱۱۰}

این رابطه در میان عناصری از ماهیت یکسان برقرار می‌شود و متعلق به یک برداشت مشترک و

^{۱۰۴} - reflexive

^{۱۰۵} - symmetric

^{۱۰۶} - transitive

^{۱۰۷} - extracting

^{۱۰۸} - paradigmatic

^{۱۰۹} - Syntagmatic Relations

^{۱۱۰} - Paradigmatic Relations

همگانی از یک پدیده است.

در کتاب کروز (Cruse و 1986) روابط جانشینی^{۱۱۱} با روابط هم ارزی^{۱۱۲} فیما بین طبقات واژگانی مانند ماهیت^{۱۱۳} شمول معنایی^{۱۱۴}، همپوشانی^{۱۱۵} انفصال^{۱۱۶} مرتبط هستند.

در واقع بهترین روابط جانشینی شناخته شده در روابط معنایی واژگانی^{۱۱۷}، مترادف، تضاد، جزء واژگی^{۱۱۸} و شمول معنایی است. در هستی‌شناسی‌ها، روابط مفهومی مرتبط با "تقابل مفهومی"^{۱۱۹} به طور صوری تعریف می‌شوند. هرچند، در این زمینه روابط معنایی بیشتری را می‌توان تصور کرد، به عنوان مثال روابط خواهر-برادری^{۱۲۰} در یک تاکسونومی مانند: سرخ / سیاه / آبی، سگ / گربه) نمونه‌های بارزی هستند.

و فور در روابط جانشینی (Murphy, 2003) سبب شده که به طرح تحقیقاتی کلان تحت عنوان:

«اصول روابط تقابل»^{۱۲۱} که در برگیرنده روابط جانشینی است، تبدیل شود

«هوانگ و دیگران ۲۰۰۲» رابطه معنایی جدیدی را تحت عنوان پارانومی^{۱۲۲} مطرح کردند که روابط جانشینی میان مفاهیم را در رده‌بندی معنایی ارائه می‌دهد (Huang et al., 2002).

بسیاری از متخصصین «پردازش زبان طبیعی»، هستی‌شناسی‌ها را بکار می‌برند، زیرا کاربرد صرف تاکسونومی کار آنان را محدود می‌سازد. روابطی که بنا بر چارچوب نظری موردنظر، نامهای مختلفی بر آنها اطلاق شده است.

۲-۲- روابط^{۱۲۳} همنشینی

همانگونه که در بالا ذکر شد، روابط همنشینی میان مدخل‌های واژگانی دارای ماهیت متفاوت ایجاد می‌شود، که به دنبال هم وقوع می‌یابند اما نمی‌توانند جایگزین یکدیگر شوند.

اغلب آنها در قالب واژگانی ظهور می‌یابند و دارای مقولات نحوی متفاوتی هستند. در معنی‌شناسی

^{۱۱۱} - Paradigmatic

^{۱۱۲} - Congruence

^{۱۱۳} - Identity

^{۱۱۴} - Inclusion

^{۱۱۵} - Overlap

^{۱۱۶} - Disjunction

^{۱۱۷} - Lexical Domain

^{۱۱۸} - Metonymy

^{۱۱۹} - Conceptual Opposition, part-of, is-a-kind of

^{۱۲۰} - Siblings

^{۱۲۱} - Principle of Relation by Contrast

^{۱۲۲} - Paronymy

^{۱۲۳} - Syntagmatic relations

واژگانی، روابط همنشینی بیشتر مرتبط با مطالعات نحوی / معنایی و وجوه مشترک آنهاست که بر روی اسناد و نقش‌های موضوعی^{۱۲۴} تمرکز می‌یابند. روابط جانشینی در برگیرنده روابط میان^{۱۲۵} (فاعل^{۱۲۶}، مفعول^{۱۲۷})^{۱۲۸}، شامل پدیده‌ها و فرآیندها و نیز سطح واژگانی (روابط اسم و فعل) و یا به عنوان مثال میان مقوله و صفات (رابطه اسم و صفت در سطح واژگانی) برقرار می‌گردد. بسیاری از این روابط همانند^{۱۲۹} (روابط حالت) را دارا هستند، که در مطالعات عمیق «فیلمور»^{۱۳۰} در گسترش نظریه دستور حالت^{۱۳۱} دیده می‌شوند (Fillmore, 1998) و اکثر روابط «فریم‌نت»^{۱۳۱} (Baker et al., 1998) را تشکیل می‌دهند. در حالی که «وردنت» و دیگر کارهای مبتنی بر هستی‌شناسی بر روابط جانشینی استوار است، روابط همنشینی کمتر مورد توجه توسعه دهندگان منابع واژگانی قرار گرفته است، در حالی که اهمیت آن برای کاربردهای «پردازش زبان طبیعی» به تدریج ضرورت بیشتری می‌یابد. روابط مکمل میان جانشینی و همنشینی در «وردنت» و «فریم‌نت» تلفیق مؤثر این دو عنصر مهم در این رشته تحقیقات کاربردی محسوب می‌شود.

«تمایز»^{۱۳۲} میان جانشینی و همنشینی در بخشهایی همپوشانی دارند مانند تمایزی که نیرنبرگ و راسکین (Nirenberg and Raskin 2001) میان معنی‌شناسی واژگانی بر گرفته از نحو و بر گرفته از هستی‌شناسی قائل می‌شود. اولی مطابق است با روابط همنشینی که مطالعات آن با تعامل میان نحو و معنی‌شناسی وابستگی نزدیکی داشته است (Levin, 1993). اخیراً توجه خاصی به روابط جانشینی (بویژه در زمینه تاکسونومی (رده‌بندی) و جزء واژگی^{۱۳۳} که در برگیرنده روابطی مربوط به مبحث همنشینی (مانند: مشارکت مفعول در فرآیندها) است، دیده می‌شود.

در حوزه‌هایی که در بازنمایی دانش دخالت دارند، بدون توجه به این که آیا هدف روشنی از کارآمدی^{۱۳۴} روان‌شناختی آن دارند یا خیر، در رویکردهای زیادی قابل تمایز هستند، مانند: از مطالعات فلسفی که به گذشته‌ای دور (دوران ارسطو) تعلق دارند تا برسد به مطالعات روان‌شناختی

^{۱۲۴} - Thematic Roles

^{۱۲۵} - Endrants

^{۱۲۶} - Agents

^{۱۲۷} - Objects

^{۱۲۸} - Perdurants Luculent

^{۱۲۹} - Care relations

^{۱۳۰} - Case Grammar

^{۱۳۱} - Frame Net

^{۱۳۲} - Distinction

^{۱۳۳} - Metonymies

^{۱۳۴} - Adequacy

که متمرکز بر ارائه دانش ذهنی و مطالعات زبان شناختی هستند.

از اینرو موضوعات اصلی تعامل میان هستی‌شناسی و منابع واژگانی، یک نوع تجربه جدید موضوعات گذشته روانشناسی زبان، زبانشناسی، هوش مصنوعی و فلسفه از نظردیدگاه‌های جدید در این گرایش‌هاست و درصدد پاسخگویی به علایق جدید در این موضوعات و رابطه آن با رویکردهای «وب معنایی»^{۱۳۵} می‌باشد.

در کل، اهمیت رویکردهای میان رشته‌ای به رسمیت شناخته شده و برای گسترش منابع زبانشناختی و «بازنمایی دانش» از سوی کسانی که سهم به سزایی در تأثیر گذاری در این حوزه داشته‌اند، مورد تایید قرار گرفته است.

(Hobbs et al., 1987; Sowa, 2000; Pustejovsky, 1995; Nirenberg and Raslin, 2004; Guarino, 1998b)

این مباحث نشان‌دهنده وجود زمینه‌ای غنی و فرصتی بسیار مناسب برای تحقیقات آتی در حوزه پردازش زبان و بازنمایی دانش و معنی‌شناسی واژگانی، بویژه زمینه جالب تعامل میان هستی‌شناسی صوری^{۱۳۶} و منابع واژگانی یا هستی‌شناسی‌های واژگانی است. به این منظور هستی‌شناسی‌های صوری و هستی‌شناسی‌های مبتنی بر بررسی‌های فلسفی و کارآمد در بازنمایی دانش کاربرد مجدد می‌یابند.

منابع واژگانی یا هستی‌شناسی‌های واژگانی ساختاری مبتنی بر زبان طبیعی بویژه براساس واژگان زمینه‌ای علوم شناختی^{۱۳۷} دارند

۳-۷- مقوله بندی^{۱۳۸}

مطالعات بر روی مبحث مقوله‌بندی از سوی دانشمندان علوم شناختی بسیار مورد توجه قرار گرفته است. به عنوان مثال معنی‌شناسی تحلیلی^{۱۳۹} از سوی کاتز و فودور (Katz & Fodor, 1963) که در آن «مقوله» یک واژه با مجموعه از مشخصه‌ها (مختصات معنایی و نحوی) تعیین می‌شود که واژه‌ای را از واژه دیگر متمایز می‌کند و در زبان یکی از معتبرترین نظریه‌هاست. به نظر می‌رسد این نظریه تا حدود زیادی منطبق با تحلیل مفهومی صوری است^{۱۴۰} و توسعه یافته آن بواسطه گانتز و ویل (Ganter & Wille, 1998) و برای اولین بار در مورد داده‌های واژگانی در اثر «پریس» بکار برده شده است و (Priss 1998, 2005) امروز این الگوی نظری مورد کاربرد در رویکردهای هستی‌شناسی متعددی

^{۱۳۵} - Semantic Web

^{۱۳۶} - Formal Ontology

^{۱۳۷} - Cognitive ground

^{۱۳۸} - Categorization

^{۱۳۹} - Componential Semantics

^{۱۴۰} - Formal Concept Analysis (FCA).

قرار گرفته است. هرچند «معنی شناسی تحلیلی» در زمینه «نظریه پیش نمونه‌های زبانی»^{۱۴۱} پیشرفت‌های زیادی داشته است (Rosch 1973, 1978). به طور تجربی ثابت شده است که ارتباطات (پیوندهای) میان واژه‌ها و حتی مفاهیم با یک مقوله، از درجه بندی معانی مفاهیم نشأت یافته است. از این رو مشکل تعیین مرزهای مقولات، موجودیت آنها و تعریف بالقوه آنها مسئله‌ای بسیار جدی است و عوامل ساختاری نقشی تأثیر گذار بر تعیین مرزهای مقولاتی دارند. مسئله دیگر کاربرد یک فهرست مختصات (ویژگی‌های معنایی) است که بایستی بسیار با صراحت بیان شود و این سوال را بر می‌انگیزد که حدود و ثغور خود این مختصات کجاست؟ البته این بحث‌ها مربوط به تعیین تعاریف از دیدگاه فلسفی است. در کنار مسئله «فرضیه پیش نمونه‌ها»، زمینه مشترک دیگر مطالعاتی مربوط به «الگوهای مفهومی» از قبیل: دسته‌بندی انواع^{۱۴۲}، ربط دهنده^{۱۴۳}، مفردات^{۱۴۴} و مفاهیم نقش دهنده^{۱۴۵}، ساختار مقوله‌ای و روابط مورد نظر یا قالب‌ها معنایی^{۱۴۶} است. شواهد همگرایی^{۱۴۷} زیادی نشان می‌دهند که زبان تمایل زیادی به مقوله بندی دارد. چنانچه واژه‌ای بر یک مفهوم در زبانی خاص برچسب گذاری شود، یادگیری مطابق با مقولات مورد نظر در زبان‌های دیگر را توسط کودکان سریعتر و آسان‌تر می‌کند. البته مباحث زیادی را می‌توان در مورد مقوله‌بندی ارائه داد، اما این بحث را به منابع معتبری که در این زمینه وجود دارند، ارجاع می‌دهیم مانند: (Wierzbicka, 1990; Croft & Cruse 2004; Murphy 2003 and Schalley and Zaefferer, 2007 ;)

در این منابع رویکردهای مختلف و تعیین حدود و ثغور آنها مفصل به بحث گذاشته شده است.^{۱۴۸} در زبان شناسی میزان زیادی از مطالعات و تحقیقات متمرکز بر موضوع «اسناد» متمرکز است زیرا نوع تفسیر و تعبیر جمله را تعیین می‌کند. فیلمور (1976) پیشرو در ارائه این مطالعات است، که بر تحلیل واژگان براساس روابط فیما بین و در «قالب معنایی» مبتنی بر اسناد که در آن ظاهر می‌شود، توصیه می‌کند. این مطالعات متمرکز بر روابطی است که در قالب دستور حالت (فیلمور 1968) بیان شده است.

^{۱۴۱} - Prototypes

^{۱۴۲} - Sortal

^{۱۴۳} - Relational

^{۱۴۴} - Individual

^{۱۴۵} - Functional Concept

^{۱۴۶} - Frames

^{۱۴۷} - Converging Evidence

^{۱۴۸} - Predication

دانشگران^{۱۴۹} در ایجاد دامنه موضوعی براساس^{۱۵۰} ساختار استدلالی (Grimshaw, 1990) نقش‌های معنایی^{۱۵۱}، محدودیت‌های گزینشی^{۱۵۲} (Dowty, 1990) و اجبار انواع^{۱۵۳} (Pustejovsky, 1995) را مد نظر قرار داده‌اند. در این حوزه در منابعی از قبیل فریم‌نت^{۱۵۴} (Borker et al. 1998) و شبکه افعال^{۱۵۵} (Kipper et al. 2000) تحقیقاتی وسیعی انجام گرفته است.

حال سوال این است که: آیا سطوح معنایی و مفهومی دارای یک هویت هستند؟ بسیاری از طرفداران رویکرد علوم شناختی زبان این فرض را مسلم می‌دانند که سطوح معنی‌شناسی و مفهومی از برابری و تساوی برخوردار هستند.

جکندوف اظهار می‌دارد، «رو ساخت»^{۱۵۶} رابطه مستقیم با مفاهیم دارد که در قالب یک سری از قواعد بیان می‌شود. این مفاهیم در برگیرنده اطلاعات بیشتری مربوط به دانش جهانی (دانش دایره المعارفی) است. هرچند بنا به گفته وی، تفکیک اطلاعات معناشناختی از دایره المعارفی، بدون از دست دادن همگانی‌ها (کلیات) غیر ممکن به نظر می‌رسد، اما سطوح معنایی و مفهومی بایستی یک سطح واحد را تشکیل دهند (Jackendoff, 1983).

هر چند لوینسون و دیگران^{۱۵۷} (Gumperz & Levinson, 1996; Levinson, 1997, 2003) استدلال‌های جدی و پیشرفته‌تری، در چارچوب نظریه کاربردی خاص با عنایت به تفکیک میان بازنمایی معنایی و مفهومی ارائه داده‌اند، این اختلاف نظرها از دیدگاه‌های مختلف زبان و از جنبه‌های مختلف نظری بیان شده است. در حالی که جکندوف تمرکز بر توافق زبانی چامسکی^{۱۵۸} دارد، لوینسون بر اهمیت ماهیت اجتماعی زبان و از اینرو بر مسائل برون زبانی^{۱۵۹} تأکید دارد. در توجیهات زبانی جکندوف تأکید دارد:

زبان که از ابتدا ابزاری برای ایجاد ارتباطات اجتماعی محسوب می‌شد، بیش از آن که ابزاری برای بازنمون دانش ذهنی هر فرد باشد با ابعاد اجتماعی سروکار دارد.

^{۱۴۹} - Knowledge Worker

^{۱۵۰} - Argument Structure

^{۱۵۱} - Thematic Roles

^{۱۵۲} - Selectional Restrictions

^{۱۵۳} - Type Coercion

^{۱۵۴} - Frame Net

^{۱۵۵} - Verb Net

^{۱۵۶} - Surface Structure

^{۱۵۷} - Levinson

^{۱۵۸} - Chomsky's I Internal Language

^{۱۵۹} - External Language

از دیدگاه کاربردی، «باتمن» (۱۹۹۷) براساس زمینه‌های روشمند استدلال نموده که یک هستی‌شناسی (نظام مفاهیم) نیاز به تعامل مفهومی و روساخت دارد. وی تصریح کرده که چنین منبعی نباید خیلی نزدیک و یا خیلی دور از سطح زبان شناختی باشد و جزء به جزء مدل پایه و بالا به پائین را^{۱۶۰} را به عنوان نمونه‌ای از این تعامل متوازن هستی‌شناسی می‌داند (Batemon et al, 1995).

این خط مطالعاتی از سویی توسط نیرن برگ و دیگران (Nirenberg & Roch 2004) دنبال می‌شود و از سویی دیگر «پوستیوسکی»^{۱۶۱} و همکارانش نیز این جهت‌گیری ساختار مشترک را برگزیده و همزمان پویایی را در واژگان زایشی، دنبال می‌کنند (Pustejovsky, 1991, 1995).

در تعریف نظام مقولات هستی‌شناسانه و بمنظور فراهم سازی یک نگرش، در پاسخ به این سوال که مفاهیم به چه نوع پدیده‌هایی (اشیایی) تعلق دارند؟ یکی از مهمترین وظایف هستی‌شناسی‌ها ارائه مقولات متافیزیکی (ماوراءالطبیعه) است، که بیشترین (و در برخی موارد بصورت متناقض یا متضاد) گزاره‌های منطقی، را در بر می‌گیرد. از ارسطو گرفته تا فیلسوفان معاصر به این مسئله توجه داشته و دارند که در علم فلسفه این گزاره‌ها، هم در ماهیت مقوله‌های هستی‌شناسانه و هم در بازنمایی دقیق و روشن مقولات و نیز در چارچوب واقع‌گرایانه بودن تصمیمات هستی‌شناسانه یعنی: پدیده‌ها، روابط منطقی و نقصانها^{۱۶۲} تأثیر می‌گذارند.

نبود اجماع آراء در این زمینه به ایجاد شک و تردید نسبت به حدود و ثغور و ثبات چنین نظام هستی‌شناسی منتهی شده است. هرچند هدف رویکردهای اخیر، برداشت درست و صحیح از علم فلسفه بوده است و در عین حال پرهیز از خطاها و اشکالات و وسواس فلسفی، تحقیق و پژوهش در این زمینه را سخت و با مشکل مواجه کرده است. این وسواس علمی می‌تواند از یک جنبه فلسفی برای ایجاد هستی‌شناسی کارساز باشد و نه تنهادر تعیین موقعیت یک مفهوم در یک تاکسونومی، بلکه در شیوه تصمیم‌گیری در این زمینه تعیین کننده باشد.

دومین جنبه مهم مبتنی بر تمایز^{۱۶۳} میان متافیزیکی توصیفی و بازاندیشی^{۱۶۴} (بازنگری) است که تعبیر^{۱۶۵} موشکافانه فلسفی سبب شده که اتفاق نظر در برداشت از مفاهیم بوجود نیاید. برای مثال مباحثی که در مورد دشواریهای یک چنین رویکرد اصلاح طلبانه پدید می‌آید، زمانی است

^{۱۶۰} - General-Upper- Model

^{۱۶۱} - Pustejovsky

^{۱۶۲} - Holes

^{۱۶۳} - Strawson

^{۱۶۴} - Revisionary

^{۱۶۵} - Paraphrases

که سعی می‌شود از نام گذاری اشیاء و پدیده‌ها به سادگی اشکال گرفته شود، از اینروست که رویکردهای اصلاح طلبانه گرایش به کنار گذاشتن زبان طبیعی و شم زبانی مشترک در تحلیل هستی‌شناسانه مبادرت می‌ورزد.

از سویی دیگر دیدگاه توصیفی به عنوان دیدگاهی فلسفی تر مطرح شده و نقطه شروع حرکت روشمند و قابل اعتماد برای تحقیقات کاربردی هستی‌شناسانه را فراهم می‌سازد.

اگر بخواهیم دقیق‌تر بگوئیم، با مجاز دانستن هستی‌شناسی‌های مختلف (همانند توصیفات و برداشت‌های گوناگون از جهان هستی) جهت ایجاد تفاهم، قادر خواهیم بود که از شکل‌گیری هستی‌شناسی نامحدود و بی‌انتهای در مورد رخدادهای طبیعی و ماهیت واقعی اشیاء و پدیده‌های جهان هستی پرهیز کنیم، و این گونه برخورد به ایجاد تمایز میان هستی‌شناسی با گرایش فلسفی و هستی‌شناسی با گرایش ثبت دانش بشری منتهی می‌گردد (Guarino, 1998b).

طراحان جدید هستی‌شناسی‌ها هم و غم‌شان ارائه یک هستی‌شناسی کامل نیست، بلکه سعی در ارائه هستی‌شناسی‌های بالقوه‌ای دارند که دارای ساختارهای دامنه‌های موضوعی متفاوت هستند و از دیدگاه‌های و زوایای مختلف به یک پدیده می‌نگرند. در مطالعات فلسفی بیش از کسب موفقیت در این کار، سعی در ارائه تعاریف دقیق دسته‌بندی مقولات می‌شود.

توماسون در تحقیق خود به سه شیوه عمده در این زمینه اشاره دارد (Thomasson, 2004).

۴-۷- ارزش‌گذاری مختصات معنایی

این شیوه، هر مقوله را به دو دسته مختصات تقسیم می‌کند که اشیاء یا پدیده‌ها را با داشتن یا نداشتن یک وجه معنایی معین متمایز می‌گرداند. این نظام مختصات را می‌توان به یک ماتریس چند بعدی (وجهی) و بالقوه بسیار غنی توسعه و گسترش داد، به طوری که از لحاظ ممیزه‌های معنایی بیش از یک درختواره معنایی را ارائه دهد.

دو نتیجه از این رویکرد فلسفی حاصل می‌شود، اولاً مختصات منفی برای بسیاری از مفاهیم که دارای یک مقوله مشترکند، اشتراک در ویژگی لاینفک آن^{۱۶۶} ثبت می‌شود و ثانیاً ناکارآمدی بسیاری از مختصات دوگانه مقولات بوسیله آزمونهای زبان شناختی به اثبات می‌رسد.

۵-۷- تشخیص بی‌معنایی^{۱۶۷}

این روش مبتنی است بر آزمون «متن دستوری» که در آن یک مفهوم می‌تواند بدون عرضه معنایی بکار گرفته شود. به عنوان مثال: «این خانه قهوه‌ای رنگ است» دارای بار معنایی است. در حالی که :

^{۱۶۶} - Intrinsic Property

^{۱۶۷} - Absurdity Detection

«این فرضیه قهوه‌ای رنگ است». بی معنا می‌باشد. از اینرو بایستی ماهیت متفاوت این دو مفهوم از لحاظ معنایی تعیین شود. این روش را می‌توان برای اثبات آن که این دو عنصر به دو مقوله متفاوت تعلق دارند بکار گرفت، اما نمی‌توان اثبات کرد که این دو پدیده متعلق به یک مقوله یکسان هستند، زیرا برای اثبات بی‌معنایی بایستی دو بافت متمایز کننده در نظر گرفته شوند (Ryle, 1971).

۶-۲- روش تمایز میان ماهیت و شرایط پدیدارشناسی

این روش نیز گونه‌ای از شیوه قبلی است که توسعه و گسترش یافته و برای بیان عبارات که بر اشیاء و پدیده‌ها دلالت می‌کنند، مورد آزمون قرار می‌گیرند. بنابراین نظریه، فرگه (Frege, 1884) معتقد است: اسامی خاص^{۱۶۸} متمایز و متفاوت از دیگر اصطلاحات هستند که شرایط ماهوی آنها به مقولات کاربردی‌شان افزوده می‌شود. نام‌ها (عام) وابسته به دسته‌بندی و تفکیک اشیاء و پدیده‌هاست. این نام‌ها، با تفکیک مقولات اشیاء (پدیده‌ها) متناظر با این نام‌ها قابل تمایز هستند و بر اساس مقولات هستی‌شناسانه متمایز می‌گردند.

سپس شرایط کاربردی و معیارهای ماهیتی بر روی این رده‌بندی آزموده می‌شود. دومین عنصر سازنده ماهوی و بنیادی برای شناسایی مقولات شرایط همانندی است. در این مسئله توصیه «کواین»^{۱۶۹} این است: «هیچ مداخلی نباید و نمی‌تواند بدون هویت مستقل باشد» (Quine, 1960) و توصیه می‌کند که قبل از آنکه موجودیت یک مداخل را به رسمیت بشناسیم، وجوه تمایز آن را از دیگر مداخل‌ها روشن و واضح بیان نماییم.

اکنون کاربرد روش آنتوکلین^{۱۷۰} بطور گسترده و در موقعیت‌های کاربردی گوناگون بصورت رایج درآمده است، و شیوه‌ای قدرتمند برای ساخت و روزآمد سازی آنتولوژیها محسوب می‌گردد. این شیوه از مطالعات فلسفی گذشته شامل مجموعه‌ای از ویژگیهای «فرا داده» گرفته شده که برای تشخیص مقولات و ویژگیهای معنایی آنها استفاده می‌شود. این ویژگیهای فرا داده‌ای شامل شرایط ماهوی نیز می‌گردد. همانگونه که قبلاً هم تأکید شد، هدف از به کارگیری این شیوه تنها در نظر گرفتن یک سری از مقولات پذیرفته شده نیست، بلکه کمک به تدوین کنندگان آنتولوژی است که هستی‌شناسی‌های مخصوص به خود را طراحی و موضوعات هستی‌شناسانه جدید را در فرآیند تدوین آنتولوژی کشف کنند.

به عبارتی دیگر آنتوکلین به طراحان هستی‌شناسی نمی‌گوید که «مقولات کامل» کدامند، بلکه به

^{۱۶۸} - Names

^{۱۶۹} - Quine

^{۱۷۰} - Onto Clean

روشن نمودن عواقب و پیامدهای گزینه‌های انتخابی در حین مفهوم سازی و تعیین مختصات مفاهیم کمک می‌کند (Guarino & Welty, 2000).

۲-۲- زمینه واژگان نگاری^{۱۷۱}

واژگان نگاری، گرایشی است که با واژه‌نامه‌ها و فرهنگ نویسی سروکار دارد و بطور سنتی به دو بخش نظری و کاربردی که همانا تدوین واژه‌نامه‌هاست تقسیم می‌شود. هرچند، فعالیت‌های واژگان نگاری جدید معمولاً ابعاد نظری و کاربردی را تلفیق می‌نمایند. از آنجا که واژه‌نامه‌ها و فرهنگ‌های سنتی سلف منابع واژگانی رایانه‌ای بوده‌اند، جای تعجب ندارد که مطالعات واژگان نگاری قبلاً نیز با بسیاری از دشواریهایی که رویکردهای جدید با آن مقابله می‌کنند، درگیر بوده است. این مطالعات در گذشته در مورد مختصات و ویژگی‌های معنایی‌ها انواع گوناگون فرهنگ‌ها و واژه‌نامه‌ها بوده است.

۲-۸- دایره المعارف در مقابل واژه‌نامه‌ها (فرهنگ‌ها):

دایره المعارف‌ها تأکید بر اهمیت اسامی عام دارند و نشان می‌دهند که بسیاری از آنها بایستی در واژه‌نامه‌های عام آورده شوند. بر اساس این تمایز شهربا^{۱۷۲} نیز بحثی پیرامون واژه‌هایی دارد که در واژه‌نامه‌ها و یا دایره المعارف‌ها براساس سطوح مختصات معنایی، مدخل‌های متفاوتی را تشکیل می‌دهند.

تمایز میان واژه‌نامه‌های صورتگرا^{۱۷۳} در تعامل با واژه‌نامه‌های معناگرا^{۱۷۴} که نوع دوم مترادف‌ها و دیگر روابط معنایی (مثلاً در تزاروس روژه^{۱۷۵}) مورد توجه قرار می‌گیرند.

در تدوین و ترجمه واژه‌نامه‌ها، دقت در ماهیت پیچیده روابط میان مدخل‌های واژگانی بین‌زبانی توصیه می‌شود. واژگان نگاری کاربردی، که در واقع شامل واژگان منابع علمی معتبر می‌شود، در مهیا نمودن منابع پیشرفته‌تر واژگانی ضرورت می‌یابد (Sheherba, 1995).

۸- هستی‌شناسی و واژگان

هنگام توسعه و گسترش واژه‌نامه (فرهنگ‌نامه)، اغلب نیاز به تأمل و بررسی بیشتر و صرفه جویی اقتصادی و ممانعت یا چشم‌پوشی از مدخل‌های غیر ضرور دارد. هرچند فارغ از سنگینی و فشار

^{۱۷۱} - Lexicographic Ground

^{۱۷۲} - Sheherba

^{۱۷۳} - Form-Based

^{۱۷۴} - Meaning-Based

^{۱۷۵} - Roget Thesaurus

مشکلات و محدودیت‌ها، واژگان نگار را مجبور به انتخاب واژگان معنی‌گرا بجای صرف اطلاعات زبانشناختی واژگان زبان عمومی می‌کند و به دایره‌المعارف‌ها و واژگان تخصصی روی می‌آورد. اهمیت تفکیک و تمایز میان دانش واژگانی و دانش دایره‌المعارفی از نظر «نورن برگ و زینن» (Nuernberg & Zaenen 1992) نیز بنیادی است. آنها توضیح می‌دهند که چون زبان بعنوان ابزار ارتباطات اجتماعی که در آن «بافت فرهنگی» عنصر تأثیرگذار است، بسیاری از قواعد دانش دایره‌المعارفی استحقاق آنرا دارند که در واژه‌نامه‌ها (فرهنگ‌ها) بکار گرفته شوند. برای کاربردی‌تر شدن، واژگان نگار می‌تواند از نظریه زبان شناختی «آپرس جان» بهره‌برداری کند (Apresjan, 2000)، که بر اساس آن چهار گرایش از نظریه‌های زبان شناختی معاصر مناسب کاربرد مستقیم در «واژگان نگاری نظام‌مند»^{۱۷۶} شناخته شده‌اند:

- الف - جستجوی تصویر مشترک از جهان براساس شعور مشترک همگانی (معرفت عمومی)،
- ب - انتقال از بررسی و مطالعه فهم و درک واژه‌های منفرد به طرف منابع نظام‌مند واژگان نگاری،
- ج - مطالعه و بررسی دقیق و موشکافانه تک تک واژه‌ها در تمامی ابعاد زبان شناختی آنها،
- د - همگرایی مطالعات دستوری و واژگانی بسوی شکل‌گیری نظریه نظام یافته توصیف زبان شناختی. تمامی این عوامل در ایجاد زمینه‌ای مناسب برای «نظریه واژگان زایشی» موثر خواهد بود (Pustejovsky, 1991, 1995).

۱-۸- زمینه زبان شناختی معنی‌شناسی واژگانی معاصر

در حال حاضر، هدف اصلی مطالعه زبان توسعه «سیستم‌های پردازش زبان طبیعی» است. با توجه به زمینه فلسفی، معنی‌شناسان زبانهای طبیعی و فیلسوفان زبان عمیقاً به آنتولوژی (هستی‌شناسی) زبان‌های طبیعی و یا مقولات هستی‌شناسانه که بطور اخص در مطالعات زبانهای طبیعی صورت می‌پذیرد، علاقه‌مند شده‌اند.

مطالعاتی پیرامون «زمان» و جنبه «یا «وجه» و نیز پیرامون «علائم جمع» «اسامی قابل شمارش و غیر قابل شمارش» حوزه‌های مطالعاتی را فراهم ساخته که در برخی مواقع اصطلاح «متافیزیک زبان» به آن تعلق می‌یابد.

هرچند این مطالعات سرنوشت‌ساز زمینه نظری محکم و قاطعی را برای پیشرفت‌های آتی فراهم می‌سازد، این بخش از تحقیقات تأثیر مستقیم کمی بر توسعه منابع کاربردی دارند.

یکی از استثنائات، «نظریه واژگان زایشی» متعلق به جیمز پوستیوسکی است. (James Pustojevsky 1991, 1995) «واژگان زایشی» تشریک مساعی در مطالعات واژگان و رابطه آن با هستی‌شناسی دارد

و زمینه فلسفی را با نظریه معنی‌شناسی واژگانی غنی که تبلوری از بحث موشکافانه روابط منطقی است، مرتبط می‌سازد. «واژگان زایشی» به دو موضوع مهم وجدی ما را هدایت می‌کنند و با دستاوردهای سنتی مقابله می‌کند. این مطالعات عموماً براساس احصاء مفاهیم بی‌شمار و انبوه و شامل فهرستی از مفاهیم و معانی واژه‌ها صورت می‌گیرد.

پوستیوسکی استدلال می‌کند که مفاهیم و معانی یک واژه نامحدود است، زیرا گویشوران زبان می‌توانند به راحتی معانی جدید به یک واژه اختصاص دهند.

در این بافت مدخل‌های واژگانی بیشتر پویا هستند و مطابق ترکیبات جنبه‌ها (نمودهای) مختلف معنا و براساس نظریه ارسطویی برای نام‌ها (اسامی) وجوه مختلف معنا را تشکیل می‌دهند، که عبارتند از:

عناصر سازنده^{۱۷۷} (اجزاء یک پدیده یا شیئی چیست). صوری^{۱۷۸} (موقعیت یک شیئی و پدیده در رده‌بندی چیست)، کنادی^{۱۷۹} (منشاء آن پدیده چیست) و نقشی^{۱۸۰} (نقش‌های معنایی آن چیست).

این نگرش چند بعدی معنا در برخی موارد برداشت‌های گوناگون وراثتی در سلسله مراتب^{۱۸۱} را موجب می‌گردد. هرچند این عملکرد به بی‌ثباتی دیدگاه باز نمایی دانش می‌انجامد، زیرا ساختار آن ابهام‌آمیز می‌شود و نگهداری انسجام کل آن مشکل‌تر می‌گردد. واژگان زایشی این آشفتگی را بر طرف می‌کند، زیرا دیدگاهی چند وجهی از معنا دارد و سازماندهی دانش محکم و مستدل با کاربرد روابط معنایی مختلف بجای درگیر شدن در یک ساختار درختی واحد را ایجاد می‌کند.

واژگان زایشی بسیار مورد توجه قرار گرفته و در برخی از منابع کاربردی واژگان طبق اصول و معیارهای این نظریه عمل کرده‌اند، مانند سیمپل^{۱۸۲} (Lenci et al., 2000) که در برگیرنده ساختار^{۱۸۳}

توسعه یافته‌ای است (Busa et al., 1999) و دقیقاً ویژگی‌های معنایی را تعریف و نوع مفهوم را تعیین می‌نماید. این تحقیق گسترده بر روی گزاره‌ها و دال و مدلول که نشانه‌های چند وجهی چهارگانه^{۱۸۴}

دارد، همه وجوه معنا را در بر نمی‌گیرد. به عنوان مثال «آشرودنيس» (Asher & Denis 2004)

معتقدند که بایستی نظریه واژگان زایشی را به یک مجموعه نشانه‌های معنایی قراردادی و نه فقط نشانه‌های چهارگانه گسترش داد، زیرا براساس روابط بافت زبانی اطلاعات واژگانی به شرایط گفت‌و

^{۱۷۷} - Constitutive

^{۱۷۸} - Formal

^{۱۷۹} - Agentive

^{۱۸۰} - Telic

^{۱۸۱} - Hierarchy

^{۱۸۲} - SIMPLE

^{۱۸۳} - Qualia

^{۱۸۴} - Qualia aspects

را نیز مربوط می گردد.

آنچه که باید مد نظر قرار گیرد، روابط دو جانبه و دو طرفه میان هستی شناسی ها و منابع واژگانی است. ما مقدم بودن مطالعات هستی شناسی بر تحقیقات واژگانی و یا بالعکس را مردود می دانیم و به یک پیوند متوازن که پیرو اصول روشن و برپایه مطالعات و بررسیهای نظری است، معتقدیم و توصیه می کنیم که روشهای هستی شناسانه و دانش مرتبط با آن می تواند پشتیبان منابع اطلاعات واژگانی باشد که این منابع واژگانی می توانند متقابلاً در ترسیم هستی شناسی ها سودمند واقع شوند (Prevot, L. et. al. 2010).

۲-۸- از هستی شناسی ها به واژگان

این جهت گیری بر پیشیانی هستی شناسی ها از منابع واژگانی در بکارگیری دانش توصیه می کند، که در مهندسی دانش و ارزیابی ساختار دانش واژگانی و معنایی منابع واژگانی و نیز توسعه آنها کاربرد دارند. این جهت گیری همچنین تأکید بر بدیهی بودن تهیه فهرست اصطلاحات این منابع واژگانی دارند. بسیاری از محققین در اولین گام به یکپارچه سازی مفاهیم و در گام دوم به قرارداد دادن داده های واژگانی در هستی شناسی توصیه و تأکید می کنند.

۳-۸- از واژگان به هستی شناسی ها

جهت گیری دیگر حرکت از منابع واژگانی به سوی دانش هستی شناسانه است، که در واقع مربوط به کثرت اصطلاحات واژگانی در یک منبع هستی شناسی است. این فرآیند راهی مؤثر در ساختن هستی شناسی های گسترده و کلان است که شامل دانش یک دامنه موضوعی^{۱۸۵} و واژگان آن نیز می گردد. تجربیات زیادی از این نوع برای تعمیم سطوح بالا به پائین و یا ابداع کامل هستی شناسی های دامنه وجود دارند. منابع واژگان موجود نیز می توانند به عنوان اصول راهنما در ساخت و تدوین این هستی شناسی بکار گرفته شوند، در نتیجه می توان برای توسعه این دو نوع منبع اطلاعاتی رابطه سیستماتیک و منظمی میان آنان ایجاد کرد.

۴-۸- محوریت پردازش زبان طبیعی NLP

بر اساس رابطه دو جانبه و دو طرفه هستی شناسی و واژگان، حوزه «پردازش زبان طبیعی» آنها را به عنوان ابزاری در ایجاد میانجی آنتولکس می شناسد و بسیاری کاربرد فناوری پردازش زبان طبیعی را بصورت نیمه خود کار^{۱۸۶} در تدوین هستی شناسی ها مؤثر می دانند. این فناوری طیفی از پردازش نحوی

^{۱۸۵} - Domain Knowledge

^{۱۸۶} - Semi-automatic

زبان^{۱۸۷} تا «تحلیل معنایی اصطلاحات» و روابط آنها را در بر می گیرد. و نیز می توان به کاربرد تحلیل های نحوی و معنایی بعنوان نتایج تحقیقات هستی شناسانه در «آموزش» اشاره کرد. همچنین حوزه پردازش زبان طبیعی منابع واژگانی و هستی شناسی را به عنوان ابزار متنوع در بازیابی اطلاعات و جستجوی پرسش و پاسخ و تحلیل های معنایی در فهم و درک عمیق زبان ها نیز بکار می گیرد (Prevot, L. et. al 2010).

۹- مهندسی هستی شناسی

همانگونه که گفته شد، هستی شناسی ها درباره واژگان و معانی آنهاست که از دیدگاه معنی شناختی دقیق و صریح تعریف شده و در قالب الگوهای تفسیر و تعبیر معنایی برای ماشین (رایانه) تدوین یافته است. در واقع، هستی شناسی ها به نوعی توسعه یافته تاکسونومی ها هستند. هستی شناسی ها در مقابل تاکسونومی ها که در فضایی دو بعدی ارائه می شوند، از فضایی سه بعدی یا بیشتر برخوردارند. به عبارتی دیگر هستی شناسی ها ما را مقدور می سازند تا دامنه موضوعی خودمان یا جامعه تخصصی و یا جوامع تخصصی را در ابعادی گسترده تر و با بیان جزئیات معنایی آن تعریف و به اشتراک گذاریم. یک هستی شناسی سعی در ذخیره معانی (که ما به آن معنی شناسی می گوئیم) یک حوزه موضوعی و تخصصی خاص و یا حوزه ای از دانش دارد، که با دانش بشری در آن دامنه موضوعی مطابقت دارد. هستی شناسی نیز این معانی را بصورت اصطلاحاتی وابسته به مفاهیم و روابط میان آنها ترسیم می کند. بعلاوه، یک هستی شناسی اغلب بصورت طبقات، روابط مفهومی، ویژگیها، نمونه ها، مختصات و ارزشگذاری آنها بازنمایی می شود.

هستی شناسی به یکی از زبانهای بازنمایی دانش (مانند زبان وب معنایی OWL/DAML+ (OIL/RDF/s و یا زبان هستی شناسان مانند (Prolog/ QKBC CYCL/ Common Logic/ KIF/ Onto ling ua ارائه می شود. مضاف بر این که این زبانهای هستی شناسی نوعاً بر مبنای منطقی خاص و با توجه به این که این زبان دارای نحو و معنی شناسی خاص خودش است، استوار می باشد. از اینرو در برخی موارد زبانی که هستی شناسی بوسیله آن بازنمایی می شود را زبان مبتنی بر منطق می نامیم.

و سرانجام اگر هستی شناسی بصورت نمودار و تصویر و یا ترجمه متنی (واژگان) ارائه گردد مسئله ای ایجاد نمی کند، زیرا هر دو شیوه کاملاً هم ارزش هستند. مهم قدرت بیان زبانهای است که برای بازنمایی هستی شناسی بکار گرفته می شود. نتیجه منطقی آن است که یک زبان هستی شناسی قوی از

پشتوانه منطق صوری برخوردار است که بواسطه آن هستی‌شناسی ماشین خوان یعنی برای ماشین، قابل تفسیر و تعبیر معنایی می‌شود. منظور از ماشین خوان بودن یعنی آنکه معنی‌شناسی یک الگوی هستی‌شناسی از نظر معنایی قابل فهم و تعبیر و تفسیر توسط ماشین باشد، به عبارتی دیگر رایانه و نرم افزار آن بتوانند معنی‌شناسی الگو را مستقیماً و بدون واسطه انسانی تعبیر و تفسیر کنند. نرم افزاری که از سوی هستی‌شناس‌ها پشتیبانی می‌شود، به سطح مفهومی دانش بشری ارتقاء می‌یابد و نه این که انسان به سطح فهم و درک ماشین تنزل یابد.

تعامل با رایانه در سطح انسانی شکل می‌گیرد و نه در سطح ماشین. این نکته‌ای بسیار حائز اهمیت است و تأکیدی بر ارزشمند بودن هستی‌شناسی‌هاست. در واقع هستی‌شناسی واژگان و مفاهیم مشترک یک حوزه تخصصی را تعریف می‌کند و بدین ترتیب حوزه‌ای از دانش بشری را توصیف و بازنمایی می‌کند که دو ویژگی دارد:

الف - گرایشی از متافیزیک (ماوراء الصبیحه) که بر ماهیت و روابط پدیده‌ها و موجودات متمرکز است.

ب - نظریه‌ای خاص درباره ماهیت وجود و یا انواع هستی‌ها. این تعاریف هستی‌شناسی نشان می‌دهد که این اصطلاح نشأت گرفته از حوزه فلسفه است و بخشی از ماوراء الطبیعه که مطالعه نظام‌مندی از اصول حاکم بر یک حوزه دانشی است و نشانگر ماهیت وجود و حاصل تجربه است.

اغلب در دوران ما تمایزی میان دو نوع هستی‌شناسی مطرح می‌شود، که یکی را با حرف اول بزرگ (Ontology) و دیگری را با حرف اول کوچک (ontology) می‌نگارند و منظور از اولی گرایش فلسفی آن و دومی گرایش مهندسی فناوری اطلاعات آن است که در دهه اخیر مطرح شده است. فناوری اطلاعات تعاریف ذیل را برای هستی‌شناسی ارائه می‌دهد:

هستی‌شناسی واژگان و مفاهیم مشترک بکار گرفته شده در بازنمایی حوزه خاصی از دانش بشری را شامل می‌گردد.

هستی‌شناسی حاصل مهندسی مبتنی بر واژگان و اصطلاحات خاص یک دامنه موضوعی است که در توصیف (بخشی از) واقعیت بعلاوه مجموعه‌هایی از فرضیات مربوط به معانی واژگان و یا به عبارتی مختصات مفاهیم^{۱۸۸} بکار گرفته می‌شود.

اولین تعریف هستی‌شناسی دو قسمت عمده را در بر می‌گیرد:

- توصیف و ارائه حوزه‌ای از دانش
- تعریف اصطلاحات و مفاهیم مشترک

در اینجا تعریف «دامنه» را بار دیگر مرور می‌کنیم:

«دامنه» حوزه‌ای از موضوعات و یا حوزه‌ای از دانش را در بر می‌گیرد که برخی نمونه‌ها از حوزه‌دانشی عبارتند از: پزشکی، برنامه‌ریزی اقتصاد، مدیریت بازرگانی، فیزیک، نساجی، جغرافیای سیاسی و غیره. توصیف حوزه‌ای از دانش در واقع «بیان» کتبی یا شفاهی نکات مهم یک حوزه خاص دانش است. به عنوان مثال در توصیف «تعمیرات خودرو»^{۱۸۹}، احتمالاً درباره موضوعات ذیل صحبت خواهیم کرد:

الف- نوع ماشین: ماشین سواری، ماشین باربری، ماشین‌های خارجی، ماشین‌های اسپرت...

ب- نوع موتور ماشین: براساس نوع سوخت مصرفی شامل: گازوئیلی، دیزلی، الکتریکی...

ج- برحسب سال ساخت موتور: ۹۶-۱۹۹۵، فورد کاروس، ...

ه- قطعات خودرو (موتور، دستگاه خنک کننده، سیستم الکتریک، بدنه، چرخ و غیره)

و- ویژگی‌ها و مختصات (موتور ۱۲ سلندر متالیک و...)

هنگامی که به توصیف حوزه‌ای از دانش (دامنه) می‌پردازیم، در واقع مهمترین پدیده‌های یک دامنه و مختصات و نیز روابط میان آنها را توصیف می‌کنیم، مثلاً وقتی به نوشتن مقاله‌ای تخصصی و یا کتابی در مورد "خودرو" مشغول هستیم، بایستی به قواعد حاکم بر دامنه موضوعی مانند قوانین تشخیص نقایص، که مشخص می‌کند چه چیزی در سیستم خودرو دارای نقص (خراب) است و نیاز به تعمیرات دارد مسلط باشیم (مثلاً بیان این مسئله که وقتی خودرو استارت نمی‌زند و یا روشن نمی‌شود، باتریها و ارتباط آنها با موتور معاینه فنی و تمیز کنید).

از اینرو توصیف پدیده‌ها و جزئیات آن می‌تواند نوعی هستی‌شناسی تلقی گردد مانند:

- طبقات (مفاهیم عام) در بسیاری از دامنه‌های موضوعی

- نمونه‌ها (مثال‌ها) اجزاء خاص

- روابط معنایی میان آن نمونه‌ها

- مختصات (ارزش مختصات) پدیده‌ها

- کارکرد و فرآیند دربرگیرنده آن

- محدودیت‌ها و قواعد در برگیرنده آن اشیاء

لازم به ذکر است که علاوه بر توصیف یک حوزه دانشی، بایستی آنها همچنین دقیق بازنمایی کنیم. معنی بازنمایی چیست؟ بازنمایی یعنی رمز گذاری یک توصیف، به طریقی که کاربری این توصیف را برای دیگران مقدور سازیم. یک توصیف بستگی به واژه‌ها، عبارات موجود در یک زبان طبیعی (مانند: چینی، انگلیسی، عربی و فارسی) دارد که در واقع همان واژگان، اصطلاحات و جملات است

که اصطلاحات را بهم دیگر متصل و روابط میان این اصطلاحات را بیان می کند. (در اینجا ما واژگان و اصطلاح شناسی را بصورت مترادف بکار گرفته و کلمه اصطلاح را برای یک واژه منفرد بکار می گیریم). بنابراین بازنمایی بدین معنی است که ما برای بازنمایی یک توصیف از اصطلاحات و عبارات و جملات استفاده می کنیم. ما اصطلاحات را در «واژگان ذهنی» خود تعریف می کنیم و یا قبلاً تعریف کرده ایم و سپس تلفیقی از این اصطلاحات را به طریقی انجام می دهیم، به طوری که بیشترین معنا را در یک حوزه دانشی ارائه دهیم.

هرچند در فناوری اطلاعات ما روشهای کمی پیچیده تری برای بازنمایی مفاهیم بکار می گیریم. اما در بازنمایی کاربرد یک توصیف و تعاریف را نیز در فناوری اطلاعات بکار می گیریم. به عبارتی دیگر، ما الگویی معنایی را برای نرم افزار ابداع می کنیم تا از بهترین و بیشترین اطلاعات بهره برداری نماییم. از اینرو طبقات، اجزاء و عناصر، نمونه ها، روابط معنایی و قواعد حاکم بر «حوزه دانشی» را ارائه می دهیم. ما اصطلاحات زبان طبیعی را برای برچسب گذاری مفاهیم مبنایی آنها بکار می گیریم. معانی حوزه دانشی مبتنی بر طبقات، عناصر و اجزاء روابط معنایی آنها استوار است. آنتولوژیها را نوعاً براساس زبان منطق رمز گذاری می کنیم که زبان بازنمایی دانش است و بیش از زبان طبیعی بکار می آید، زیرا می خواهیم توصیف خود را حتی المقدور روشن، دقیق و بدون ابهام ارائه دهیم. همچنین هدف ما دسترس پذیر نمودن معانی و مفاهیم، برای کاربرد در فناوری اطلاعات می باشد. بازنمایی که با تعریف اصطلاحات (واژگانی که به عنوان برچسب مفاهیم عمل می کنند) سرو کار دارد و این در واقع به معنای تعریف مفاهیم و روابط آنهاست که در پشت این برچسب ها قرار دارد و ساختار معنایی الگوی مفهومی حوزه دانش مورد علاقه ما را نشان می دهد.

اولین تعریف با توصیف و بازنمایی یک حوزه دانشی سروکار دارد. اما در تعریف دوم این سوال مطرح است که از آنتولوژی چه می دانیم؟ تعیین مختصات در مفهوم سازی به چه معناست؟ برای روشن نمودن این تعریف بایستی به بخش هایی از تعریف اول پردازیم. مفهوم سازی نوعی تفکر در مورد بخشی از جهان هستی است. هنگامی که از جهان هستی و یا بخشی از جهان ادراکی داریم، جزئیات الگوی معنایی آن بخش از جهان هستی را در ذهنمان مجسم کرده ایم. به عنوان مثال وقتی موضوع «تعمیرات ماشین» را ادراک می کنیم، احتمالاً تعدادی از تصورات ذهنی را در مورد اتومبیل، سیستم های آن، اجزاء سازنده آن، محل نگهداری اتومبیل و یا تعمیرگاه آن، مکانیسم های واحدهای آن و غیره را در ذهن مرور می کنیم. چنانچه مجبور به توصیف این تصورات شویم، احتمالاً به اولین تعریف مراجعه می کنیم و با بیان آن در قالب اصطلاحات مربوط به آن اجزاء سازنده اتومبیل و تعمیرات آن و مختصات و روابط معنایی آن می پردازیم. با ارائه طریق خاصی از تفکر در مورد بخشی

از جهان هستی (دامنه موضوعی) یا به عبارتی، مفهوم سازی، طریقه خاصی از توصیف را برای بیان این موضوعات انتخاب می‌نمائیم و چون جزئیات را به طریقی دقیق و جزء به جزء بیان می‌کنیم در واقع مختصات آن پدیده را تعیین می‌کنیم.

«سیک» اولین سیستم بازنمایی دانش مبتنی بر هستی‌شناسی است که اجرا و توسعه آن در سال ۱۹۸۴ در شرکت «میکرو الکترونیک و فناوری رایانه»^{۱۹۰} (MCC) در اوستین / تگزاس شکل گرفت و هم اکنون بصورت تجاری در سایت ([Http://www.Cyc.com](http://www.Cyc.com)) در دسترس می‌باشد.

مسئله حساس و مهم در ایجاد هستی‌شناسی‌ها این است که چه چیزی سبب برتری یک هستی‌شناسی نسبت به دیگری است و کدام اجزاء و عناصر هستی‌شناسی بویژه بعنوان الگوی منطقی مفاهیم ارائه می‌گردد و آنرا از دیگر الگوهای معنایی متمایز می‌گرداند و از دیدگاه مدیران فناوری اطلاعات و مهندسين اطلاعات و از نظر کاربرد در فناوری «وب معنایی» و برنامه‌ریزی آتی مدیریت استراتژیک اطلاعات یک سازمان تأثیرگذار خواهد بود. در دنیای «وب» ما نوعاً با اسناد و مدارک سروکار داریم، از این رو ساختار نحوی و معنایی آن و نیز نوع نرم‌افزاری که این اسناد و داده‌ها را سازماندهی می‌کند، جزو علایق اولیه ما محسوب می‌شود. ساختار نحوی اسناد بوسیله زنجیره‌ای از کاراکترها یعنی الفبای ساختار متون و یا رمزگذارهای دیگر مانند گرافیک، منابع صوتی - تصویری و غیره بازنمایی می‌شود.

معنی‌شناسی این متون شامل نشانه‌ها و سمبولهایی است که محتوایی معنایی آنرا بصورت یک دامنه موضوعی و محتوای اسناد و اطلاعات مورد علاقه کاربران منعکس می‌نماید. در دنیای «وب» معنایی XML زبانی است که دارای ساختار نحوی است و می‌تواند محتوای نحوی و معنایی اسناد را بازنمایی کند.

در هر صورت پایگاههای اطلاعاتی، مورد استفاده در «وب معنایی»، به الگوهای معنایی و هستی‌شناسی‌ها دارای ساختار و نظام معنایی نیاز دارند. بعنوان مثال در قالب یک پایگاه داده ابزاری برای توصیف و تعیین ساختار نحوی و معنایی پایگاه داده ارائه می‌شود. این ساختار معمولاً بوسیله مدیریت سیستم پایگاه داده تعیین و توصیه می‌شود که می‌توان او را بعنوان تعبیر کننده معنایی و یا ارزشیابی کننده داده‌ها با عنایت به الگوی ساختاری پایگاه داده شناخت.

الگوهای مفهومی، آنگونه که در UML طراحی می‌شوند نیز مبتنی بر ساختار هستند. ساختار الگوهای مفهومی در ساختار سلسله مراتبی و روابط اعم و اخص (طبقه، زیر طبقه) منعکس می‌شوند. این ساختار منطقی در درجه‌بندی تعبیر معناشناختی عناصر سازنده آن تأثیرگذار است. دانش (آنگونه که

^{۱۹۰} - Micro Electronics & Computer Technology Corporation (MCC)

در هستی‌شناسی‌ها رمز‌گذاری شده‌اند) با ارائه نشانه‌های معنایی پیچیده در بازنمایی، جنبه‌هایی از عالم مقال را منعکس می‌کند که ما آنرا حوزه و یا دامنه موضوعی می‌نامیم. تعبیر و تفسیر معنایی بصورت طراحی نقشه معنایی میان داده‌های ساختارمند و الگوهای معنایی تک‌تک عناصر و نیز روابط معنایی میان آن عناصر شکل می‌گیرد و نوعاً در ذهن انسان به عنوان الگوهای معناشناختی قرار دارد. ما به عنوان انسان این الگوهای معناشناختی را به عنوان معنای نشانه‌های مبتنی بر پدیده‌های جهان هستی و روابط میان آن اشیاء و پدیده‌ها ادراک می‌کنیم. وقتی یک متن یا یک سند را می‌بینیم، با توجه به الگوهای ذهنی خودمان تعبیر و تفسیر می‌کنیم و معنای آنرا در می‌یابیم. اگر بخواهیم در اشاعه دانش نهفته در یک سند دخیل باشیم، بایستی آن سند را برای دیگر انسانها دسترس پذیر نمائیم. قابل پیش‌بینی است که آنها نیز تعبیر معنایی براساس الگوهای ذهنی خاص خود از این سند و نشانه‌های معنایی آن خواهند داشت. بنابراین در یک سند علمی هیچ دانشی یافت نمی‌شود که بدون تعبیر و تفسیر ذهن انسان بماند در نتیجه آن را فارغ از سمبل‌های بدون معنی می‌گرداند (Obrst & Liu, 2003).

اگر بخواهیم در اشاعه دانش نهفته در اسناد علمی در وب معنایی از رایانه کمک بگیریم، لازم است که بخشی از فرآیند تعبیر معنایی را بصورت خودکار (اتوماتیک) درآوریم. در این راه نیازمند آنیم که به طریقی بخشی از الگوهای ذهنی خود از دامنه موضوعی را بصورت کاربردی برای رایانه تعریف و بازنمایی کنیم. و در این راه هستی‌شناسی‌ها این توانایی را در بازنمایی دامنه موضوعی دارند و این مهمترین بخش ساختار "وب معنایی" است.

نرم افزارهایی که در آینده (مبتنی برعوامل هوشمند، خدمات و سرویس‌دهی شبکه‌های معنایی و غیره) طراحی می‌شوند، قادر خواهند بود که دانش بشری را در قالب هستی‌شناسی‌ها رمزگذاری کنند تا بتوان آنها را در ادراک و تفسیر معنایی اسناد در شبکه بکار گرفت. در نظریه زبان صوری که شامل نحو و معناشناسی اشیاء و پدیده‌ها می‌شود، در کاربرد واژگان و نیز برنامه‌ریزی زبانها و پایگاههای داده و ساختارمند نمودن آنها تأکید شده است. هستی‌شناسی‌ها سعی می‌کنند تعبیر و تفسیر معنایی این الگوهای صوری را در قالب معنی‌شناسی واژگان و محتوای معنایی آنها محدود سازند، هیچکدام از ساختارهای معنایی دیگر مانند تاکسونومی‌ها (رده‌بندی) تزاروس‌ها و غیره در این حدا توانا نیستند. این الگوهای معنایی منطبق بر الگوهای ذهنی و مبتنی بر واژگان ذهنی انسان است. هستی‌شناسی‌ها می‌توانند این الگوهای معنایی را برای ماشین (رایانه) بازنمایی کنند و ادراک ماشین را به ادراک انسان نزدیک کنند.

از اینرو مهندسین نرم‌افزار و علوم کامپیوتر زبانهای سطح بالایی را طراحی کرده‌اند که به سطح

ادراک انسان نزدیک و تا حدود زیادی همدریف آن باشد. در تعبیر معنایی ماشینی و به زبانی منطقی و ساختارمند که در بازنمایی دانش مؤثر است، نشانه‌ها و سمبولهایی که انسان درک می‌کند بایستی بایک فرآیند بازسازی و براساس الگوهای منطقی و شرایط تطبیقی با ادراک انسان در قالب‌های صوری برای ماشین تعریف گردند.

همانگونه که گفته شد، هستی‌شناسی معمولاً در قالب زبان بازنمایی دانش و مبتنی بر الگوهای منطقی و به صورت ارائه طبقات، اجزاء و عناصر، ویژگیها و روابط معنایی ارائه می‌گردند. برخی از ابزارهای مهندسی هستی‌شناسی کاربرد خودکار (اتوماتیک) هستی‌شناسی‌ها را مقدور می‌سازد و در نتیجه آن کاربردهای هوشمند جستجو و بازایی اطلاعات از طریق نظام مفاهیم، تصمیم‌سازی‌های مدیریتی، بازشناسی گفتار زبان طبیعی، مدیریت دانش، پایگاههای داده هوشمند و تجارت الکترونیک را امکان پذیر می‌سازد.

آنتولوژی‌ها در تاکسونومی (دانش ساختارمند سلسله مراتبی)، در اصطلاحنامه (ساختارمند بودن واژگان و مترادف‌ها)، در الگوهای مفهومی (دارای ساختار پیچیده‌تر معنایی) و در نهایت نظریه منطق مدار (دارای ساختار غنی و بسیار پیچیده معنایی) کاربرد دارند.

بنابراین یک هستی‌شناسی شامل هر دو بعد، یعنی واژگان که برای توصیف و بازنمایی یک حوزه دانشی بکار میرود و نیز معانی این واژگان که از نظر نحوی زبانی دارای انواع اصطلاحات است که متناظر با معنی‌شناسی صوری است، معانی ساختارمند زبان و ترکیبات زبانی را نیز شامل می‌شود.

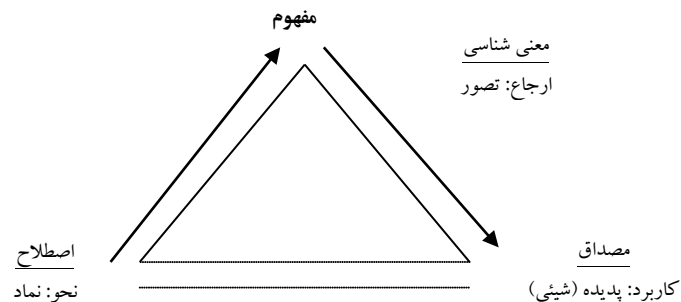
آخرین گرایش رایانه‌ای که به توسعه و مدیریت هستی‌شناسی‌ها اشاره دارد تحت عنوان «مهندسی هستی‌شناسی» معروف شده است. «مهندسی هستی‌شناسی» معمولاً ویژگیهای یک هستی‌شناسی را (مانند نظریه منطقی) در قالب اصطلاحات و براساس «نظام گزاره‌های منطقی» و قواعد منطقی و الگوهای منطق صوری و به صورت یک نظریه منطقی ارائه می‌دهد. از دیدگاه «مهندسی هستی‌شناسی»، هستی‌شناسی شامل واژگانی است که مدخل‌های واژگانی و روابط معنایی آنها را براساس مدل مفهومی یک دامنه موضوعی و زنجیره‌های معنایی آن تعریف می‌کنند. هر دو آنها یعنی واژگان و زنجیره‌های معنایی مورد نیاز هستند تا الگوهای اطلاعاتی منطبق با دامنه موضوعی در جهان هستی بازنمایی گردد.

۹-۱- اصطلاح در مقابل مفهوم / اصطلاحنامه در مقابل هستی‌شناسی

به منظور فهم بیشتر هستی‌شناسی و ویژگیهای آن، به تفصیل به تمایز میان اصطلاح و مفهوم می‌پردازیم (ISO 704, 2000).

یکی از راههای توصیف آن، ایجاد تمایز میان اصطلاحنامه و هستی‌شناسی (بوژه هستی‌شناسی مبتنی

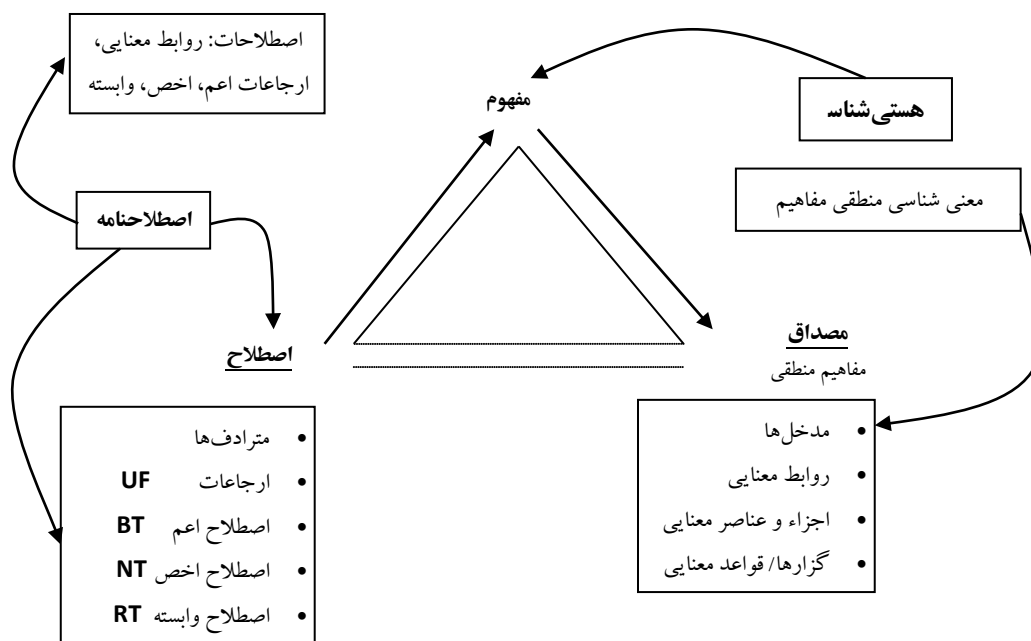
بر نظریه منطقی) است. ابتدا به توضیح در مورد «مثلث معنا شناختی» می‌پردازیم:



نمودار ۴: مثلث معنی‌شناختی (Daconta et. al)

اولین عنصر «اصطلاح» است که در واقع نماد (برچسب گذاری مفهوم) و یا واژه است و قواعد نحوی حاکم بر آنها می‌تواند بصورت عبارت یا جمله نیز ارائه شود. نمادها به تنهایی بی معنی هستند، مگر آن که در یک مثلث معنایی با مفاهیم و مصداق‌های آنها در جهان هستی پیوند حاصل نمایند. اصطلاحنامه به مفهوم نیازمند است تا بتواند میان اصطلاح و پدیده‌های جهان هستی ارتباط برقرار کنند و ارجاعات را ایجاد نماید. اصطلاحنامه معمولاً با اصطلاحات و مفاهیم سروکار دارد، در حالی که هستی‌شناسی با مفاهیم و ارجاع به مصداق‌ها در جهان هستی پیوند می‌یابد. اصطلاحنامه در ابتدا با طبقه بندی مفاهیم دامنه موضوعی و روابط معنایی آنها شکل می‌گیرد، مانند: اصطلاحنامه روزه ۱۹۱۶، که بمنظور طراحی نظام مفاهیم جهت جستجو و بازیابی اطلاعات علمی - فنی طراحی شده و از نظر معنی شناختی در حد ارائه ویژگیهای معنایی روابط بیان اصطلاحات (برچسب‌های مفاهیم) و رده‌بندی اعم و اخص مفاهیم بود.

آنچه که در یک اصطلاحنامه باید بدانیم این که یک اصطلاح بصورت گره‌ای در اصطلاحنامه و برای تفکیک معنایی آن از دیگر اصطلاحات و رفع ابهامات معنایی ایجاد شده است و بصورت اعم یا اخص‌تر از دیگر اصطلاحات نمایش داده می‌شود.



نمودار ۵: اصطلاحنامه در مقابل هستی‌شناسی (Daconta et. al. 2003)

هستی‌شناسی سعی در بازنمایی معنایی مفاهیم و روابط پیچیده میان آنها را دارد و بدین جهت اجزاء عناصر، ویژگیها، ارزشگذاری، زنجیره‌های معنایی و قواعد حاکم بر آنها را تعیین می‌کند. بنابراین هدف هستی‌شناسی از اصطلاحنامه متمایز است. هستی‌شناسی در صدد اکتساب و بازنمایی معانی یک دامنه موضوعی است که با پدیده‌های جهان هستی منطبق باشد، زیرا هدف آنها شبیه سازی مفاهیم ذهنی انسان و ساختار دامنه موضوعی آن در جهان هستی است. علاوه بر این هستی‌شناسی مستقیماً مورد کاربرد در برنامه‌ریزی نرم‌افزارهایی است که عملکرد آنها بازیابی دانش مورد نظر براساس دامنه موضوعی است.

البته هستی‌شناسی نیز همانند اصطلاحنامه بایستی گسترش، روزآمدسازی، ویرایش و کاربرد مجدد یابد و علاوه بر کاربرد تزاروس به معنی‌شناسی دقیق پیچیده و غنی مفهومی بپردازد. بنابراین هستی‌شناسی‌ها دو نوع دانش را ارائه می‌دهند:

۱- دانش و اطلاعات عام مربوط به توصیف طبقه‌بندی و کاربرد الگوی معنایی یک دامنه موضوعی.

۲- دانش و اطلاعات مربوط به نمونه‌ها و مثالها که شامل توصیف ویژگیهای معنایی الگوی معنایی می‌شود.

در یک پایگاه داده و زبان صوری یا طبیعی اولین نوع دانش و اطلاعات معنایی مربوط به «مفهوم»^{۱۹۱} و اطلاعات ثانوی مربوط به «گستره»^{۱۹۲} آن است. در زبان طبیعی یا صوری، یک توصیف و ارائه ویژگیهای معنایی، مفهومی است در حالی که توصیف اشیاء واقعی (مفردات) در جهان هستی که توصیف منطبق با آن است، گستره مفهومی را می‌رساند. بدین معنی که در یک هستی‌شناسی یک مجموعه عناصر ساختارمند که معانی ویژه‌ای دارند توصیف می‌شود که به آن الگوی معنایی گفته می‌شود و اطلاعات مربوط به جنبه‌های (وجوه) خاصی از جهان هستی را در قالب یک الگوی معنایی بازنمایی می‌کنند. به عنوان مثال در یک هستی‌شناسی می‌توان اطلاعات معنایی مربوط به دامنه‌های موضوعی متفاوت را ارائه کرد، مانند: اشخاص، موقعیت‌ها، رخدادها و غیره و نیز روابط میان این اجزاء مثلاً موقعیت یک فرد، زمان رخداد و روابط را بازنمایی کرد.

در بحث پیرامون هستی‌شناسی‌ها نیازمند به تفکیک سطوح بازنمایی اطلاعات هستیم. این تفکیک از آن جهت لازم است که هستی‌شناسی را می‌توان بعنوان زبان بازنمایی اطلاعات شامل واژگان، نحو و معنی‌شناسی یک متن در نظر گرفت. علاوه بر این در هستی‌شناسی‌ها ارائه محتوا صورت می‌گیرد و از اینرو به زبانهایی برای بازنمایی محتوایی نیازمند است که معمولاً «زبان بازنمایی دانش»^{۱۹۳} نامیده می‌شود.

از اینرو به دو سطح از بازنمایی دانش قائل هستیم: ۱- سطح «زبان بازنمایی دانش» که سطح فرداده^{۱۹۴} و ۲- سطح بازنمایی اشیاء (پدیده‌ها) که «سطح محتوایی»^{۱۹۵} گفته می‌شوند.

همچنین به سطح ۳ نیازمندیم که در آن ویژگیها، اجزاء و عناصر معنایی در طبقات هستی‌شناسی موجودیت می‌یابند و در آن دانش جهانی به شکل مفاهیم یک هستی‌شناسی عرضه می‌شوند.

سطح ۱- بازنمایی دانش

سطح ۲- بازنمایی مفاهیم هستی‌شناسی.

سطح ۳- بازنمایی اجزاء معنایی هستی‌شناسی

در سطح اول زبان بازنمایی دانش (سطح فراداده)، تعابیر معنایی مورد استفاده در سطح مفهومی هستی‌شناسی ارائه می‌شوند و شامل طبقات، روابط، عناصر و اجزاء معنایی و غیره می‌شوند. نمونه‌هایی از زبانهای بازنمایی دانش (KR) که در وب معنایی بکار گرفته می‌شوند عبارتند از: KL-ONE،

^{۱۹۱} - Intension

^{۱۹۲} - Extension

^{۱۹۳} - Knowledge Representation Language

^{۱۹۴} - Meta Level

^{۱۹۵} - Object (Content) Level

LOOM، Classic، Ontolingua، فرمات تبادل اطلاعات (KIF)، Cycl و UML و نیز RDF/s، DAML+OIL و OWL، (Mc Guinness & Van Harmelen 2002).

در سطح دوم یا سطح مفهومی هستی‌شناسی (OC)، ساختارها و تعابیر معنایی سطح بازنمایی دانش (KR) بکار گرفته می‌شوند. در این سطح محتوای مورد توافق و مشترک میان متخصصین بصورت الگوهای معنایی در می‌آید و دامنه موضوعی پیرامون اشخاص، موقعیت‌ها، رخدادها و غیره طراحی می‌شود.

در سطح سوم که پائین‌ترین سطح محسوب می‌گردد، تعابیر معنایی شامل عناصر معنایی سازنده هستی‌شناسی و سطح مفهومی است، از اینرو این سطح شامل دانش پایه واقعی است که عناصر و مفردات عینی را بیان می‌کند. مسئله مهم در ادراک (شناخت) و توسعه هستی‌شناسی‌ها، ارائه نقشه مفهومی آنهاست، مسئله‌ای که بر همه چیز در فناوری اطلاعات تأثیر گذار است، و آن مسئله بازنمایی «معنا» برای سیستم‌های کاربردی نرم‌افزاری در پایگاههای داده بنیان و گردآوری اسناد است.

همواره باید این مسئله را مد نظر قرار دهیم که بازنمایی «معنا» در سیستم‌های بازیابی اطلاعات و کاربرد در پایگاههای اطلاعاتی و گردآوری اسناد و بازنمایی «معنا» در جامعه اطلاعاتی و متخصصین موضوعی و جهان پیرامون آن مطابقت یافته و نمایانگر «نقشه مفهومی» آنها باشد و در میان مجموعه هستی‌شناسی‌های متعدد یکسان سازی و تلفیق مفاهیم صورت پذیرد، که به معنای واقعی کاری است بسیار دشوار و زمان بر و هزینه‌های گزافی را می‌طلبد.

اغلب زبانها و ابزار پشتیبان آنها تسهیلاتی را برای تعریف و تلفیق میان هستی‌شناسی دارند. ساده‌ترین مکانیسم شرایط ادغام هستی‌شناسی‌ها است که طی آن یک هستی‌شناسی به طور کامل به هستی‌شناسی جاری ورود می‌یابد و تمامی مفاهیم و روابط معنایی هستی‌شناسی اقتباس شده در فضای جدید دسترسی پذیر می‌گردد. هرچند این مفاهیم وارداتی و روابط معنایی آنها کاملاً قابل فهم و درک برای برای هستی‌شناسی که از قبل ایجاد شده نیست، اما می‌توان با یکدست کردن و ویرایش معنایی این تلفیق و ادغام را میان مفاهیم جدید و قدیم و روابط معنایی آنها ایجاد کرد. این ادغام و تلفیق سبب تثبیت مفاهیم و روابط معنایی آنها می‌شود. و در این تلفیق می‌توان با حفظ و نگهداری هر دو هستی‌شناسی، مفاهیم و روابط آنها را بطور مجزا اما مرتبط با یکدیگر حفظ کرد.

ابزار پشتیبان چون پروتژه (Protégé) و آنتولینگوا (Ontolingua) این فرآیند تلفیق و ادغام را بهتر انجام می‌دهند. زبانهای «وب معنایی» مانند DAML + OIL و OWL ساختارهای معنایی دارند که اقتباس هستی‌شناسی‌ها مقدور می‌سازند و نیز آشکار می‌سازند که دو مفهوم در معنا همانند هستند یا خیر.

علاوه بر این زبان OWL ما را در آشکار ساختن همانندی دو ویژگی در معنی همانند یاری می دهد که قدم بزرگی برای پشتیبانی از طراحی نقشه مفهومی خواهد بود.

بطور کلی زبانهای وب معنایی که هستی شناسی ها را بازنمایی می کنند شامل^{۱۹۶} RDF و RDF/s است که الگوی زبانی توصیف منابع محسوب می شود و در پروژه توسعه تحقیقات دفاعی آمریکا (DARPA) زبان DAML^{۱۹۷} بکار گرفته شده و بجز زبان OIL^{۱۹۸} که برای بازنمایی هستی شناسی بکار میرود. و سرانجام زبان OWL که هستی شناسی را در سطح وب معنایی بازنمایی می کند.

در مورد زبان DAML که زبان هستی شناسی در وب معنایی است و در سال ۲۰۰۰ میلادی در آمریکا مقدمات آن شروع شد، می توان گفت که تاکنون در سطح کشورهای اروپایی توسعه و پیشرفت زیادی یافته است (Bechhofer et. al. 2000).

نکته قابل توجه آن است که تمامی این زبانها براساس نحو XML بصورت زیر بنایی برای تعامل هستی شناسی و در محیط وب عمل می کند، و در ایجاد پایگاههای داده و تطبیق با دیگر پایگاهها تأثیر گذار است. در مرحله بعد RDF/s است که زبان ساده ای برای بیان مفاهیم و روابط میان عناصر معنایی هستی شناسی است و باز هم براساس نحو XML عمل می کند. در مراحل بالاتر زبانهای OIL+ DAML و OWL که هستی شناسی را با بیان دقیق تر و موشکافانه ای بیان می کند، اما در میان ساختارهای معنایی اجزاء و عناصر از RDF/s استفاده می کند.

هر دوی زبانهای فوق الذکر مستقیماً در قالب XML داده ها را تعریف می کنند. باید تأکید شود که علیرغم این که همه این زبانها براساس نحو XML ساخته شده اند، لازم است که ابزار تعبیر معنایی خاص هر زبان برای درک نتایج حاصله ایجاد گردد. از اینرو بالاترین سطح زبانی یعنی تعبیر معنایی OWL برای ایجاد ساختارهای معنایی خاص خود بهره گیرد.

تاکنون دقیق ترین تعبیر معنایی بواسطه زبان OWL بدست آمده است. از بحث گذشته می توان نتیجه گیری کرد که ادغام و تلفیق و ترسیم نقشه معنایی با استفاده از اقتباس و توسعه هستی شناسی ها که نمایانگر دامنه های موضوعی و نیز میان رشته ای هستند، می توان فناوری هستی شناسی ها را توسعه داده و نگرش مفهومی واژگان ذهنی انسان را برای ماشین (رایانه) قابل درک و خوانا کرد، به شکلی که بتوان در جستجو و بازیابی اطلاعات از این ساختار منطقی بیشترین و بهترین استفاده را در حوزه فناوری اطلاعات نمود.

^{۱۹۶} - Resource Description Framework (RDF)

^{۱۹۷} - Defense Agent Mark up Language (DAML)

^{۱۹۸} - Ontology Inference Layer

دو عامل اصلی در مدیریت دانش، «کاربرد فناوری» و «نیروی انسانی» است. این نکته قبلاً در آثار "داون پورت" ذکر شده است. هم اکنون فناوری اطلاعات پشتیبان نظام‌های مدیریت دانش به حافظه سازمانی مراکز سپرده شده است. حافظه سازمانی تلفیقی از دانش صوری، نیمه صوری و غیر صوری است که به منظور تسهیل در دستیابی، اشتراک منابع و کاربرد مجدد حافظه اعضا یک سازمان و برای حل مشکلات فردی و جمعی سازمان بکار گرفته می‌شود. و به عنوان مثال بخشی از فرآیند شغلی افراد را تشکیل می‌دهد.

ساختار دانش که زیر بنای این نظام مدیریت دانش را شکل می‌دهد نوعی هستی‌شناسی است (Staab and Suder 2004)، که طبق آن می‌توان روش‌ها و شیوه‌ها را انتخاب نمود (Davies et. al 2006).

این شیوه‌ها، رویکردی مشترک به سوی ساختارهای مهندسی دانش و ابزاری برای مهندسین دانش، متخصصین موضوعی و دیگران است، تا بتوانند با روش‌هایی مانند تجزیه و تحلیل پرسشنامه‌ها و مصاحبه‌ها به نتیجه مطلوب برسند. در حالی که گروه کاربران چنین هستی‌شناسی‌ها می‌توانند زیاد باشند، تدوین و طراحی خود آن توسط گروه کوچکی خبره شامل متخصصین موضوعی که الگوهای معنایی را برای ارائه مفاهیم دانش و نیز مهندسین هستی‌شناسی که ساختار و فرمول بندی آنرا به عهده دارند، انجام می‌پذیرد.

از اینرو مدیریت دانش غیر متمرکز به طور گسترده اهمیت می‌یابد. «وب معنایی» توسعه یافته (Berners-lee et. al. 2001) و گسترش آنرا در موارد بسیاری برای ارائه این الگوهای جدید پشتیبانی می‌کنند. بهمین دلیل شیوه‌ها بر اساس نظام مدیریت دانش سنتی و متمرکز دیگر قابل کاربرد نیستند.

برخی راه‌حل‌های فنی پا به پای نظام مدیریت دانش عمل می‌کنند.

در مهندسی دانش اعتمادی راسخ و همه جانبه حاکم است که شیوه‌هایی برای ساختار نظام‌های دانش - بنیان و طرح‌های مهندسی دانش در به ثمر رسیدن تلاش‌ها و دسترسی به موقع به اهداف موثر است. با استفاده از هستی‌شناسی‌ها در نظام‌های دانش - بنیان، همان شیوه‌های دسترس پذیری اعمال شده در فرآیند مهندسی آنتولوژی از طریق رویکردهای مختلف ذیل اعمال می‌شود:

الف- فرآیندهای مدیریت هستی‌شناسی شامل تعاریف برنامه‌ریزی و زمانبندی مراحل مهندسی هستی‌شناسی است که طی آن بایستی مراحل مکانیسم کنترل و تضمین کیفیت نیز در نظر گرفته شود.

ب- هنگامی که یک هستی‌شناسی تدوین و توسعه می‌یابد، باید فرایندهایی برای سازگاری با محیط و تسهیل اجرای آن تعریف شوند. پس از تصمیم‌گیری پیرامون ساخت یک

هستی‌شناسی، مهندسين هستی‌شناسی نیاز به فرآیندهایی برای تخصیص ویژگی‌ها، مفهوم‌سازی، فرمول‌بندی و اجرای هستی‌شناسی دارند. سرانجام کاربران و مهندسين به راهنمایی برای حفظ و نگهداری، تکثیر، کاربرد و توسعه هستی‌شناسی نیازمند هستند.

ج- برای کمک به ایجاد یک هستی‌شناسی، شماری از فعالیت‌های حمایتی مهم بایستی انجام پذیرد که شامل آموزش و فراگیری دانش لازم، ارزیابی، ابداع، ادغام و نظام‌مند نمودن و ترکیب بندی مدیریت آن است. این فعالیت‌ها در تمامی مراحل فرآیند ایجاد و مدیریت آن انجام می‌پذیرد. آموزش دانش لازم می‌تواند بصورت متمرکز یا غیر متمرکز صورت پذیرد. یادگیری هستی‌شناسی راهی است برای حمایت و پشتیبانی آموزش دستی دانش با کمک فناوری آموزش الکترونیک.

د- مهندسی هستی‌شناسی، بایستی معیارهای ارزیابی سنجش کیفی آنتولوژی ایجاد شده را فراهم نماید. که البته برای هستی‌شناسان این تصمیمگیری با مشکلاتی مواجه است، زیرا الگوسازی در بسیاری از موارد براساس ذهنیات افراد صورت می‌پذیرد. بررسی کلی معیارهای سنجش و ارزیابی هستی‌شناسی‌ها را می‌توان در مطالعات گومز - پرز یافت. علاوه بر این می‌توان به شیوه ارزیابی و معیارهای سنجش براساس داده‌های آماری و برگرفته از اصول فلسفی رجوع کرد. یکی از این رویکردهای موجود ارزیابی هستی‌شناسی است (Guarino and Welty, 2002).

۹-۲- انواع روشهای مهندسی هستی‌شناسی

همانطور که گفته شد روشهای مهندسی هستی‌شناسی به طور مفصل در کتاب گومز - پرز (۲۰۰۳)^{۱۹۹} آمده است. کریستانی و کوئل (۲۰۰۵)^{۲۰۰} نیز چارچوبی برای مقایسه شیوه‌های مهندسی هستی‌شناسی و ارزیابی هر یک از آنها را تعیین کرده‌اند. در پروژه «آنتو وب»^{۲۰۱}، اعضاء پروژه، اصول راهنمایی و کاربرد آنها در صنعت را جمع‌آوری کرده‌اند، که متمرکز بر کاربرد هستی‌شناسی در تجارت الکترونیک^{۲۰۲}، بازیابی اطلاعات، شبکه‌های اجتماعی در وب است. توصیفی کاربردی برای شروع ساخت آنتولوژیها را می‌توان در مطالعات «نوی و مک کوائین ۲۰۰۱» یافت. برخی از رویکردها از این قرارند:

^{۱۹۹} - Gomez-Perez et al. (2003)

^{۲۰۰} - Cristani-Cuel (2005)

^{۲۰۱} - Onto Web Project

^{۲۰۲} - E-Commerce

KADS - ۹-۲-۱

این رویکرد در واقع روشی برای ایجاد هستی‌شناسی نیست ، اما ابعادی از اشتراک در مدیریت دانش با استفاده از تحلیل و مهندسی دانش و طراحی و استقرار نظام‌های اطلاعاتی دانش - بنیان را پوشش می‌دهد. این رویکرد متمرکز بر مراحل اولیه کار کردهای توسعه مدیریت دانش دارد. این روش مراحل تسهیل در فرآیند ساخت هستی‌شناسی از طریق جمع‌آوری نظرات کاربران و ابزارهایی مانند مصاحبه و توزیع پرسشنامه جهت اعمال موفقیت‌آمیز مدیریت دانش دارد و از اینرو پشتیبان مهندسين دانش و متخصصين موضوعی می‌باشد. (Schreiber et al. 1999)

DOGMA - ۹-۲-۲

پایگاه اطلاعاتی است که منابع اطلاعاتی را بر مبنای هستی‌شناسی دسته‌بندی کرده و قواعد و محدودیت‌های آنرا بیان می‌نماید. (Jarrar and Meeresman. et al. 2002)

Enterprise Ontology - ۹-۲-۳

این رویکرد سه مرحله مهم و اصلی در مهندسی دانش دارد:

الف - شناسایی اهداف

ب- دریافت و کشف مفاهیم و روابط میان آنها و اصطلاحاتی که به این مفاهیم و روابط ارجاع می‌دهند.

ج- رمزگذاری و برچسب دهی هستی‌شناسی.

در واقع اصول حاکم بر این روش بسیاری از فعالیت‌های هستی‌شناسی را تحت تأثیر قرار می‌دهد. ابزار پشتیبان آن از طریق سرور آنتولینگو^{۲۰۳} است ولی در واقع این اصول عمدتاً بر طراحی بسیاری از ویرایشگران پیشرفته هستی‌شناسی‌ها تأثیرگذار است. (Uschold and King, 1995, Uschold et. al. 1998).

KACTUS - ۹-۲-۴

این رویکرد پایگاه‌های دانش موجود را برای ایجاد و استقرار آنتولوژیها تجهیز می‌نماید. چنین هستی‌شناسی براساس الگوهای دانش موجود ایجاد می‌شود و روشی از پائین به بالا دارد. (Bernaras et al. 1996)

METHONTOLOGY - ۹-۲-۵

روشی است برای طراحی و ساخت هستی‌شناسی‌ها با استفاده مجدد از آنتولوژیهای موجود و یا مهندسی مجدد آنها. چارچوب آن ساخت آنتولوژیها در سطح دانش یعنی سطح مفهومی تا ایجاد و

استقرار آن را مقدور می‌سازد. چارچوب کاری آن مبتنی بر شناسایی فرآیند ایجاد و گسترش هستی‌شناسی و فعالیت‌های اصلی (ارزیابی، ساختاربندی، مدیریت، مفهوم‌سازی، تلفیق و استقرار نظام و غیره) که مشخصات هر یک مراحل را برای انجام هر فعالیت فراهم می‌سازد. روشهای فنی که بکار گرفته می‌شوند، نتایج حاصله و چگونگی ارزیابی آنها را شامل می‌گردد. (Fernandez- Lopez et al. 1999)

SENSUS -۹-۲-۶

رویکردی است با روش بالا به پایین جهت استخراج آنتولوژیهای دامنه موضوعی خاص از هستی‌شناسی‌های کلان و گسترده بکار می‌رود. این روش بسیار خاص است و همه مباحث مهندس هستی‌شناسی را نمی‌پوشاند. (Swartout et al., 1997)

TOVE -۹-۲-۷

با هدف ارائه شیوه‌ای صوری برای ساخت و تدوین هستی‌شناسی‌ها براساس سوالات کارآمدی و کیفی است. این رویکرد سوالاتی را توصیف می‌کند که هستی‌شناسی قادر به جوابگویی آن است و ثابت شده که این شیوه کاربردی بویژه در مورد همکاری با متخصصین موضوعی که از دانش کمتری در الگو سازی درخور دارند بسیار کارآمد است. (Uschold & Grueninger 1996)

HOLSAPPLE -۹-۲-۸

با هدف ارائه شیوه‌ای مبنی بر تشریک مساعی در مهندسی آنتولوژی است. هدف از این کار حمایت از ابداع و خلق آنتولوژی است. مهندسی دانش هستی‌شناسی اولیه را طراحی و پس از آن براساس باز خورد از متخصصین موضوعی (domain experts) تغییرات و تحولات را ایجاد می‌نماید. این بازخورد با توزیع پرسشنامه و جمع‌آوری نتایج آن حاصل می‌گردد. مهندسی دانش پرسشنامه‌ها را بررسی و نیازهای جدید را در پرسشنامه جدید الحاق و مجدد توزیع می‌کند و این کار تا زمانی که همه مشارکت کنندگان با نتایج حاصله موافقت نمایند، ادامه می‌یابد.

(HOLS APPLE & JOSHI, 2002)

HCONE^{۲۰۴} -۹-۲-۹

رویکردی است که اخیراً برای توسعه هستی‌شناسی‌ها بکار گرفته شده است. این رویکرد از توسعه هستی‌شنای‌هایی بصورت غیر متمرکز پشتیبانی می‌کند. سه وجه متفاوت معرفی می‌شوند که در آنها آنتولژیها قابل ذخیره هستند. یکی فضای اطلاعات شخصی^{۲۰۵} که در آن فرد می‌تواند هستی‌شناسی‌هایی را ابداع و تلفیق کند، نسخه‌های مختلف را کنترل، اصطلاحات و معانی واژه‌ها را با

^{۲۰۴} - Human Centered ontology environment

^{۲۰۵} - Personal Space

مفاهیم مرتبط و هستی‌شناسی به روش بالا به پائین بکار گیرد. توسعه و گسترش هستی‌شناسی‌های شخصی^{۲۰۶} به اشتراک گذاشته می‌شود. این فضای اشتراک منابع برای تمامی مشارکت‌کنندگان دسترس پذیراست و در این فضا امکان بحث و بررسی تصمیمات مربوط به هستی‌شناسی برای تمامی کاربران امکان پذیر است. پس از مباحث و توافقات هستی‌شناسی مزبور به فضای مورد توافق^{۲۰۷} منتقل می‌گردد. (Kotis & Vouros, 2003; Kotis et al., 2004)

OTK-۹-۲-۱۰

این شیوه در پروژه آنتولوژی دانش اروپا^{۲۰۸} توسعه یافته است. این روش فرآیند مهندسی آنتولوژی را به پنج مرحله تقسیم می‌کند و هر مرحله زیر مرحله‌های متعددی دارد و تصمیم‌گیری اصلی و نهایی آن منجر به نتایج ویژه و خاصی می‌شود. این مراحل عبارتند از: مرحله مطالعاتی و امکان‌سنجی^{۲۰۹}، مرحله شروع^{۲۱۰}، مرحله اصلاح و بهبود^{۲۱۱}، مرحله ارزیابی^{۲۱۲} و مرحله اجرا و توسعه^{۲۱۳}. این مراحل یعنی «اصلاح»، «ارزیابی»، «اجرا» و «توسعه» بصورت دور و تسلسل تکرار می‌شوند. بایستی به این نکته توجه کرد که توسعه مراحل اجرایی نیازمند فرآیندهای دیگری مانند مهندسی و برنامه‌ریزی نرم‌افزاری و تربیت نیروی انسانی متخصص نیز می‌باشد. (Sure, 2003)

UPON^{۲۱۴}-۹-۲-۱۱

هرچند این روش هنوز کامل در پروژه‌های مختلف آزموده نشده است و ابزارهای پشتیبانی آن نیز در مراحل مقدماتی سیر می‌کند، از نظر مفهومی به خوبی پایه ریزی شده است، و براساس فرآیند نرم‌افزاری یکپارچه^{۲۱۵} و یو.ام.ال^{۲۱۶} با «زبان برنامه‌ریزی یکپارچه» پشتیبانی می‌شود. این شیوه نیز دارای مراحل مختلفی است که به صورت زنجیره‌وار انجام می‌پذیرد: ۱- شناسایی شرایط و لوازم، ۲- تجزیه و تحلیل منابع موجود و روند اجرایی، ۳- طراحی نظام مفهومی، ۴- پیاده‌سازی ۵- آزمون پوشش و کاربرد دامنه موضوعی که اجرای آن طی چهار مرحله عمده صورت می‌پذیرد: الف- مرحله آغازین،

^{۲۰۶} - Shared space

^{۲۰۷} - Agreed Space

^{۲۰۸} - Euproject On-To- Knowledge

^{۲۰۹} - Feasibility Study

^{۲۱۰} - Kick off

^{۲۱۱} - Refinement

^{۲۱۲} - Evaluation

^{۲۱۳} - Application and Evolution

^{۲۱۴} - The Unified Process for Ontology Building

^{۲۱۵} - Unified Software Development Process)

^{۲۱۶} - Unified Modeling Language (UML)

ب- توسعه و گسترش ج- ساخت، د- انتقال. این مراحل نیز بصورت دور و تسلسل انجام می‌پذیرد و پس از هر دور یک هستی‌شناسی کاربردی تولید می‌شود. (Nicola et al., 2005)

۱۰- شیوه طراحی و ساخت هستی‌شناسی^{۲۱۷}

در طراحی هستی‌شناسی علوم پایه و تبدیل پایگاه اطلاعات آن از واژه-بنیاد به مفهوم-بنیان پس از طی مرحله بررسی و مطالعه متون گردآوری شده از شیوه مت آنتولوژی استفاده شده، که ساخت هستی‌شناسی را بر اساس سطوح دانش میسر می‌سازد و ریشه در فعالیت‌های اصلی توسعه برنامه‌ریزی نرم‌افزاری دارد و بر اساس روشهای مهندسی دانش استوار است.

از نظر استراتژی طراحی هستی‌شناسی بر اساس استراتژی بالا به پایین ساخت آنتولوژی صورت می‌گیرد، که مراحل اصلی آن عبارتند از:

۱- ساخت واژه‌نامه مجموعه اصطلاحات^{۲۱۸}

۲- یکسان سازی مفاهیم و تعریف روابط میان مفاهیم

۳- ساخت تاکسونومی مفاهیم^{۲۱۹}

بنابراین طراحی آنتولوژی براساس تاکسونومی است^{۲۲۰}، زیرا با داشتن مجموعه متناهی از اصطلاحات آنتولوژیست‌ها تاکسونومی مفاهیم را شکل می‌دهند تا نظام سلسله مراتبی مفاهیم را تعیین نمایند.

۱۰-۱- شیوه ارزیابی هستی‌شناسی

همانند دیگر برنامه‌های نرم‌افزاری محتوای هستی‌شناسی‌ها نیز بایستی قبل از بکارگیری و مهندسی مجدد مورد ارزیابی قرار گیرد. هر چند هستی‌شناسی ارزیابی شده نیز خالی از اشکال نمی‌تواند باشد. اولین ارزیابی هستی‌شناسی در سال ۱۹۹۴ (گومز-پرز ۱۹۹۴ الف و ۱۹۹۴ ب) و در سالهای ۱۹۹۶ و ۲۰۰۱ گومز-پرز در چارچوب مت آنتولوژی و گوآرینو و همکاران (۲۰۰۱) با شیوه آنتوکلین^{۲۲۱} ارزیابی را انجام داده‌اند.

این ارزیابی بایستی از دو جنبه مدنظر قرار گیرد:

۱- ارزیابی فنی از دیدگاه تولیدکنندگان^{۲۲۲}

۲- ارزیابی از نظر کاربران آن^{۲۲۳}

^{۲۱۷} - Method for Building Ontology

^{۲۱۸} - Glossery of Terms

^{۲۱۹} - Concept Taxonomies

^{۲۲۰} -Taxonomic

^{۲۲۱} - Onto Clean

^{۲۲۲} - Logical-based Evaluation

ارزیابی فنی از محتوای هستی‌شناسی شامل :

۱- بازنگری^{۲۲۴}

۲- اعتباردهی^{۲۲۵}

الف- بازنگری و اعتباردهی هر یک از تعاریف مفاهیم

ب- مجموعه تعاریف مفاهیم و روابط آنها

ج- تعاریف مفاهیمی که از هستی‌شناسی‌های دیگر اقتباس شده‌اند

د- تعاریفی که منتج از دیگر مفاهیم و موضوعات شده و به آنها ارجاع داده می‌شود.

۲-۱۰- ارزیابی تاکسونومی^{۲۲۶}

۱- در ارزیابی تاکسونومی بایستی از لحاظ عدم تطابق (ناهمگونی) خطاهای دور باطل و اجزاء

و عناصر و نیز خطاهای معنایی مورد بررسی قرار گیرد.

۲- بررسی نواقصی که ممکن است در طبقه‌بندی مفاهیم و یا عناصر واجزاء و روابط نظام مفاهیم

وجود داشته باشد.

۳- بررسی موارد حشو در روابط دستوری و زیر طبقات و نیز یکسان بودن و همپوشانی تعاریف

برخی از طبقات مفاهیم و یا یکسان بودن تعاریف برخی نمونه‌ها ایجاد شده باشد.

۴- شیوه ارزیابی OntoClean بوسیله گروه طراحان هستی‌شناسی CNR در پادوای ایتالیا ابداع

شده است و دیدگاه این گروه بررسی هستی‌شناسی از نظر فلسفی است که از نظر عدم

انعطاف‌پذیری، ماهیت و همسانی و یکپارچگی مفاهیم مورد بررسی قرار می‌گیرند.

بر اساس مطالعات انجام‌شده ارزیابی کیفیت آنتولوژی OntoQA بر اساس استاندارد RDF/OWL و

استاندارد SKOS و براساس روابط منطقی مفاهیم logical-based و با تلفیق این دو روش مورد بررسی

قرار خواهد گرفت.

۱۱- مقولات هستی‌شناسی

در منطق، سور وجودی^{۲۲۷} نشانه‌ای است برای ارائه آن چیزی (شیئی یا پدیده) که وجود دارد. اما

منطق به تنهایی واژگان مناسب برای توصیف هستی‌شناسی را ندارد و آنتولوژی این خلاء را پر

^{۲۲۳} - Metric-based Evaluation

^{۲۲۴} - Verification

^{۲۲۵} - Validation

^{۲۲۶} - Taxonomy Evaluation

^{۲۲۷} - Existential Quantifier

می‌کند، هستی‌شناسی مطالعه هستی از هر دو نوع انتزاعی یا عینی است که جهان هستی را بواسطه گزاره‌ها و برجسب‌های زبانی نشانگر مفاهیم و نمودارهای مفهومی، بازنمایی می‌کند. منطق و هستی‌شناسی پیش نیاز بازنمایی دانش هستند و بدون هستی‌شناسی و منطق نمی‌توان در مورد هستی صحبت کرد و بدون آنها فقط در کلیات و ابهامات مباحث صورت می‌گیرد. اولین قدم در طراحی پایگاه اطلاعاتی دانش محور، گزینش مقولات هستی‌شناسی آن است. در سیستم‌های اطلاعاتی این مقولات به صورت «دامنه‌های موضوعی»^{۲۲۸}، در هوش مصنوعی به صورت «انواع»، در سیستم‌های موضوع - محور، «طبقات» و در منطق انواع یا «گونه‌ها» نامیده می‌شوند. به هر اسمی که نامیده شوند، گزینش مقولات بیان‌کننده پدیده‌ای است که می‌تواند در سیستم کامپیوتری بازنمایی شود.

فلاسفه معمولاً ساختار هستی‌شناسی‌های خود را از بالا به پایین ترسیم می‌کنند. اما برنامه ریزان کامپیوتری ترجیح می‌دهند که از پایین به بالا هستی‌شناسی را تدوین کنند. برای شکل‌گیری پایگاه داده‌ها و سیستم‌های هوش مصنوعی، با هستی‌شناسی‌های کوچکتر و محدود شروع می‌کنند و ابتدا با تعداد محدودی مفاهیم «دامنه موضوعی» خاص آغاز می‌نمایند. ترسیم هرم مفهومی در رباتیک^{۲۲۹}، برنامه‌ریزی رایانه‌ای و یادگیری الکترونیک بسیار متداول است.

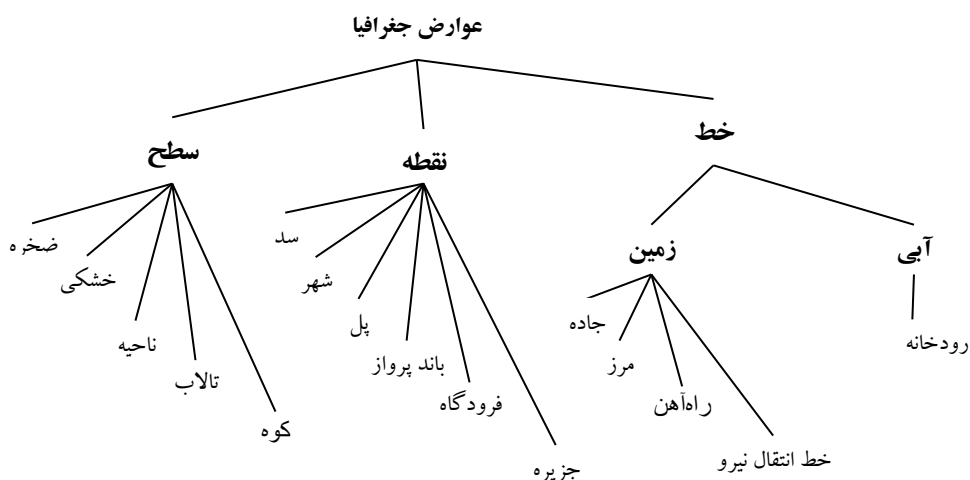
به عنوان مثال دیوید وارن^{۲۳۰} و فرناندو پیرا^{۲۳۱} (۱۹۸۲) برای سیستم جستجو و پاسخگویی به پرسشهای جغرافیایی (Chat-80) رده بندی مطرح در این زمینه را طراحی نمودند (Sowa, 2008).

^{۲۲۸} - Domains

^{۲۲۹} - Robotics

^{۲۳۰} - David Warren

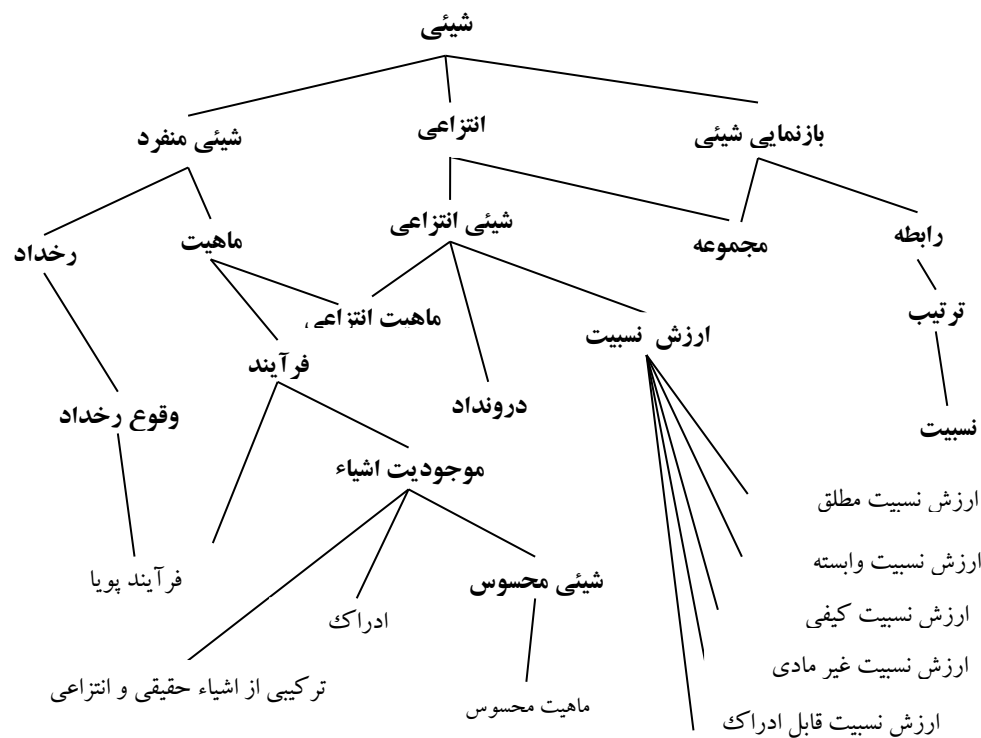
^{۲۳۱} - Fernando Pereira



نمودار ۶. رده‌بندی جغرافیایی سیستم چات-۸۰ (Sowa, 2000)

این تقسیم‌بندی مقوله‌ها سبب می‌شود که در جستجوی اطلاعات و پاسخ به پرسش‌های تحلیل گران و نیز ماشین جستجو، اطلاعات دقیق‌تری به برنامه‌ریزان و متخصصین ارائه گردد. مثلاً برای برنامه ریزان شهری ویژگی‌های یک منطقه، برای خلبانان ویژگی‌های دقیق خطوط هوایی و مختصات محل فرود هواپیما، برای کشتیرانان مشخصات رودخانه در سه بعد طول، عرض و عمق را نشان می‌دهد زیرا عمق رودخانه‌ها و فاصله آنها تا ساحل بسیار حائز اهمیت است. این اطلاعات دقیق برای برنامه‌ریزان کشور و نیز تجارت و حمل و نقل از اهمیت خاص برخوردار است.

مثال دیگر دسته بندی مقولات در سیستم CYC است (Lenat 8 Guha 1990) که بر خلاف دیدگاه جزئی نگر نمونه قبلی، دیدگاهی کلان نگرداشته و تقریباً تمامی دانش بشری را در بر می‌گیرد.



نمودار ۷. کاربرد رده‌بندی سطوح بالا در سیستم Cyc (Sowa, 2000)

نمودار فوق سلسله مراتب مقولات و دسته بندی کلان این سیستم بازیابی اطلاعات را نشان می‌دهد. این سلسله مراتب مقوله‌ای پیش نیاز طراحی پایگاههای داده – بنیان کلان نگر و دانش محور است. در واقع این تقسیم‌بندی کلان مقولات، در گذشته نیز از سوی فلاسفه به طریقی انجام شده بود. در تقسیم‌بندی مقوله‌ها توسط ارسطو که از سوی برنتانو^{۲۳۲} (۱۸۶۲) به صورت نمودار درختی درآمده است و هستی را به ماهیت و رخداد تقسیم کرده، زیر مقوله‌های آن مانند فرآیند، علیت، موقعیت، شرایط و غیره را نیز نشان داده است. (Sowa, 2000)

پیرس^{۲۳۳} این مقولات را به سه بخش عمده تقسیم‌بندی نموده است

اطلاعات ساختاری ناب و محض شامل خلاصه‌ای از توصیف اشیاء و پدیده‌ها و رسانه‌ای که آنرا ثبت و ضبط و یا انتقال می‌دهد.

کیفیت این اطلاعات که کاملاً وابسته به پیچیدگی ساختار آن است و ظاهراً به معنی شناسی آن وابسته نیست. کلود شانون^{۲۳۴} (۱۹۴۸) در «نظریه اطلاعات»^{۲۳۵} خود، بزرگترین اشتباه یعنی نادیده گرفتن

^{۲۳۲} - Brentano

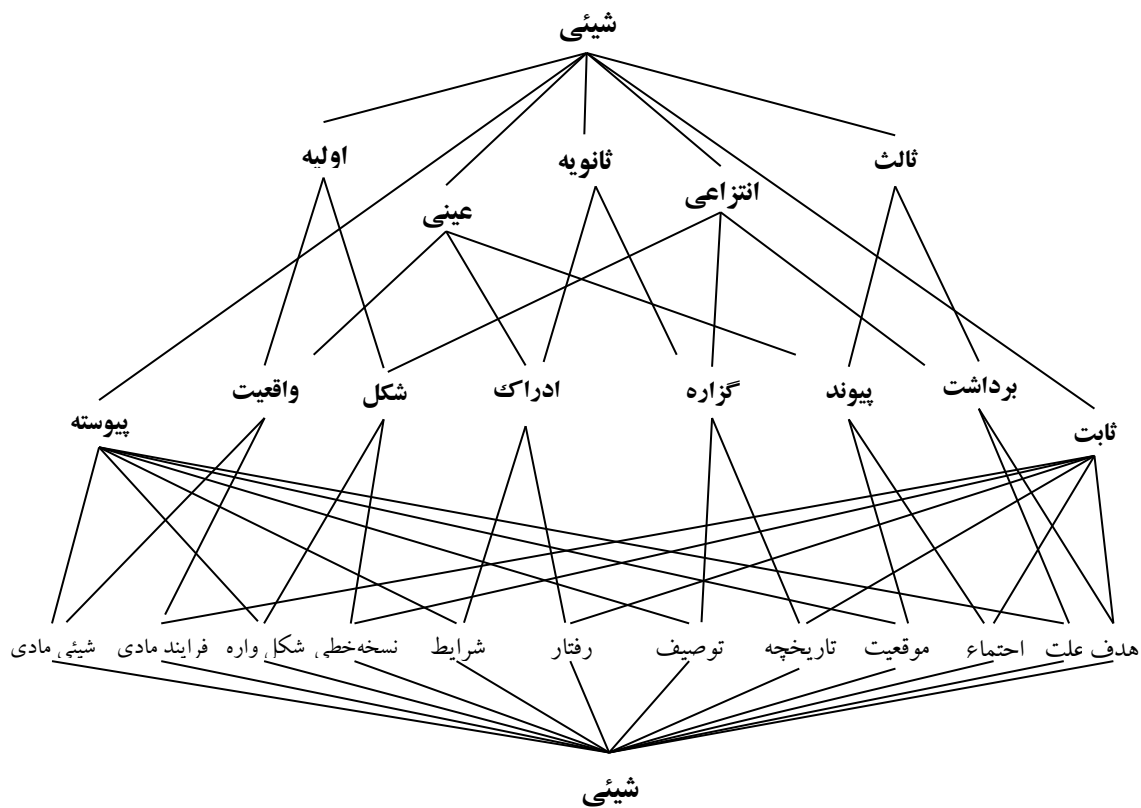
^{۲۳۳} - Peirce

^{۲۳۴} - Claude Shanon

^{۲۳۵} - Information Theory

«معنا» در این فرآیند را مرتکب شد.

وی معتقد بود که: «مشکل اساسی در ایجاد ارتباطات باز تولید دقیق یک پیام و رساندن آن از نقطه‌ای به نقطه‌ای دیگر است. هرچند این پیام‌ها دارای «معنا» هستند؛ که آنها را ارجاع به نظام‌های مفهومی می‌دهد اما این ابعاد معنی‌شناختی ارتباطات، خارج از موضوع و مسائل مهندسی آنهاست.» اما ابعاد معناشناختی که «شانون» در تحلیل مسائل و مشکلات مهندسی اطلاعات کاملاً نادیده گرفته بود، از دیدگاه فلسفی بسیار اساسی و زیر بنایی است. تمرکز بر آنتولوژی، همبستگی اطلاعات با مدخل‌های مفهومی و سیستم مفاهیم را نشان می‌دهد. (نمودار ۸) ترکیبی از نظریات گسترش یافته از «هراکلیتوس»^{۲۳۶} تا «پیرس» و «وایتد»^{۲۳۷} است.



نمودار ۸. شبکه مفاهیم براساس تمایزات سه گانه (Sowa, 2000)

بنیادی‌تر از خود مقولات ذکر شده، معیارها و ضوابط ایجاد تمایز میان این مقولات است و تعریف این که آیا به مدخل خاصی تعلق دارند یا خیر. مقولات هشت گانه «وایتد» منبعی است برای به دست آوردن این معیارها: هرگاه تضادی میان نمونه‌ها و مصداق‌های یک مقوله مشاهده شود، می‌تواند به

^{۲۳۶} - Heraclitus

^{۲۳۷} - Whitehead

عنوان یک تمایز تعمیم یافته و آن مقوله را به دو یا چند زیر مقوله تقسیم نماید. برای وایتد، مقولات و روابط یک هستی‌شناسی (آنتولوژی) مدخل‌های انتزاعی محسوب می‌شوند که نشأت گرفته از تضادهای تعمیم یافته هستند. تقابل‌های تعریف شده از سوی «وایتد» که در مدخل‌های فیزیکی مشاهده شده‌اند، داده‌های آغازین قاعده‌های اصلی نشأت گرفته از نظریات لایبنس^{۲۳۸} است. هر دو شیوه مشاهده تقابل‌های وایتد و نیز شیوه استخراج قواعد اصلی لایبنس می‌توانند تلفیق یافته و سلسله مراتب مقولات را تولید نمایند. هر چند «وایتد» مشاهدات خود را با روش پائین به بالا و لایبتین با روش بالا به پائین سلسله مراتب مقولات را ترسیم نموده است. اما روش تقابل‌ها از انعطاف پذیری بیشتری برخوردار است و هر تقابلی از مشاهدات حاصل شده و هر مقوله می‌تواند زیر شاخه‌هایی براساس تقابل‌های جدید کشف شده داشته باشد. «وایتد» در این موضوع که هر شیئی در حال تغییر مداوم است با هراکیتوس موافق بود، اما معتقد بود که در برخی اشیاء این تغییر شتاب بیشتری داشته و در برخی دیگر به نسبت ثبات بیشتری وجود دارد. برای سازگاری اشیاء در یک هستی‌شناسی فرآیند محور وی میان «اشیاء پایدار» که در دوره‌های زمانی از ثبات بیشتری برخوردارند، و «واقعیت‌ها» که مدام در حال تغییر هستند و فقط در فواصل زمانی منظم بازشناسی می‌شوند، تمایز قائل گردید و اولی را «پیوستار» و دومی را «رخداد» نامید.

اجرای یک برنامه موسیقی و سمفونی «ثابت» است زیرا حداکثر یک ساعت طول می‌کشد، اما ضبط کردن و اجراهای مجدد «پیوسته» و قابل ضبط اطلاعات آن بر روی نوار مغناطیسی می‌باشد که سالها می‌توان آنرا به طور مکرر به اجرا درآورد. تمایز میان «پیوسته» و «ثابت» این است که می‌توان سوال کرد: «این اطلاعات در مورد چه چیزی است؟» و این که «بر روی چه رسانه‌ای این اطلاعات ذخیره شده است؟» اما در اصطلاح شناسی «وایتد»، اطلاعات به خودی خود شیئی یا پدیده‌ای است فنا ناپذیر و دائمی که می‌تواند بطور نامحدود نگهداری شود و نسخه‌هایی از آن بطور مکرر و از زمانی به زمانی دیگری تهیه و نگهداری شود. هرچند ساختار اطلاعات تغییر نمی‌یابد، اما می‌تواند ساختار «پیوسته» و یا «ثابت» را رمز گذاری کند:

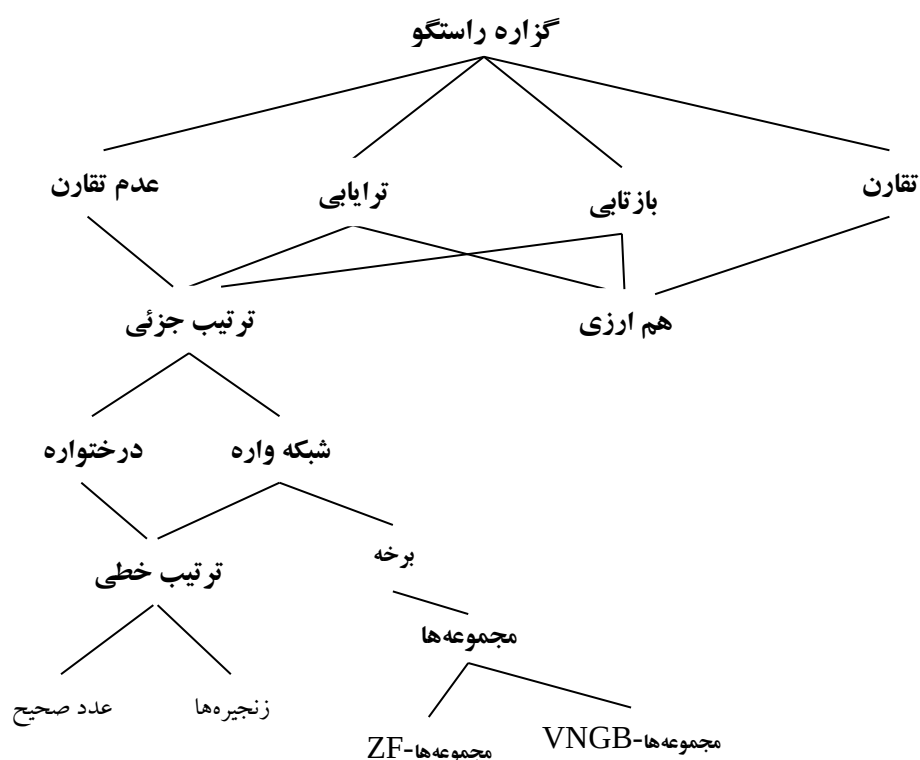
الگوی ایستا بر روی نوار مغناطیسی می‌تواند فرآیندی پویا مانند یک فیلم یا اجرای یک سمفونی را ثبت و ضبط کند. با تمایز دو گانه‌ای میان «پیوسته» و «ثابت» و شش مقوله به صورت دوازده زیر مقوله ارائه می‌شوند. نمودارهای سلسله مراتب ترسیم شده براساس شیوه لایبنس و تلفیق آنها «شبکه‌ها»^{۲۳۹} نامیده می‌شوند (Sowa, 2000).

^{۲۳۸} - Leibniz

^{۲۳۹} - Lattices

نمونه‌ای دیگر ساختارهای ریاضی هستند که می‌توان به عنوان «صریح»^{۲۴۰} و یا «غیر صریح و تلویحی»^{۲۴۱} تعریف نمود. در یک تعریف صریح، هر یک ساختار جدید کوتاه نوشته‌ای^{۲۴۲} است برای ترکیبی که قبلاً به عنوان یک ساختار کامل تعریف شده است. به عنوان مثال برای هر ساختاری می‌توان نشانه گذاری کرده و در متون آن علامت یا نشانه را بجای کل ساختار بکاربرد مانند «فرمولهای جبر» که چنین علائمی را بکار می‌گیرند.

در تعاریف غیر صریح و تلویحی هنوز اصطلاح یا علائمی برای بکارگیری ساختارها و تعاریف نیامده است، مثلاً نظریه‌های جدید که بایستی با اصطلاحات جدید کاملاً توضیح داده شوند، زیرا اصول حاکم بر ماهیت آنها هنوز روشن نمی‌باشد. در نمودار ذیل سلسله مراتبی از نظریه‌های ریاضی را نشان می‌دهد. هر نظریه تعمیم یافته زیر مجموعه خود و از طرفی دیگر اخص از نظریه‌های فوقانی است. نظریه‌های صدق (راستگو)^{۲۴۳} شامل تمامی گزاره‌های منطقی صادق که می‌تواند مورد آزمون قرار گیرد.



نمودار ۹. سلسله مراتب نظریه‌ها

^{۲۴۰} - Explicitly

^{۲۴۱} - Implicitly

^{۲۴۲} - Abbreviation

^{۲۴۳} - Tautologies

هر یک از نظریه‌های صدق منشعب از رده‌بالا تر از خود با اضافه نمودن اصول موضوعی دیگر است، که قضیه‌ها شامل همه قضیه‌هایی است که وارث ویژگی‌های قضایای فوقانی به علاوه موضوعات جدید می‌باشد که آنها نیز برهان پذیر هستند. با افزودن اصل موضوعات بیشتر، یک نظریه گسترش یافته، بدین معنا که در برگیرنده گزاره‌های بیشتری خواهد بود. این اصل زیر بنایی که برای اولین بار توسط ارسطو مشاهده گردید به رابطه وارون^{۲۴۴} میان مفهوم^{۲۴۵} و مصداق^{۲۴۶}: اشاره می‌کند در حالی که معنا و مفهوم بمعنای افزایش تعداد موضوعات و یا شرایط و حدود تعریف گسترش می‌یابد، مصداق‌ها از نظر تعداد نمونه‌ها و شواهد کوچکتر می‌شود. (Sowa: 2000)

زیر مجموعه گزاره‌های منطقی صادق، چهار نظریه به نام «عدم تقارن»^{۲۴۷}، «تعدی و ترایی»^{۲۴۸}، «بازتابی»^{۲۴۹} و «تقارن»^{۲۵۰} قرار می‌گیرند. هر کدام از آنها گزاره‌های صادق را در برمی‌گیرند و علاوه بر آن هر کدام اصل موضوعی خاص را دارد که رابطه دوتایی ایجاد می‌کند. و بهمین ترتیب تقسیم‌بندی‌هایی براساس اصل موضوع انجام می‌گیرد و انشعاب‌هایی ایجاد می‌گردد که تمامی آنها زیر مجموعه نظریه گزاره‌های منطقی صادق محسوب می‌شوند. نمودار ۱۳ تنها بخش کوچکی از شبکه نامتناهی نظریه‌های ممکن هستند و نشان دهنده تعدادی از نظریه‌هایی هستند که ریاضیدانان علاقه‌مند به مطالعه و بررسی آنها هستند، اما شبکه نامحدود به اندازه کافی نظریه‌های مختلف را در برمی‌گیرد که بتوان بر اساس آنها برنامه‌ریزی‌های کامپیوتری که تابه حال توسط انسان، روبات و سیستم‌های هوش مصنوعی انجام یافته را در برگیرد.

براساس این نظریات متخصصین موضوعی همکار طرح «هستی‌شناسی» به مطالعه و تحقیق در زمینه تقسیم‌بندی شبکه مفاهیم هر رشته تخصصی در حوزه علوم پایه پرداختند و براساس نظریات و یافته‌های محققین این رشته به طراحی هستی‌شناسی‌ها پرداختند. هر یک از رشته‌های علوم پایه (شیمی، فیزیک، زیست‌شناسی، زمین‌شناسی و ریاضی) با بهره‌گیری از نظرات متخصصین رشته تخصصی خود شبکه‌ای از مفاهیم را براساس تقسیم‌بندی موضوعی و ایجاد رابطه منطقی میان مفاهیم انجام داده و به صورت نمودارهای خورشیدی ترسیم نمودند.

^{۲۴۴} - Inverse Relationship

^{۲۴۵} - Intension

^{۲۴۶} - Extention

^{۲۴۷} - Antisymmetry

^{۲۴۸} - Transitivity

^{۲۴۹} - Reflexivity

^{۲۵۰} - Symmetry

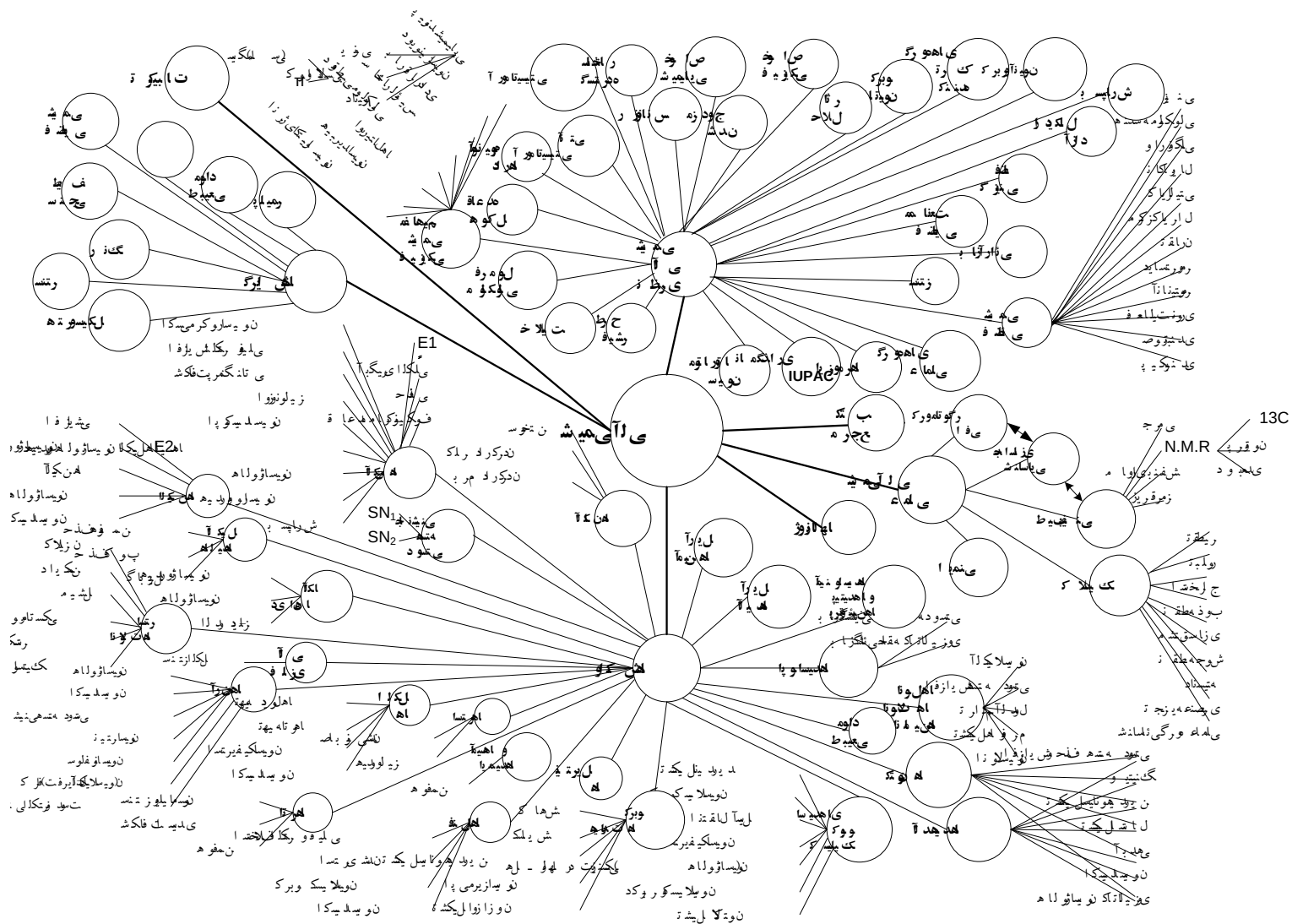
۱۲ - تدوین ساختار درختی

معمولاً هستی‌شناسی‌های واژگانی به یکی از دو روش قیاسی و یا استقرایی تدوین می‌گردند. در روش قیاسی ابتدا چارچوب کلی زمینه موضوعی ترسیم میگردد، یعنی مشخص می‌شود که محدوده آن چیست و چه موضوعهای کلی را در برمیگیرد. سپس هر یک از این موضوعها به موضوعهای فرعی تر تقسیم می‌گردد و این کار تا مرحله تعیین اخص‌ترین مفاهیم ادامه پیدا می‌کند. در این روش، مفاهیم اصطلاحات معمولاً از منابع متعددی نظیر طرحهای رده‌بندی، اصطلاحنامه‌های موجود، دایره‌المعارفها، فرهنگها و لغت‌نامه‌های تخصصی، نمایه‌نامه‌های ادواری، فهرست‌مندرج و متون کتابهای درسی پایه و غیره استخراج می‌شوند. این واژه‌ها در کمیته‌هایی متشکل از متخصصان موضوعی بررسی و ارزشیابی می‌شوند تا آنهایی که مناسب با نیاز زمینه موضوعی هستند انتخاب و پیشنهاد شوند در روش استقرایی ابتدا مجموعه نمونه‌ای از انواع مدارکی که در عمل زیر پوشش نظام ذخیره و بازیابی مورد نظر قرار خواهند گرفت انتخاب می‌شود، سپس از نمایه‌سازی این مدارک، مجموعه‌ای از مفاهیم و اصطلاحات فراهم می‌آید.

در هر دو روش مشارکت عملی متخصصان موضوعی در گزینش اصطلاحات رکن اصلی کار است. در واقع روش استقرایی شکل دهی سلسله‌مراتبی به یک توده از واژه‌های یک حوزه است و در روش قیاسی تدوین یک ساختار مفهومی پایه در یک یا چند حوزه تخصصی و بعد تزریق واژه‌های مرتبط در هر یک از شاخه‌ها و زیر شاخه‌های مربوط به این ساختار پایه و بسط و توسعه آن است. (ساختار زیست‌شناسی، نمودار ۱۰)

به عنوان مثال برای تدوین هستی‌شناسی فارسی شیمی (علوم پایه) همکاران ما روش قیاسی را برگزیدند. با استفاده از کتب معتبر درسی در سطح دوره های کارشناسی و کارشناسی ارشد شیمی و نیز با استفاده از برخی از هستی‌شناسی های الکترونیکی موجود بر روی پایگاه های اینترنتی، ابتدا با رسم یک نمودار به تقسیم بندی رشته شیمی به شاخه های اصلی آن پرداختیم سپس به ترسیم ساختار درختی در هر یک از شاخه های اصلی همت گماشتیم. برای نمونه ساختارهای درختی پایه برای گرایش های اصلی شیمی فیزیک آورده می شود.





نمودار ۱۴. بخشی از «اصطلاح‌نامه شیمی» تدوین: تقی رجبی

برای بسط و توسعه و روز آمد سازی این ساختار پایه و نزدیک کردن آن به یک ساختار هستی شناسی واژگانی قابل قبول، باید به این ساختار به صورتی روشمند واژه تزریق نمود. واژه های تزریقی را به دو طریق عمده جمع آوری کرده ایم. یک بخش از واژه های مورد استفاده از مجموعه واژگانی که در بانک اطلاعات جامع مرکز اطلاعات و مدارک علمی ایران موجودند و بخش دیگر آن واژه هایی هستند که در اصطلاحنامه های شیمی موجود بر روی اینترنت، یا مجموعه واژگان ها و واژه نامه های مرکز نشر دانشگاهی و کتب معتبر درسی یافت می شوند.

۱۳- همسانسازی اصطلاحنامه

برای استفاده از چند اصطلاحنامه به صورت یک ماکرو تزاروس یا آنتولوژی، در اولین قدم باید کلیه مفاهیم مشترک در اصطلاحنامه ها همسان شوند. در یک مجموعه اصطلاحنامه با دهها هزار واژه که در زمانهای مختلف و توسط متخصصین موضوعی مختلف تولید شده اند، همسانسازی بدون استفاده از تواناییهایی که کامپیوتر در اختیارمان میگذارد تقریباً ناشدنی خواهد بود. حتی استفاده از ابزارهایی عمومی مانند «اکسل»^{۲۵۱} یا «اکسس»^{۲۵۲} یا... که میتوانند تا حدی پروسه همسانسازی را آسان نمایند، نیاز به زمان و نیروی بسیار زیادی از طرف نیروی انسانی متخصص خواهند داشت تا بتوانند فرآیند همسانسازی را تا حدی به نتیجه مطلوب نزدیک سازند.

همسانساز اصطلاحنامه^{۲۵۳} ابزاری است برای نشان دادن سریع موارد اختلاف بین مفاهیم مشترک بین دو اصطلاحنامه که با استفاده از نرم افزار تزاروس بیلدر^{۲۵۴} ساخته شده اند. این ابزار توانمند ضمن ارائه موارد اختلاف، برای هر یک مورد راهکارهای موجود را به متخصص موضوعی پیشنهاد میدهد. در بسیاری از موارد راهکار انتخاب شده توسط متخصص موضوعی میتواند مستقیماً توسط همسانساز به اجرا درآید تا مفهوم مورد نظر بین دو اصطلاحنامه کاملاً یکسان شوند، یا کاملاً از یکدیگر متمایز گردند.

موارد اختلاف بین اصطلاحنامه ها بر اساس استاندارد ایزو ۲۵۹۶۴ بررسی می شوند. اگر بیش از دو اصطلاحنامه نیاز به همسانسازی داشته باشند با همسانسازی همه اصطلاحنامه ها بصورت دو به دو، یک مجموعه اصطلاحنامه همسان خواهیم داشت. همسانساز اصطلاحنامه موارد زیر بین دو اصطلاحنامه را مورد بررسی قرار می دهد:

^{۲۵۱} . Excel

^{۲۵۲} . Access

^{۲۵۳} . Thesaurus Synchronizer

^{۲۵۴} . Thesaurus Builder

- ۱- اختلاف نوشتاری بین یک مفهوم در دو اصطلاحنامه
- ۲- اختلاف در واژه‌های اخص برای یک مفهوم در دو اصطلاحنامه
- ۳- اختلاف در واژه‌های غیر مرجح برای یک مفهوم در دو اصطلاحنامه
- ۴- اختلاف بین ترجمه‌های یک مفهوم در یک زبان خاص در دو اصطلاحنامه
- ۵- اختلاف در واژه‌های وابسته برای یک مفهوم در دو اصطلاحنامه
- ۶- نبود یک واژه وابسته که معادل مفهوم مورد نظر در اصطلاحنامه دیگر آمده است
- ۷- استفاده از ترجمه یکسان برای دو مفهوم مختلف در یک زبان خاص
- ۸- اختلاف در انتخاب یک واژه مرجح برای یک مفهوم
- ۹- اختلاف در نوع رابطه بین دو مفهوم
- ۱۰- ایجاد حلقه نامحدود میان مفاهیم (شبکه مفاهیم)
- ۱۱- عدم یگانگی یک مفهوم در برابر یک اصطلاح (چند معنایی).

۱۴- روش استفاده از نرم‌افزار همسانساز

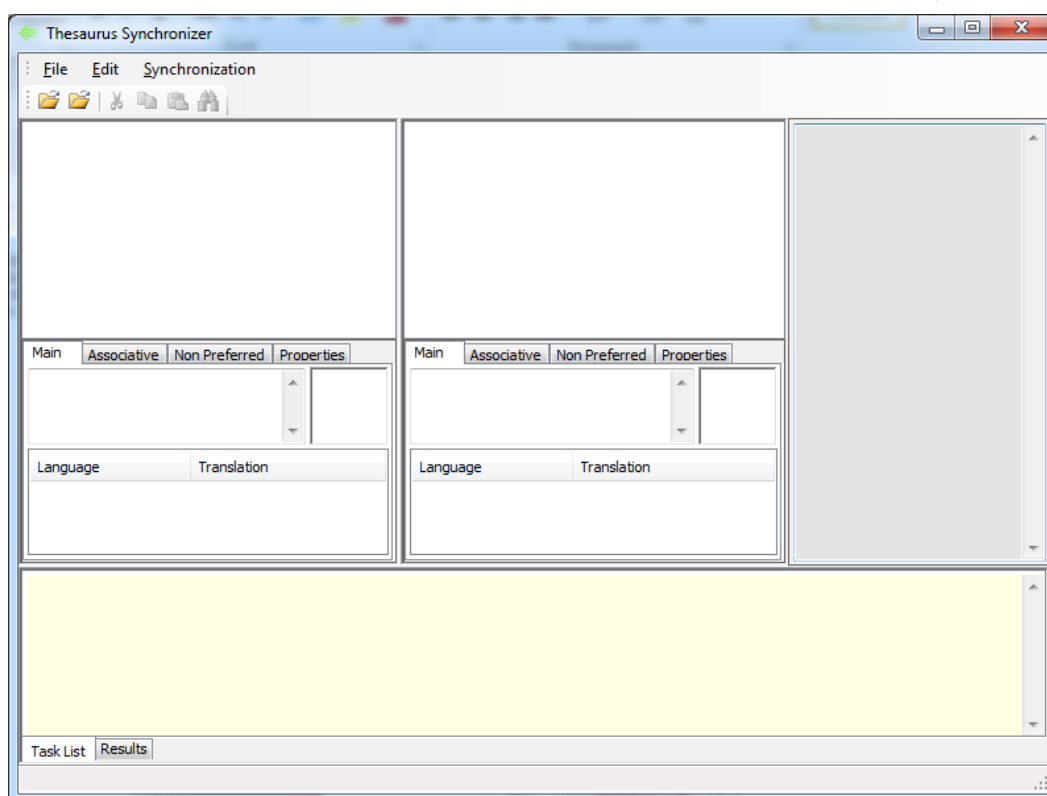
فرآیند تبدیل اصطلاحنامه به هستی‌شناسی مستلزم انجام مراحل خاصی توسط متخصصین موضوعی برای انطباق مفاهیم با ساختار هستی‌شناسی است. یکی از این مراحل تبدیل اصطلاحنامه متن‌بنیان به اصطلاحنامه مفهوم‌بنیان است، که خود دارای پیچیدگی‌های وسیعی است. مهمترین عاملی که باعث می‌شود ساختار یک هستی‌شناسی با مشکل مواجه گردد، وجود مفاهیم یکسان با روابط ناهمگون و یا وجود مفاهیم متفاوت با زیر مجموعه‌های یکسان است. یکی از وظایف متخصصین موضوعی یافتن اینگونه ایرادات در ساختار اصطلاحنامه یا هستی‌شناسی است. از آنجایی که متخصص موضوعی عموماً با استفاده از روشهای تجربی و ذهنی (ضمن داشتن احاطه علمی) عهده‌دار انجام این مرحله است، معمولاً درصدی از اینگونه اشکالات در زوایای پنهان اصطلاحنامه باقی میماند. حال اگر بیش از یک متخصص موضوعی روی بخشهای مختلف یک اصطلاحنامه کار کرده باشند، تفاوت ذهنیت هر یک از آنها در مورد مفاهیم مزید بر علت خواهد شد، تا انطباق مفاهیم در اصطلاحنامه بطور صحیح انجام نشود.

بطور خاص، می‌توان شرایط بسیار ویژه‌ای برای اصطلاحنامه‌های پژوهشگاه علوم و فناوری اطلاعات به عنوان تنها سازمانی که در سطح دنیا اقدام به تولید مجموعه‌ای از چندین اصطلاحنامه در حوزه علوم پایه کرده است، در نظر گرفت. اصطلاحنامه‌های این پژوهشگاه هر یک بطور جداگانه از نظر ساختار درونی کاملاً منطبق با استانداردهای جهانی ساخت و توسعه اصطلاحنامه هستند و ایرادی از لحاظ روابط متن‌بنیان به آنها وارد نیست. با این وجود وقتی صحبت از انتقال آنها به ساختارهای مفهوم‌بنیان

میشود، داستان کاملاً متفاوتی در پیش رو خواهیم داشت. چند عامل باعث پیچیدگی و ناهمگونی بسیار زیادی در مفاهیم موجود در اصطلاحنامه‌های پژوهشگاه شده است که از آن جمله میتوان به تعدد اصطلاحنامه‌ها، تعدد متخصصین موضوعی، فعالیت غیر متمرکز، و گستردگی موضوعات علوم پایه اشاره کرد.

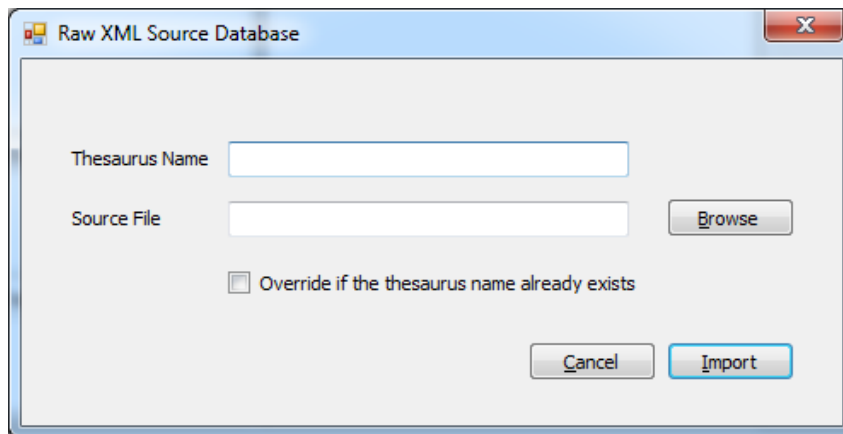
با توجه به عوامل یاد شده، و اینکه موارد ناهمگونی بین اصطلاحنامه‌ها تا چندین هزار مورد پیش‌بینی می‌شود، به نظر نمی‌رسد که بتوان بیش از درصد اندکی از این اشکالات را حتی با صرف هزینه و زمان زیاد، و با استفاده از روشهای متداول، برطرف کرد. به همین دلیل، طرحی برای تولید نرم‌افزاری که بتواند بخش مهمی از ناهمگونی‌های بین اصطلاحنامه‌های پژوهشگاه را به متخصص موضوعی بنمایاند مطرح گردید. حاصل این طرح، به صورت یک نرم‌افزار هوشمند تحت عنوان Thesaurus Synchronizer ارائه شده است که در ادامه روش استفاده از آن بطور مختصر بیان میگردد.

در این نرم‌افزار تا حد امکان سعی شده است تا با حذف زوائد و انجام عملیاتها در پس زمینه، حداکثر سادگی هم از نظر روش اجرا و هم از نظر ظاهری برای کاربر فراهم گردد. با اجرای نرم‌افزار، صفحه اصلی به شکل زیر باز خواهد شد:



Thesaurus Synchronizer با استفاده از فرمولهای پیچیده سعی در تشخیص روابط ناهمگون بین اصطلاحنامه‌ها دارد. این برنامه اصطلاحنامه‌ها را دو به دو بارگذاری کرده و با نمایش ناهمگونی‌ها

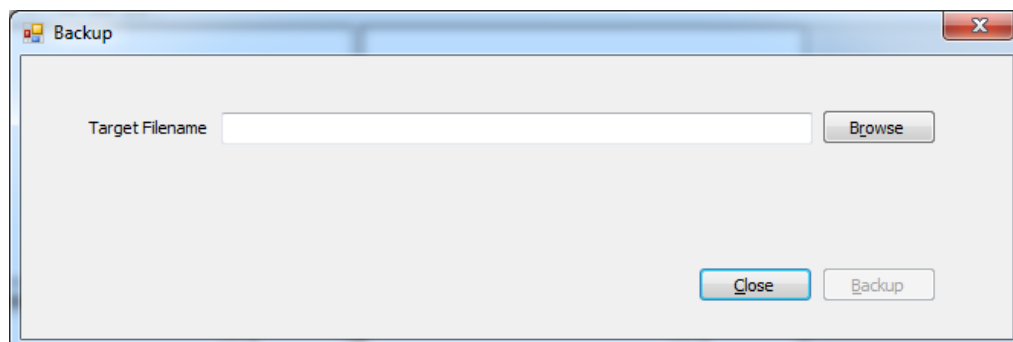
منتظر تصمیم متخصص موضوعی برای راه حل آنها می‌شود. برای اینکه اصطلاحنامه‌ها قابل بارگذاری باشند، ابتدا باید آنها را از نرم‌افزار Thesaurus Builder به SQL منتقل نمود. برای این کار، فایل‌های XML مربوط به اصطلاحنامه‌های پژوهشگاه ارائه شده‌اند. این فایل‌ها با استفاده از تزاروس بیلدر نسخه ۶,۹,۵,۵ و یا بالاتر قابل تولید است. از منوی File گزینه Import Raw Data را انتخاب نموده تا پنجره زیر نمایان گردد:



ابتدا نام اصطلاحنامه را انتخاب، سپس مسیر و نام فایل مربوط به اصطلاحنامه را در قسمت Source File وارد کرده، کلید Import را زده و تا پایان کار ادامه می‌یابد. عمل فوق را برای دیگر اصطلاحنامه‌ها نیز انجام می‌دهیم. باید توجه داشته باشیم که بعد از یکبار انجام این مرحله دیگر نیازی به تکرار آن نیست.

۱-۱۴- نسخه پشتیبان

با توجه به پیچیدگی عملیات مربوط به یکسان‌سازی اصطلاحنامه‌ها، برای جبران اشتباهات احتمالی، حتماً سعی شود در فواصل زمانی معین (مثلاً روزانه) از کل اطلاعات نسخه پشتیبان تهیه کنیم. برای این کار از منوی File گزینه Backup Database را انتخاب نموده تا پنجره زیر باز شود:

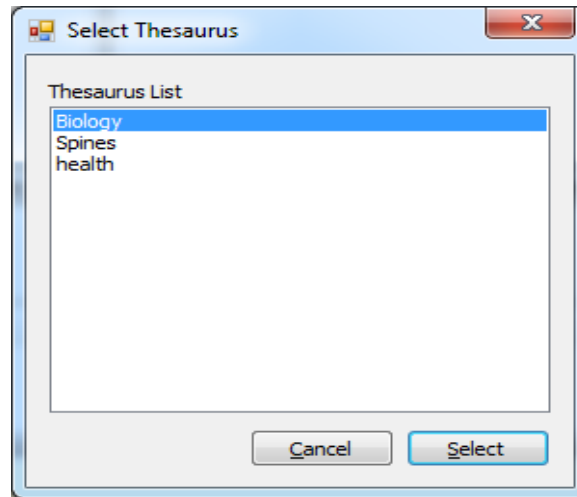


نام و مسیر فایل پشتیبان را وارد کرده و کلید Backup را می‌زنیم. توضیح اینکه که برای این عمل، لازم است SQL Server دسترسی نوشتن به مسیر یاد شده را داشته باشد. از مدیر شبکه درخواست

نموده تا این دسترسی را برقرار گردد.

۱۴-۲- بارگذاری

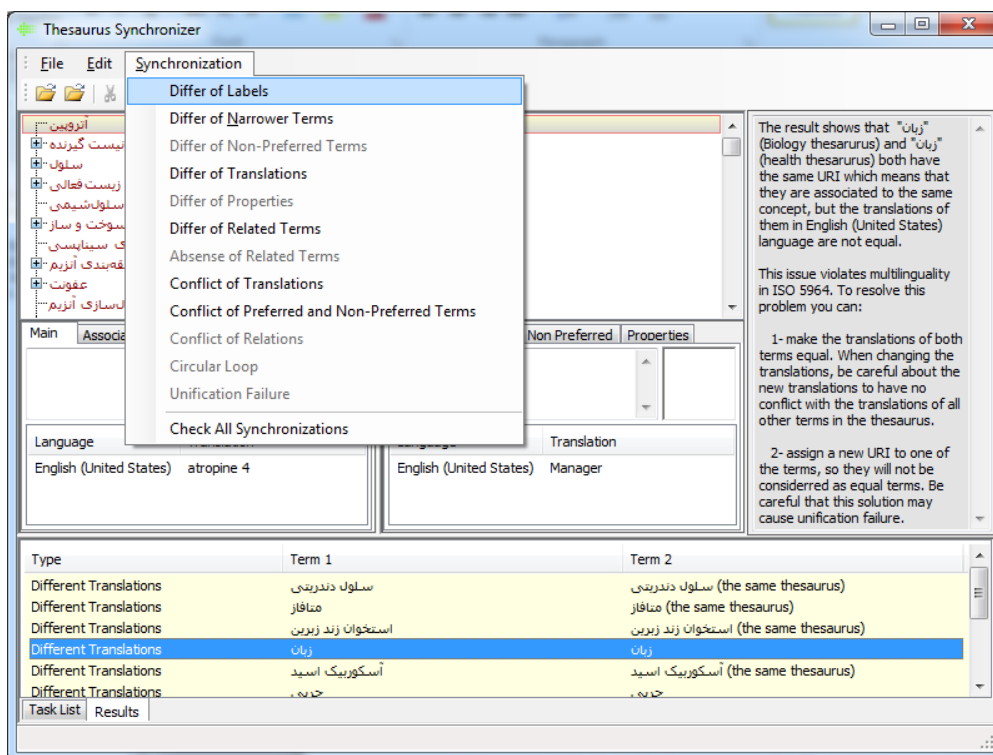
قبل از یکسان سازی لازم است دو اصطلاحنامه بارگذاری گردند. از منوی File گزینه Open Thesaurus 1 را انتخاب نموده تا پنجره زیر باز شود:



اصطلاحنامه مورد نظر را انتخاب نموده و کلید Select را فشار می دهیم تا بارگذاری آغاز گردد. همین عمل را برای اصطلاحنامه دوم تکرار می نماییم.

۱۴-۳- آشکارسازی ناهمگونی ها

با استفاده از آیتم های موجود در منوی Synchronization می توانیم موارد ناهمگونی را مشخص نماییم. فهرست ناهمگونی ها در پایین صفحه اصلی در قسمت Result قابل مشاهده خواهد بود:



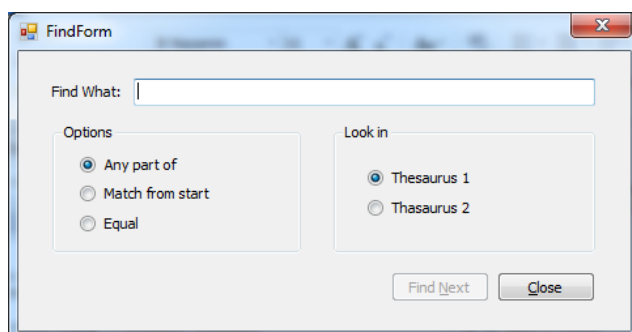
با انتخاب هر یک از نتایج، توضیح مختصری در مورد ناهمگونی و راه کارهای آن در قسمت بالا سمت چپ پنجره اصلی قابل مشاهده است. با دوبار کلیک کردن روی هر جواب، واژه‌های مورد نظر در هر دو اصطلاحنامه انتخاب خواهند شد.

۴-۱۴- ویرایش

متناسب با هر یک از نتایج بدست آمده، متخصص موضوعی راه کارهای مخصوص آن نتیجه را در اختیار دارد که در منوی Edit فهرست شده است. شرح مختصر هر یک در ادامه آمده است:

Find برای جستجوی یک واژه بکار می‌رود. در پنجره جستجو، شکل جستجو و

اصطلاحنامه مورد جستجو باید مشخص باشد:



از این دستورات برای کپی یا انتقال یک واژه به محلی دیگر از یکی از دو اصطلاحنامه استفاده می‌شود. باید توجه کرد که این عمل بین دو اصطلاحنامه

Paste، Copy، Cut

ممکن است باعث ایجاد ناهمگونی‌های جدید گردد.

Edit Label

برای ویرایش شکل نوشتاری یک مفهوم از این دستور استفاده می‌شود. ویرایش شکل نوشتاری یک مفهوم باعث می‌شود که شکل نوشتاری تمام مواردی که این مفهوم در اصطلاحنامه‌ها تکرار شده‌اند نیز تغییر نماید.

Delete Term

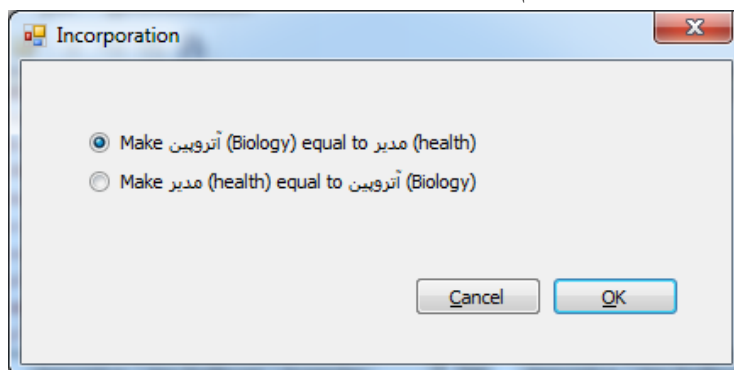
در صورتی که متخصص موضوعی وجود یک مفهوم نامرتبط یا نادرست را تشخیص دهد، با استفاده از این دستور می‌تواند مفهوم مورد نظر را حذف نماید. رابطه بین مفهوم مورد نظر و دیگر مفاهیم با توجه با ساختار اصطلاحنامه حذف خواهد شد که ممکن است موجب ایجاد ناهمگونی‌های جدید گردد.

Assign New URI

هر مفهوم با یک URI منفرد مشخص می‌شود. برای اینکه بتوانید از میان چند مفهوم مشابه، یکی که علیرغم تشابه مفهومی دیگر دارد را جدا نماییم، باید با استفاده از این دستور یک URI جدید به آن اختصاص دهیم.

Term Incorporation

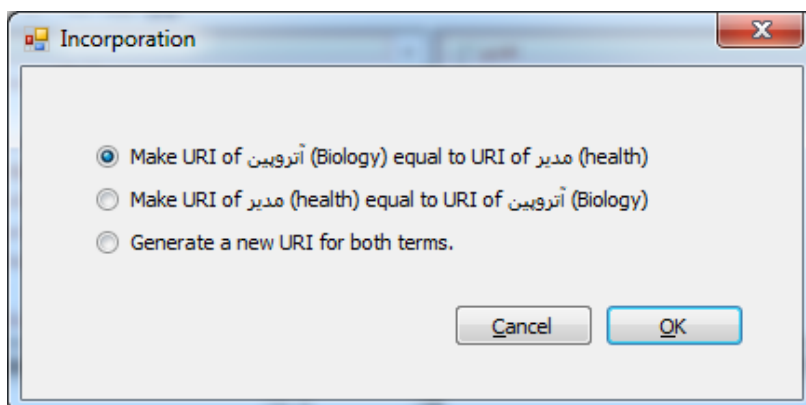
گاهی لازم است یک مفهوم که در دو اصطلاحنامه به دو شکل مختلف و یا با روابط نامشابه آمده‌اند را کاملاً یکسان نماییم. با این دستور به سادگی این عمل قابل انجام است. ابتدا در پنجره‌ای که باز می‌شود مشخص نماییم که کدام مفهوم دقیقاً مشابه چه مفهوم دیگری گردد و کلید Incorporate را فشار دهیم.



دقت نماییم که تغییرات ایجاد شده موجب ناهمگونی‌های جدید نگردد.

URI Incorporation

در مواردی ممکن است لازم باشد دو مفهوم را با وجود اختلاف ظاهری یکی در نظر بگیریم، از این گزینه استفاده نموده تا URI مربوطه را یکسان نماییم.



Related and
Opposite Terms
Incorporation

روابط وابسته بین دو مفهوم با این دستور یکسان میشود. با توجه به این که واژه‌هایی از یک اصطلاحنامه دیگر به عنوان رابطه وابسته به مفهوم مورد نظر تزریق میگردد، این عمل ممکن است موجب ایجاد اشکالاتی در ساختار اصطلاحنامه گردد.

متخصصین موضوعی در فرآیند یکسان‌سازی باید بسیار مراقب باشند تا تغییرات مورد نظر آنها با در نظر گرفتن تمام جوانب آن انجام شود. هرگونه تصمیم نادرست برای یک مفهوم میتواند موجب شود تا مشکلات عمیقتری بوجود آید.

از طرف دیگر، توجه داشته باشیم که این نرم‌افزار برای ساخت اصطلاحنامه نیست و صرفاً برای نمایان کردن ناهمگونی‌های بین اصطلاحنامه‌ها است. متخصص موضوعی می‌تواند تغییرات لازم را مستقیماً در نرم‌افزار Thesaurus Builder یا هر نرم‌افزار دیگری که برای ساخت اصطلاحنامه خود استفاده می‌کند، انجام دهد. انجام تغییرات در Thesaurus Synchronizer محدود به رفع ناهمگونی‌ها است و انجام آن به مراتب ساده‌تر خواهد بود.

پس از تکمیل عملیات یکسان‌سازی، اطلاعات توسط متخصص نرم‌افزار به Thesaurus Builder یا نرم‌افزار ساخت آنتولوژی "پروتزه" منتقل خواهد شد.

در این راهنما از روش بدست آوردن ناهمگونی‌ها به دلیل پیچیدگی‌های بسیار و دارا بودن نکات فنی و تخصصی، صرف نظر شده است.

توسعه هستی‌شناسی واژگانی هر یک از گرایش‌های زیر مجموعه علوم پایه و بکارگیری نرم‌افزار «همسان‌ساز» جهت رفع مشابهت‌ها و مغایرت‌ها در مرحله ادغام و تلفیق هستی‌شناسی‌های خرد صورت گرفته و هستی‌شناسی علوم پایه بصورت کلان طراحی می‌شود. این هستی‌شناسی کلان می‌تواند به عنوان ابزاری استاندارد، پشته‌ای قوی در ذخیره و بازیابی اطلاعات در پایگاه‌های اطلاعاتی پژوهشگاه علوم و فناوری اطلاعات بکار گرفته شده و از ریزش کاذب اطلاعات ممانعت بعمل آورد.

۱۵ - مراحل تلفیق اصطلاحنامه‌ها

نظر به این که انجام تلفیق اصطلاحنامه‌ها و تولید یک ماکرو تزاروس مقدمه طراحی کامل آنتولوژی کلان است، زمان شروع آن بلافاصله پس از یکسان‌سازی کامل و صحیح و دقیق اصطلاحنامه‌ها خواهد بود که توسط متخصصین موضوعی انجام می‌شود. فرآیند تبدیل از واژه - بنیان به مفهوم - بنیان کاملاً ماشینی خواهد بود و نیاز آن به متخصصین موضوعی کمتر است.

اما قبل از آن باید تلفیق اصطلاحنامه‌ها به طور کامل و دقیق مراحل ذیل را طی نماید:

۱. تصمیم‌گیری برای ادغام اصطلاحنامه‌ها یا ایجاد نقشه معنایی بین آنها و پیاده‌سازی آن توسط متخصص؛
۲. بررسی مفاهیم و تعریف اجتماع انواع روابط سلسله مراتبی، روابط هم‌ارزی و روابط وابسته توسط ماشین؛
۳. تبدیل هر یک از روابط موجود (سلسله مراتبی، رابطه و هم‌ارز) در اصطلاحنامه به روابط تعریف شده جدید بین مفاهیم توسط متخصص؛
۴. تعیین میزان ارتباط (وابسته و سلسله مراتبی) بین مفاهیم توسط متخصص؛
۵. ایجاد پیوند بین مفاهیم و منابع مرتبط توسط متخصص؛
۶. بررسی مفاهیم و تعریف انواع روابط سلسله مراتبی؛ وابسته و هم‌ارز توسط ماشین؛
۷. بررسی مفاهیم و تخصیص URI‌های یکسان توسط متخصص به کمک ماشین؛
۸. یکسان‌سازی مفاهیم مشترک بین اصطلاحنامه‌ها توسط متخصص به کمک ماشین؛
۹. بررسی روابط هم‌ارزی و انتخاب واژه مرجع مشترک توسط متخصص؛
۱۰. بررسی روابط سلسله مراتبی و وابسته و مشخص کردن روابط نامتجانس بین مفاهیم یکسان توسط ماشین؛
۱۱. تصمیم‌گیری برای هر یک از روابط نامتجانس توسط متخصص؛
۱۲. بررسی روابط سلسله مراتبی و وابسته و یکسان‌سازی آنها بر اساس روابط تعریف شده جدید توسط متخصص به کمک ماشین
۱۳. انتقال اصطلاحنامه از استاندارد ۵۹۶۴ به استاندارد ایزو ۲۵۹۶۴ و اعمال تغییرات لازم توسط متخصص
۱۴. تبدیل از حالت Term Base به حالت Concept Base
۱۵. استخراج آنتولوژی علوم پایه

۱۶- مراحل تبدیل اطلاعات واژگانی به مفهوم- بنیاد

مراحل تبدیل اطلاعات اصطلاحنامه‌های علوم پایه به قالب آنتولوژی برای انتقال به نرم‌افزارهای تولید آنتولوژی عبارتند از:

۱۶-۱- بررسی امکان انتقال اطلاعات

در این مرحله باید مشخص می‌شد که اساسا این انتقال قابل انجام هست یا خیر. برای این منظور پروژه انجام این پروسه توسط موسسات دیگر از جمله FAO و National Library of Medicine مورد بررسی قرار گرفت. با توجه به شباهت اصطلاحنامه‌های این موسسات (که بطور عمومی از استاندارد ایزو ۵۹۶۴ و ایزو ۲۷۸۸ اصطلاحنامه استفاده کرده‌اند) با اصطلاحنامه‌های پژوهشگاه، روش ارائه شده توسط آنها برای انتقال اطلاعات تاییدی بر قابلیت انتقال اطلاعات اصطلاحنامه‌های علوم پایه گردید.

۱۶-۲- بررسی نوع فرمت نهایی

طبق بررسی که به عمل آمد برای ذخیره اطلاعات آنتولوژی فرمت‌های مختلفی ارائه شده است. از جمله آنها می‌توان به OWL، OBO، KRSS2، RDF/XML، OWL و غیره اشاره کرد. با توجه به بررسی‌ها فرمت ارائه شده توسط W3C (MIT, ERCIM, Keio) برای OWL2 مبتنی بر RDF/XML برای فرمت نهایی انتخاب گردید. این فرمت عمومیت بیشتری در نرم‌افزارهای تولید آنتولوژی دارد و تبدیل آن به قالب‌های دیگر توسط این نرم‌افزارها امکان‌پذیر است.

۱۶-۳- نحوه تشکیل روابط اصطلاحنامه‌ای

بر اساس مقالاتی که بر نقد اصطلاحنامه‌هایی که به قالب آنتولوژی درآمده‌اند، ایرادات بسیاری بر نحوه تشکیل روابط اصطلاحنامه‌ای آنها وارد آمده است. مهمترین مسئله پیش رو این است که آیا روابط اصطلاحنامه‌ای بصورت روابط بین Individualها بیان گردد (مانند MeSH) یا بصورت روابط بین Classها (مانند AGROVOC). انتخاب هر یک از اینها باعث بیان انتقادهای مشابه میشد. بنابراین با بررسی بیشتر، استفاده از تلفیق هر دو روش برای اصطلاحنامه در نظر گرفته شد که خود نوآوری مطلوبی برای ارائه اصطلاحنامه در قالب آنتولوژی به حساب می‌آید. در این روش روابط سلسله‌مراتبی بصورت Class و سایر روابط به صورت Individual بیان می‌گردند.

۱۶-۴- ایجاد جدول تبدیل روابط

پیرو نتایج بدست آمده از مرحله قبل، در این مرحله جدول تبدیل روابط اصطلاحنامه‌ای به روابط آنتولوژی بصورت یک به یک تدوین گردید (جدول پیوست).

۵-۱۶- اصلاح کدهای غیر استاندارد نوشتاری

برای ساخت فرمت آنتولوژی در نظر گرفته شده، مشکلات زیادی وجود داشت. به عنوان مثال گاهی متخصص موضوعی یا تاپیست به جای علامت تنوین در کلماتی که شکل عربی دارند از (") استفاده کرده است. این قبیل مشکلات با توجه به ایجاد خطاهای غیر منتظره در بازخوانی فرمت نهایی، موجب صرف وقت زیادی برای رفع آنها گردید.

۶-۱۶- تبدیل به فرمت نهایی

در نهایت با کد نویسی فرآیند تبدیل، مراحل بالا (غیر از ۵) توسط کامپیوتر به اجرا درآمده و فایل‌های نهایی استخراج گردیدند. در نسخه بعدی Thesaurus Builder این فرآیند تبدیل گنجانده خواهد شد (اطلاعات مفهوم- بنیان بازنویسی شده به زبان OWL2 به علت حجیم بودن به شکل DVD پیوست گزارش ارائه شده است).

۷-۱۶- آزمایش فایل ایجاد شده در قالب نهایی

پس از تولید فایل‌های OWL2 (RDF/XML)، این فایلها توسط نرم‌افزار Protégé 4.2 به عنوان رایج‌ترین ابزار تولید آنتولوژی باز خوانی شده و صحت فرآیند مورد تایید قرار گرفت.

Narrower Term	<owl:Class rdf:about="[Term URI]"><rdfs:subClassOf rdf:resource="[Parent URI]"/>	
Concept Number	<ConceptNumber rdf:datatype="&xsd;int">[CN]</ConceptNumber>	
Status	<Status rdf:datatype="&xsd;string">[Status Value]</Status>	
Frequency	<Frequency rdf:datatype="&xsd;int">[Frequency]</Frequency>	
Input Date	<InputDate rdf:datatype="&xsd;date">[Input Date]</InputDate>	
Last Update	<LastUpdate rdf:datatype="&xsd;date">[Last Update]</LastUpdate>	
Approval Date	<ApprovalDate rdf:datatype="&xsd;date">[Approval Date]</ApprovalDate>	
Not Valid Date	<NotValidDate rdf:datatype="&xsd;date">[Not valid Date]</NotValidDate>	

Abbreviation	<Abbreviation rdf:datatype="&xsd:string">[ABB]</Abbreviation>	
Abbreviation For	<AbbreviationFor rdf:datatype="&xsd:string">[ABF]</AbbreviationFor>	
Additional Notes	<AdditionalNotes rdf:datatype="&xsd:string"><![CDATA[Additional Notes]]></AdditionalNotes>	
Related Term	<RelatedTerm rdf:resource="[URI to Related Term]"/>	
Translation	<Translation rdf:resource="[URI to Translation]"/>	
Group of Translations	<Translations xml:lang="en-us" rdf:about="[URI to Translation]">	
Original Term in Main Language	<OriginalTerm rdf:resource="[URI to Equivalent in Main Language]"/>	
Language	<Language rdf:datatype="&xsd:string">[Language of Translation]</Language>	
Non-Preferred Term	<NonPrefTerm rdf:resource="[URI to equivalent Term]"/>	
Preferred Term	<PrefTerm rdf:resource="[URI to equivalent Term]"/>	
Broader Term	<BroaderTerm rdf:resource="[URI to Broader Term]"/>	
Scope Note	<ScopeNote rdf:datatype="&xsd:string"><![CDATA[Scope Note]]></ScopeNote>	

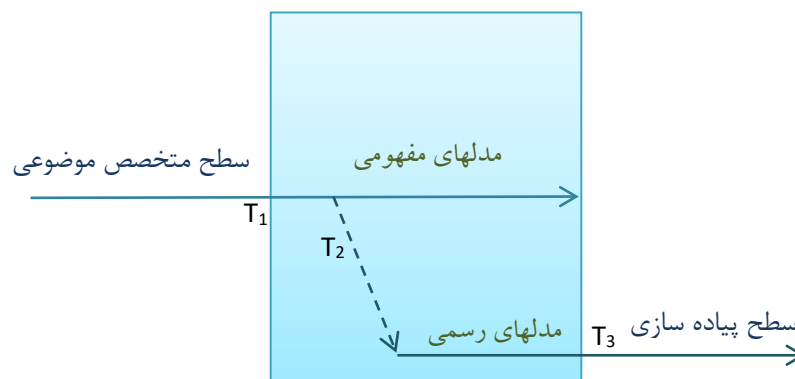
۱۷- ایجاد هستی‌شناسی

برای ساخت هستی‌شناسی از روش مبتنی بر متدولوژی «مت‌آنتولوژی» استفاده شده است. این متدولوژی، ساخت هستی‌شناسی‌ها را در سطح مفهوم بنیان مقدور می‌سازد. مبنای این روش، فعالیت‌های اصلی مشخص شده در فرآیند تولید نرم افزار و متدولوژی‌های مهندسی دانش است و شامل: مشخص کردن فرآیند تولید هستی‌شناسی و روش‌های اجرای هر یک از فعالیت‌های مدیریت، توسعه و پشتیبانی می‌شود. در این روش بیشترین میزان اکتساب دانش، مجتمع‌سازی^{۲۵۵} و ارزیابی در مرحله مفهوم‌سازی صورت می‌گیرد و این مرحله از اهمیت ویژه‌ای برخوردار است.

۱-۱۷ مدل سازی مفهومی در مت آنتولوژی

مرحله مفهوم سازی در روش مت آنتولوژی نیاز به توجه خاصی دارد زیرا نقش تعیین کننده ای در ادامه ساخت هستی شناسی دارد. هدف این مرحله سازماندهی دانش بدست آمده در طی فعالیت کسب دانش و نمایش آن به صورتی مستقل از الگوهای صوری و زبان های پیاده سازی مورد استفاده برای فرمول بندی و پیاده سازی هستی شناسی است. در نتیجه، فعالیت مفهوم سازی رابطه نیرومندی با فعالیت اکتساب دانش دارد.

وقتی مدل مفهومی ساخته شود، متدلوژی پیشنهاد می دهد که مدل مفهومی به مدل رسمی تبدیل شود. مدل رسمی با استفاده از یک زبان پیاده سازی هستی شناسی، پیاده سازی خواهد شد. در طول این فرآیند، هستی شناسی بتدریج از سطح دانش به سمت سطح پیاده سازی حرکت می کند و درجه رسمیت^{۲۵۶} مدل دانش کم کم افزایش پیدا می کند بطوریکه برای ماشین قابل فهم خواهد شد.



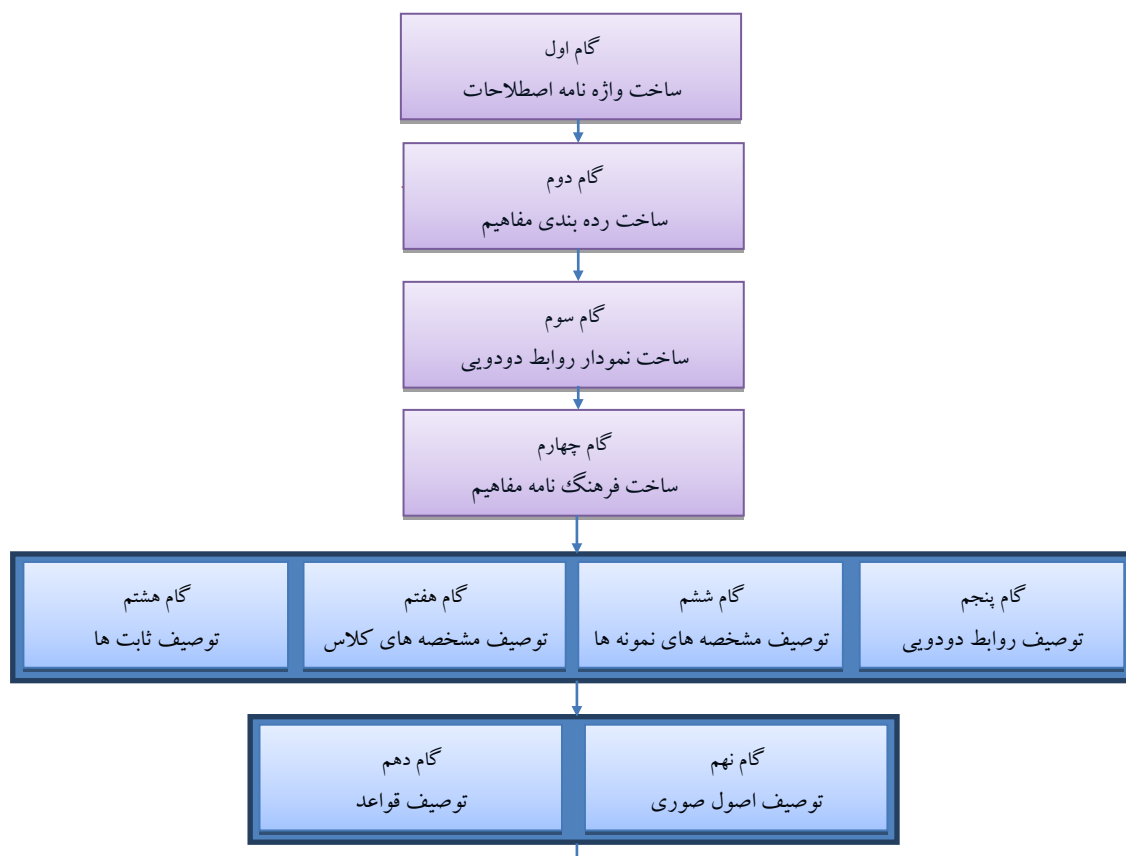
شکل ۱: مدل فرآیند اصلی در ایجاد هستان شناسی.

در شکل ۱، مدل فرآیند اصلی در مهندسی هستی شناسی نشان داده شده است. تبدیل T_1 ، نشانگر فرآیند مدل سازی مفهومی است و می توان آن را به صورت تبدیل تصویری از دامنه به یک مدل مفهومی که آن تصور را شرح می دهد در نظر گرفت. تبدیل T_2 مدل مفهومی را به یک مدل رسمی تبدیل می کند. تبدیل T_3 مدل رسمی را به یک مدلی که قابلیت اجرا در رایانه داشته باشد تبدیل می کند. در شکل، T_1 و T_3 به صورت خطهای پیوسته و T_2 به صورت خط چین نشان داده شده است.

این تفاوت به این علت است که قدری از دانش دامنه ممکن است در هنگام تبدیل از مدل مفهومی به مدل رسمی از دست برود.

مدل سازی مفهومی در مت آنتولوژی دید ادراکی غیر رسمی از دامنه را به مشخصاتی نیمه رسمی تبدیل می کند. این مشخصات به صورت مجموعه ای ارائه ی واسطه مبتنی بر گراف و جدول نشان داده می شود که برای متخصصان موضوعی و ایجاد کنندگان هستی شناسی قابل فهم است. فعالیت مدل سازی مفهومی در مت آنتولوژی شامل مجموعه ای از وظایف برای سازماندهی دانش است که در شکل ۲ نشان داده شده است. در این شکل ترتیب پیشنهادی برای ایجاد کردن مولفه های مختلف هستی شناسی در طول فعالیت مدل سازی مفهومی شرح داده شده است.

در گام اول واژه نامه اصطلاحات باید ساخته شود. این واژه نامه، مجموعه ای از اصطلاحاتی را که هستی شناسی در بر می گیرد مشخص می کند. علاوه بر این در واژه نامه، تعریف هر اصطلاح به زبان طبیعی و همچنین مترادف ها و متضادهای هر واژه وجود دارد. در گام دوم رده بندی مفهومی برای طبقه بندی مفاهیم ساخته می شود. خروجی این گام می تواند یک یا چند رده بندی باشد که در آنها مفاهیم دسته بندی شده اند. گام سوم مربوط به ساخت نمودارهای روابط دودویی بین مفاهیم برای مشخص کردن روابط بین مفاهیم در هستی شناسی است. گام چهارم ساختن فرهنگ نامه مفاهیم است که شامل نمونه های هر مفهوم، ویژگیهای نمونه ها و کلاسها و روابط بین آنها است.



شکل ۲: گام‌های فعالیت مدل‌سازی مفهومی بر طبق مت‌آنتولوژی

وقتی فرهنگ‌نامه مفاهیم ساخته شد، هستان‌شناس باید در گام‌های پنجم تا هشتم جزئیات هر یک از روابط دودویی، ویژگی‌های نمونه‌ها، و مشخصه‌های کلاس‌ها و نیز ثابت‌های اصلی دامنه را در فرهنگ‌نامه مفاهیم بتفصیل مشخص کند. در گام‌ها نهم و دهم اصول صوری و قواعدی که برای بررسی محدودیت و استنباط ارزش مشخصه‌ها بکار می‌رود توضیح داده می‌شود. در نهایت هستان-شناس می‌تواند به صورت اختیاری اطلاعات بیشتری درباره‌ی نمونه‌ها اضافه کند.

۲-۱۷ طراحی هستی‌شناسی

با توجه به روش مت‌آنتولوژی که در بخش قبل توضیح داده شد، طراحی هستی‌شناسی طی مراحل زیر صورت گرفته است:

۱. ساخت واژه‌نامه مجموعه اصطلاحات

واژه‌نامه اصطلاحات بر مبنای اصطلاحات موجود در اصطلاحنامه تشکیل داده شد. این اصطلاحات در هر حوزه توسط متخصصان موضوعی از منابع مختلف به یکی از دو روش قیاسی و یا استقرائی، استخراج و گزینش شده است.

۲. یکسان‌سازی مفاهیم

از آنجایی که اصطلاحات موجود در حوزه‌های مختلف توسط متخصصان موضوعی مختلف استخراج شده بود، نیاز به یکسان‌سازی مفاهیم مشترک در حوزه‌های مختلف وجود داشت. این کار توسط متخصصین موضوعی و با کمک ماشین انجام گرفت.

۳. ساخت رده‌بندی مفاهیم

رده‌بندی مفاهیم بر اساس نظام مفهومی موجود در اصطلاحنامه‌ها شکل گرفته است. هر اصطلاح مرجع بعنوان یک مفهوم یا کلاس در نظر گرفته شده است و روابط (اخص/اعم) به روابط (زیر کلاس/ سوپر کلاس) تبدیل شده است. یعنی اگر یک مفهوم بر اساس اصطلاحی در نظر گرفته شده باشد که نسبت به یک اصطلاح دیگر اخص باشد، این مفهوم زیر کلاس مفهومی خواهد بود که بر اساس اصطلاح دوم در هستان‌شناسی در نظر گرفته شده است. بخش‌هایی از رده‌بندی مفاهیم در هستان‌شناسی در شکل‌های ۳ تا ۶ نشان داده شده است.

۴. تعریف روابط دودویی میان مفاهیم

روابط دودویی بین مفاهیم بر اساس روابط مفهومی بین اصطلاحات در اصطلاحنامه، برای هستان شناسی در نظر گرفته شد. بین هر دو مفهوم وابسته در اصطلاحنامه یک رابطه دودویی از نوع SymmetricProperty با برچسب وابسته در هستی شناسی در نظر گرفته شد. روابط دیگر نیز به صورتی مشابه در هستی شناسی لحاظ شده است.

۵. ساخت فرهنگ نامه مفاهیم

اصطلاحات نامرجح، ترجمه اصطلاحات و عبارت های اختصاری به صورت individual تعریف شده است. هر اصطلاح نامرجح نمونه ای از کلاس مربوط به اصطلاح مرجح خود در نظر گرفته شده است. روابط دودویی بین نمونه ها بر اساس نوع روابط بین اصطلاحات مربوط به آن ها در اصطلاحنامه مشخص می شود.

۳-۱۷ تولید هستی شناسی

مرحله تولید هستان شناسی با توجه به مدل مفهومی طراحی شده صورت گرفته است. در این مرحله پس از آماده سازی اطلاعات موجود و انجام فعالیت های مربوط به یکسان سازی اصطلاحات در اصطلاحنامه های مختلف، اصطلاحنامه از استاندارد ایزو ۵۹۶۴ به استاندارد ایزو ۲۵۹۶۴ منتقل شد و با اعمال تبدیلات در نظر گرفته شده در مرحله طراحی، اصطلاحنامه از حالت واژه بنیان به حالت مفهوم بنیان تغییر وضعیت پیدا کرد. بخش هایی از هستی شناسی تولید شده در شکل های ۷ و ۸ نشان داده شده است.

۴-۱۷ غنی سازی هستی شناسی

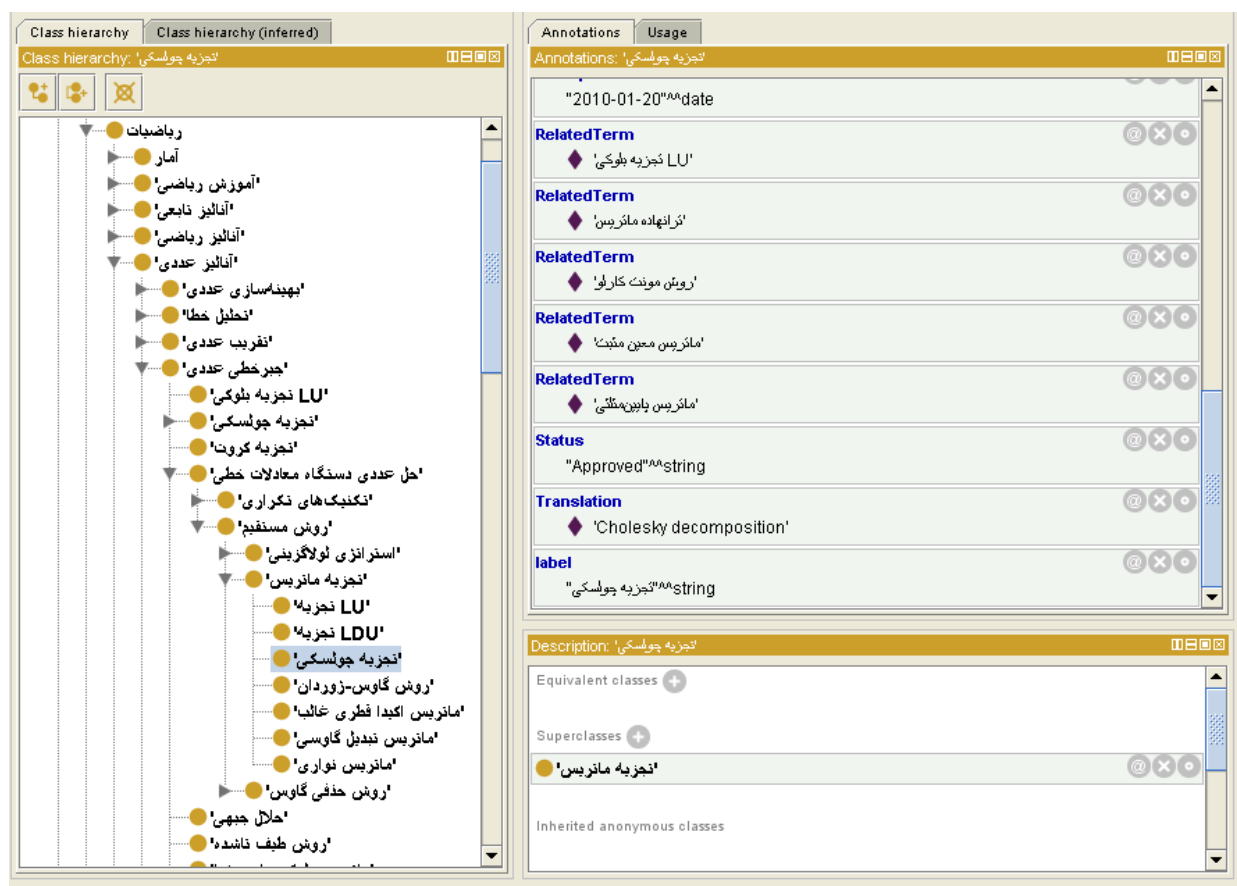
پس از تولید هستی شناسی به منظور دقیق تر کردن روابط معنایی تعریف شده در هستی شناسی، در مرحله بعد باید روابط معنایی توسط متخصص موضوعی پالایش شود. این روابط شامل روابط سلسله مراتبی و روابط وابسته هستند. در این مرحله متخصص موضوعی باید الگوهای روابط معنایی بین مفاهیم را بیابد و به صورت قاعده بیان کند تا قابل اعمال بر روی موارد مشابه باشد. این الگوها نوع خاصی از محدودیت هستند. مبنای این محدودیت ها روابط موجود در اصطلاحنامه از یک طرف و نوع موجودیت های مفاهیم مورد بحث از طرف دیگر است. جدول ۱ چند نمونه از محدودیت ها بر مبنای روابط اصطلاحنامه را نشان می دهد.

جدول ۱: نمونه‌هایی از محدودیت‌ها بر مبنای روابط اصطلاحنامه

روابط ممکن در هستان‌شناسی	روابط اصطلاحنامه
<code><hasMember>/<memberOf></code> <code><includesSpecific>/<isa></code> <code><spatiallyInclude>/<spatiallyIncludeIn></code> ...	NT/BT
<code><similarTo></code> <code><growsIn>/<EnvironmentForGrowing></code> <code><treatmentFor>/<treatedWith></code> <code><hasMember>/<memberOf></code> ...	RT

نمونه‌هایی از روابط استخراج شده توسط متخصصین در جدول ۲ نشان داده شده است. در این جدول

X و Y مفاهیم هستان‌شناسی هستند.



شکل ۷: بخشی از هستان‌شناسی تولید شده



شکل ۸: بخشی از هسته شناسی تولید شده

جدول ۲: نمونه‌هایی از روابط بین مفاهیم

روابط سلسله مراتبی (<i>Isa</i>) X <includesSpecific> Y / Y <isa> X X <inheritsTo> Y / Y <inheritsFrom> X روابط جزء-کل X <containsSubstance> Y / Y <substanceContainedIn> X X <hasIngredient> Y / Y <ingredientOf> X X <madeFrom> Y / Y <usedToMake> X X <yieldsPortion> Y / Y <portionOf> X X <spatiallyIncludes> Y / Y <spatiallyIncludedIn> X X <hasComponent> Y / Y <componentOf> X X <includesSubprocess> Y / Y <subprocessOf> X X <hasMember> Y / Y <memberOf> X مثال‌هایی از روابط دیگر X <causes> Y / Y <causedBy> X X <instrumentFor> Y / Y <performedByInstrument> X X <processFor> Y / Y <usesProcess> X X <beneficialFor> Y / Y <benefitsFrom> X X <treatmentFor> Y / Y <treatedWith> X X <harmfulFor> Y / Y <harmedBy> X X <hasPest> Y / Y <afflicts> X X <growsIn> Y / Y <growthEnvironmentFor> X X <hasProperty> Y / Y <propertyOf> X X <hasSymptom> Y / Y <indicates> X X <similarTo> Y / Y <similarTo> X X <oppositeTo> Y / Y <oppositeTo> X X <hasPhase> Y / Y <phaseOf> X X <growsIn> Y / Y <EnvironmentForGrowing> X X <ingests> Y / Y <ingestedBy> X
--

۱۷-۵ مثال‌هایی از روابط در هستان‌شناسی

• X <includeSpecific> Y / Y <isa> X

این رابطه به مفهوم این است که Y همه مشخصات X را دارا است و نوع خاصی از X محسوب می‌شود. این یک رابطه عام استاندارد بین مفاهیم است و برای نشان دادن روابط سلسله مراتبی کاربرد دارد. در مثال‌های زیر همه روابط به جز مورد آخر در اصطلاحنامه رابطه NT بوده‌اند. مورد آخر رابطه‌ای از نوع UF بوده است.

chemical soil type <includeSpecific> saline soils

خاک‌های شور <includeSpecific> انواع شیمیایی خاک

cells <includeSpecific> blood cells

سلول‌های خون <includeSpecific> سلول‌ها

۶-۱۷ روابط جزء-کل

روابط متعددی در زمره روابط جزء-کل قرار می‌گیرند.

• **X <containsSubstance> Y / Y <substanceContainedIn> X** و

X <hasIngredient> Y / Y <ingredientOf> X

مفهوم **Y** جنس یا ماده‌ای است که **X** از آن بوسیله طبیعت <containsSubstance> یا

بدست انسان <hasIngredient> ساخته شده است. **Y** وقتی در ترکیب **X** قرار می‌گیرد

ماهیت خود را از دست می‌دهد. در اصطلاحنامه این رابطه به صورت **BT/NT** یا **RT** ظاهر شده

است. چند مثال در زیر نشان داده شده است:

Blood < containsSubstance > blood lipids

NT در اصطلاحنامه

لیپیدهای خون < containsSubstance > خون

Blood < containsSubstance > blood proteins

NT در اصطلاحنامه

پروتئین‌های خون < containsSubstance > خون

Blood < containsSubstance > blood cells

NT در اصطلاحنامه

سلول‌های خون < containsSubstance > خون

Cocoa beverages <hasIngredient> cocoa powder

RT در اصطلاحنامه

پودر کاکائو <hasIngredient> نوشیدنی‌های کاکائویی

• **X <yieldsPortion> Y / Y <portionOf> X**

این نوع رابطه، ارتباط بین توده **X** و قطعه **Y** که از این توده برگرفته شده است را نشان می‌دهد. به عنوان مثال:

Roots <yieldsPortion> cuttings RT در اصطلاحنامه
 قلمه‌ها <yieldsPortion> ریشه‌ها

• **X <spatiallyIncludes> Y / Y <spatiallyIncludedIn> X**

این رابطه برای اشیائی که دارای وسعت فضایی هستند استفاده می‌شود. این رابطه زمانی استفاده می‌شود که **X** یک جزء لایتجزا از **Y** است که قابل شناسایی است ولی ذاتاً تفکیک ناپذیر است. بخش‌های بدن و موقعیت‌های جغرافیایی شامل این رابطه می‌شوند.

Asia <spatiallyIncludes> East Asia NT در اصطلاحنامه
 شرق آسیا <spatiallyIncludes> آسیا

این رابطه دارای خاصیت تعدی یا انتقالی نیز هست. یعنی

اگر **Y <spatiallyIncludes> X** و **X <spatiallyIncludes> Z** آنگاه **X <spatiallyIncludes> Z**

Asia <spatiallyIncludes> East Asia
East Asia <spatiallyIncludes> China
→Asia <spatiallyIncludes> China

• **X <hasComponent> Y / Y <componentOf> X**

این رابطه زمانی برقرار است که **Y** قسمتی از **X** است و هویت خود را بعنوان یک شیء حتی زمانی که بخشی از یک شیء دیگر است حفظ می‌کند. بعلاوه **Y** باید شمارش پذیر یا نامبردنی باشد. مثال:

Ploughs <hasComponent> ploughshare RT در اصطلاحنامه
 تیغه‌های خیش <hasComponent> خیش‌ها
nucleus <hasComponent> chromosome NT در اصطلاحنامه
 کروموزوم <hasComponent> هسته (ساختار سلول)

• **X <includeSubprocess> Y / Y <subprocessOf> X**

فرآیندهای زیادی در اصطلاحنامه وجود دارد که می‌توانند با این رابطه به یکدیگر پیوند داده شوند، بعنوان مثال:

ATP cycle <includesSubprocess> phosphorylation RT در اصطلاحنامه
فسفریل‌دار شدن (کردن) <includesSubprocess> چرخه آ تی پی

• **X <hasMember> Y / Y <memberOf> X**

X <hasMember> Y نشان دهنده رابطه عضویت در یک جمع یا گروه یا سازمان است. این رابطه خاصیت تعدی یا انتقالی ندارد.

Francophone Africa <hasMember> Benin RT در اصطلاحنامه
بنین <hasMember> آفریقای فرانسوی زبان

pesticide crops <hasMember> Artemisia absinthium RT در اصطلاحنامه
آرتمیزیا آبسیتیوم <hasMember> گیاهان آفتکش

مثال‌هایی از روابط دیگر:

• **X <causes> Y / Y <causedBy> X**

مثال:

Overgrazing <causes> desertification RT در اصطلاحنامه
بیابان‌زایی <causes> فزون‌چرایی

• **X <instrumentFor> Y / Y <performedByInstrument> X**

این رابطه این حقیقت را بیان می‌کند که مفهوم **X** وسیله‌ای برای دستیابی به مفهوم **Y** است. مثال:

plough <instrumentFor> ploughing

RT در اصطلاحنامه

شخم‌زنی <instrumentFor> خیش

وسیله مورد نظر می‌تواند توسط یک موجود زنده (انسان یا حیوان) یا یک ماشین و یا یک سیستم بکار گرفته شود. معنی این رابطه اشاره به نتیجه‌ای است که با استفاده از یک وسیله حاصل می‌شود.

• X <processFor> Y / Y <usesProcess> X

این حالتی است که X نشان‌دهنده فرآیندی است که در Y درگیر است. مثال:

soil injection <processFor> fertilization

RT در اصطلاحنامه

بارورسازی <processFor> تزریق به خاک

• X <hasProperty> Y / Y <propertyOf> X

blood circulation <hasProperty> blood pressure

NT در اصطلاحنامه

فشار خون <hasProperty> گردش خون

روابط معنایی موجود در اصطلاحنامه به ما این امکان را می‌دهد تا آن را به هستی‌شناسی تبدیل کنیم و یا به عنوان منبعی برای ساختن هستی‌شناسی از آن استفاده کنیم. برای استخراج روابط هستی‌شناختی از اصطلاحنامه باید پالایش و غنی‌سازی روابط معنایی صورت گیرد. اصطلاحنامه محتوی تعداد زیادی رابطه است که می‌تواند به صورت خاص‌تر بیان و تبدیل شود، اما به سختی ممکن است که بتوان همه انواع روابط را تبدیل کرد. در نتیجه رابطه کلی **Related Term** باید برای بیان روابط باقیمانده همچنان نگهداشته شود. به نمونه‌های این رابطه باید یک یادداشت که معنی یا منظور رابطه را مشخص می‌کند ضمیمه شود. این کار استنتاج انواع جدید روابط از این روابط باقیمانده را در آینده تسهیل خواهد کرد. (نمودارهای پیوست روابط معنایی تکمیل شده هستی‌شناسی را نشان می‌دهد).

۱۸- نتیجه گیری

روش طراحی آنتولوژیها در گروه «هستی شناسی» دانشگاه پلی تکنیک مادرید^{۲۵۷} تحت عنوان «مت آنتولوژی»^{۲۵۸} شکل گرفت که نه تنها طراحی که ساخت هستی شناسی ها را در سطح مفهوم بنیان مقدور می سازد. این روش ریشه در فعالیت های اصلی فرآیند توسعه نرم افزاری هستی شناسی ها و براساس «مهندس دانش»^{۲۵۹} دارد و شامل: شناسایی فرآیند تولید آنتولوژی، برنامه ریزی نرم افزاری، نیز اجرا و پیاده سازی هر یک از فعالیت های مدیریت و توسعه و پشتیبانی آنتولوژیها می شود. روش «مت آنتولوژی» با هدف ساخت و پشتیبانی چرخه حیات آنتولوژی و مبتنی بر تولید و توسعه «نمونه های اصلی»^{۲۶۰} یا «الگوهای نخستین» بوجود آمد، زیرا افزودن، تغییر و یا انتقال مفاهیم و به تبع آن اصطلاحات را در هر ویرایش جدیدی امکان پذیر می سازد و برای افزودن هر «نمونه اصلی» به شناسایی و رده بندی میزان زمان و هزینه مورد نیاز محاسبه می نماید.

سپس تعیین جزئیات و مختصات معنایی آنتولوژی را آغاز می کند و همزمان فعالیت های جنبی دیگری مانند کنترل کیفیت فرآیندهای ساخت و نیز پشتیبانی مانند: آموزش دانش فنی، تلفیق و ادغام، ارزیابی، مستندسازی و رده بندی ساختار آنرا انجام می دهد.

همه این فعالیت های مدیریتی و پشتیبانی بطور موازی با فعالیت های تولید و توسعه آنتولوژی (تخصیص ویژگیها، مفهوم سازی، فرمول بندی، پیاده سازی و نگهداری و پشتیبانی) در چرخه حیات یک آنتولوژی صورت می پذیرد. به محض آن که اولین «نمونه اصلی» تخصیص معنایی یافت، الگوی مفهومی آن نیز ساخته می شود و همانند یک پازل معنایی هر قطعه آن نشانگر فرآیند آموزش یک حوزه دانشی است که در طول مفهوم سازی تکامل یافته و بدان افزوده شده سپس فرمول بندی و پیاده سازی آن شکل می گیرد. بدین ترتیب اگر در یکی از این فعالیت ها کمی و کاشتی دیده شود، می توان به فعالیت های پیشین مراجعه و اصلاحات را انجام داد.

اگر ابزارهایی چون «ویرایشگر آنتولوژی»^{۲۶۱} بکار گرفته شود، الگوی مفهومی را می توان به طور خودکار برای زبانهای مختلف (زبانهای راست و چپ) بکار گرفت. در نتیجه فرمول بندی «مت آنتولوژی» انعطاف پذیر می باشد و در زبانهای مختلف کاربرد دارد.

در مرحله پشتیبانی آنتولوژیها نیز آموزش دانش فنی، ادغام و تلفیق و ارزیابی در طول مفهوم سازی و

^{۲۵۷} - Universidad Politecnica de Madrid

^{۲۵۸} - METHONTOLOGY

^{۲۵۹} - Knowledge Engineering

^{۲۶۰} - Prototype

^{۲۶۱} - Web ODE

گسترده‌تر از آن در مراحل فرمول‌بندی و تهیه الگوها و نیز پیاده‌سازی آنها تداوم دارد و عوامل پیشرفت آن عبارتند از:

الف - بیشترین دانش فنی در ابتدای ساخت آنتولوژیها صورت می‌پذیرد.
ب - ادغام و تلفیق در دیگر آنتولوژیها بمرور زمان انجام یافته و تا مرحله پیاده‌سازی به تعویق انداخته نمی‌شود.

ج - مفهوم سازی آنتولوژی بایستی به طور دقیق ارزیابی شود تا از اشاعه خطاها در مراحل بعدی چرخه حیات آنتولوژی ممانعت بعمل آید.

بمنظور ساخت هستی‌شناسی‌ها فعالیت ذیل بایستی انجام پذیرد:

۱- تهیه و تدوین اصطلاحنامه‌های تخصصی شامل مفاهیم مورد نیاز هستی‌شناسی، تعریف آنها

در زبانهای طبیعی، مترادف‌ها و سر واژه‌ها،

۲- ساخت تاکسونومی مفاهیم شامل طبقه‌بندی مفاهیم که می‌تواند منجر به ساخت

تاکسونومی‌های موضوعی (زیر رشته‌های علوم پایه) گردد،

۳- طراحی شبکه مفاهیم و روابط فیما بین مفاهیم یک آنتولوژی با مفاهیم آنتولوژیهای دیگر.

۴- ساخت فهرست مفاهیم، مختصات و اجزاء معنایی و روابط معنایی آنها،

۵- توصیف روابط معنایی و اجزاء و مختصات هر مفهوم در نظام مفاهیم،

۶- توصیف طبقات و رده‌بندی مفاهیم در یک نظام مفاهیم (دامنه موضوعی)،

۷- تدوین قواعد و الگوهای معنایی به منظور بکارگیری در نرم‌افزار،

۸- تعیین مغایرت‌ها و مشابهت‌ها در نامگذاری مفاهیم با استفاده از نرم‌افزار «همسان ساز»،

۹- رفع اختلاف‌ها و مغایرت‌ها میان اصطلاحنامه‌ها و نظام‌های مفاهیم،

۱۰- شکل‌گیری شبکه مفاهیم و اصطلاحات «علوم پایه» پس از ادغام و تلفیق هستی‌شناسی‌های

زیر رشته‌های تخصصی آن.

- Arano, Silvia (2005) Thesauruses and Ontologies , <http://www.hipertext.net,num3,2005>.
- Asanee, Kawtrakul, ...et al.(2005)Automatic term relationship cleaning and refinement for AGROVOC. (online) available:<ftp://ftp.fao.org/docrep/fao/2008>.
- Brentano, franz (1874) Psychologie vom empirischen Standpunkte, translated as Psychology from and Empirical standpoint by A. c. Rancurello, d. B. Terrell, & L.L. McAlister, routedge, London.
- Bundy, Alan (2007) Cooperationg Reasoning Processes: More than Just the Sum of their Parts, Proceedings of IJCAI-07, PP. 2-11.
- Bundy, Alan, & Fiona McNeill (2006) Representation as a fluent: An AI challenge for the next half century, IEEE Intelligent Systems 21:3, pp. 85-87.
- Croft, William, & d. Alan Cruse (2004) Cognitive Linguistics, Cambridge University Press, Cambridge.
- Cruse, D. Alan (1986) Lexical Semantics, Cambridge University Press, New York.
- Fillmore, Charles (1988) the Mechanisms of Construction Grammar, Berkeley Linguistics Society 35, 35-55.
- Frege, Gottlob (1879) Begriffsschrift, English translation in J. van Heijenoort, ed. (1967) from Frege to Godel, Harvard University Press, Cambridge, MA, 1-82.
- Gilchrist, A. (2002). "Thesauri, taxonomies and ontologies an etymological note". EN: Journal of Documentation 59 (1). 7-18.
- Grice, H. Paul (1975) Logic and Conversation, in P. Cole & J. Morgan, eds., Syntax and Semantics 3: Speech Acts, Academic Press, New York, 41-58.
- Gomes Perez (2005). Ontological Engineering: a state of the art. [en linea] Pennsylvania: School of Information Sciences and Technology (IST). Pennsylvania State University. <http://citeseer.Mj.nec.com/444416.html> [consulta: 19 marzo 2005]
- Gruber (1993b). Toward principles for the design of ontologies used for knowledge sharing. [en linea]. Pennsylvania: school of Information Sciences and technology (IST). Pennsylvania State University. <http://citeseer.nj.nec.com/gruber93toward.html> [Consulta: 19 marzo 2005]
- Guarino, N., Guiaretta (1995). Ontologies and Knowledge bases: towards a terminological clarification. [en linea]. Trento, Roma: Laboratory for Applied Ontology. <http://www.loa-cnr.it/papers/kbks95.pdf> [consulta: 19 marzo 2005]
- Guarino, N. (1998b). Formal ontology and information systems. [en linea]. Trento, Roma: Laboratory for Applied Ontology (LOA) <http://www.Ladseb.Pd.cnr.it/infor/Ontology/papers/fOIS98.pdf> [consulta: 19 marzo 2005]
- Halliday, M.A.K. & Christian M.I.M. Matthiessen (1999) Construing Experience Through Meaning: a Language-Based Approach to Cognition, Cassell, London.
- Jackendoff, Ray (1983) Semantics and Cognition, MIT Press, Cambridge, MA.
- Khosravi, F. & Vazifedoost, A.(2008) Creating a Persian Ontology through Thesaurus Reengineering for Organizing the Digital Library of the National Library of Iran. In: Faslname Ketab, 70,19-36.
- Langacker, Ronald W. (1999) Grammar and Conceptualization, de Gruyter, Berlin.
- Lopez- Huertas,M.J. (1997). "Thesaurus structure design: a conceptual approach for improved interaction". EN: Journal of documentation 53(2). 139-177.
- Majumdar, Arun K., & John f. Sowa (2009) Two Paradigms are better than one and Multiple Paradigms are even better, in S.Rudolph, F. Dau, and S.O. Kuznetsov, eds., Proceedings of ICCS'09, LNAI 5662, Sublications, London.

- Miller, George A. (1995) WordNet: A lexical database for English, *Communications of the ACM* 38:11, 39-41.
- Minsky, Marvin (1975) A framework for representing knowledge, in P. Winston, ed., *The Psychology of computer Vision*, McGraw-Hill, New York, 211-280.
- Minsky, Marvin (1987) *The Society of Mind*, Simon & Schuster, New York.
- Mohanty, J.N. (1982) *Husserl and Frege*, Indiana University Press, Bloomington.
- Montague, Richard (1970) English as a formal language, reprinted in r. Montague, *formal Philosophy*, Yale University Press, New Haven, PP. 188-221.
- Ogden, C.K., & I.A. Richards (1923) *The Meaning of Meaning*, Harcourt, Brace, and World, New York, 8th edition 1946.
- Peirce, Charles S. (1902) *Logic, Considered as Semeiotic*, MS L75, edited by Joseph Ransdell, <http://members.door.net/arisbe/menu/library/bycsp/175/175.htm>
- Peirce, Charles Sanders (CP) *Collected Papers of C. S. Peirce*, ed. By c. Hartshorne, P. Weiss, & A. Burks, 8 vols., Harvard University Press, Cambridge, MA. 1931-1958.
- Prevot, L. et. al (2010) *Ontology and the Lexicon, a multidisciplinary perspective*. In: *Ontology and the Lexicon* Cambridge Uni. Press.
- Qine,j.; Paling,S. (2001). "Converting a controlled vocabulary into an ontology: the case of GEM". [en linea]. *EN: Information Research* 6 (2). <http://informationr.net/ir/6-2/paper94.html> [consulta: 19 marzo 2005]
- Saussure, Ferdinand de (1916) *Cours de Linguistique Generale*, translated by W. Baskin as *Course in general Linguistics*. Philosophical Library, New York, 1959.
- Shamsfard, M. et.al.(2006) *An Ontology Based System for Ranking Documents*, In: *Intrnational Journal of Computer Science*, Vol. 1, N. 3.
- Shamsfard, M. et. al.(2010) *Ontology Population and Semantic Annotation of Persian Texts*, In: *Computer Sience and Engineering(CSI)*.
- Schank, Roger C., ed. (1975) *Conceptual Information Processing*, North-Holland Publishing Co., Amsterdam.
- Schalley, Andrea C. &Zaefferer, Dietmar (2007) *Ontoliguistics. How Ontological Status Shapes the Linguistic Coding of Concepts*. Mouton de Gruyter. Berlin/New York.
- Shiri, A.A.; Revie, C. (2000) "Thesauri on the Web: current developments and trends" *EN: Online Information review* 24(4). 273-279. <http://dlist.sir.arizona.edu/archive/00000163/01/thesauri.pdf> [consulta: 19 marzo 2005].
- Studer, R. et al. (2005) *Knowledge engineering: principles and methods*. [en linea] Pennsylvania: School of Information Sciences and technology (IST). Pennsylvania State University. <http://citeseer.nj.nec.com/225099.html> [consulta: 19 marzo 2005]
- Soergel, D. ...et al.(2004) *Reengineering thesauri for new applications:the AGROVOC example*. *Journal of Digital Information*, Vol.4, No.4. (2004).
- Sowa, John f. (1976) *Conceptual graphs for a data base interface*, *IBM Journal of Research and Development* 20:4, 336-357.
- Sowa, John F. (1984) *Conceptual Structures: Information Processing in Mind and Machine*, Addison-Wesley, reading, MA.
- Sowa, John f (2000) *Knowledge Representation: Logical, Philosophical, and computational Foundations*, Brooks/Cole Publishing co., Pacific Grove, CA.
- Sowa. John f. (2008) *Conceptual Graphs*, in f. van Harmelen, V. Lifschitz, and b. Porter, eds., *Handbook of Knowledge Representation*, Elsevier, Amsterdam, pp. 213-237.

- Sowa, John f. (2010) The role of logic and ontology in language and reasoning, chapter 11 of theory and Applications of Ontology. Philosophical Perspectives, edited by r. Poli & J. Seibt, Berlin: Springer, 2010, pp. 231-263.
- Slype (1991). Los lenguajes de indizacion: concepcion, construcciony utilizacion en los sistemas documentales. Madrid, Salamanca: fundacion German Sanchez ruiperez.
- Talmy, Leonard (2000) Toward a Cognitive Semantics, volume I: Concept Structuring Systems, Volume II: Typology and Process in concept Structure, MIT Press, Cambridge, MA.
- Tarski, Alfred (1933) Der Wahrheitsbegriff in den formalisierten Sprachen, English trans. As the concept of truth in formalized Languages, in A. Tarski, Logic, Semantics, Metamathematics, Second edition, Hackett Publishing co., Indianapolis, pp. 152-278.
- Villazon-Terrazas et.al.(2010) A Pattern-Based Method for Re-Engineering Non-Ontological Resources into Ontologies, In: International Journal on Semantic Web and Information Systems, 6(4), 27-63.
- Weinbrenner, S.& Engler, J. (2011) SCY Ontology and metadata scheme, DIV.2, SCY consortium.
- Whitehead, Alfred North (1937) Analysis of Meaning, Philosophical Review, reprinted in a. N. Whitehead, Essays in Science and Philosophy, Philosophical Library, New youk, 122-131.
- Wierzbicka, Anna (1996) Semantics: Primes and Universals, Oxford University Press, Oxford.
- Wilks, Yorick (2006) Thesauri and ontologies, in Rapp, Sedlmeier and Zunker-Rapp, eds., Perspectives on Cognition: a Festshrift for Manfred Wettler, Pabst.
- Wills, Yorick (2008a) The Semantic Web as the Apotheosis of Annotation, but what are its Semantics? IEEE transactions on Intelligent systems, May/June 2008.
- Wilks, Yorick (2008b) What would a Wittgensteinian computational linguistics be like? Proceedings of AISB'08, Workshop on Computers and Philosophy, Aberdeen.
- Winograd, Terry (1972) Understanding Natural Language, Academic Press, New York.
- Wittgenstein, Ludwig (1921) Tractatus Logico- Philosophicus, Routledge & Kegan Paul, London.
- Wittgenstein, Ludwig (1964) Philosophische Bemerkungen, ed. By Rush Rhees, translated as Philosophical Remarks by Raymond Hargreaves and Roger White, University of Chicago Press, Chicago, 1980.
- Woods, William A. (1968) Procedural Semantics for a Question-answering Machine, AFIPS Conference Proc., 1968 FJCC, 457-471.
- Zadeh, Lotfi A. (1975) Fuzzy logic and approximate reasoning, Synthese 30, 407-428.

Abstract

The relation between ontologies and language is in the forefront of both Natural Language Processing (NLP) and Knowledge Engineering. Ontologies, as widely used models in semantic technologies, have much in common with the lexicon. A lexicon organizes words as a conventional inventory of concepts, while an ontology formalizes concepts and their logical relations. A shared lexicon is the prerequisite for Knowledge-sharing through language, and a shared ontology is the pre-requisite for Knowledge-sharing through Information Technology. In building models of language, computational linguists must be able to map accurately the relations between words and the concepts that they can be linked to.

Keywords: Ontology, Lexicon, NLP, Ontology Engineering, Information



Iranian Research Institute for Information Science and Technology

Research Project

Designing the basic science ontology and its concept-based databank

Final Report

By:

Moluk S. Hosseini Beheshti

2013