فصلنامه علمی - پژوهشی فرماندهی و کنترل، سال دوم، شماره سه ، پاییز ۱۳۹۷

گسترش شبکه واژگان زبان فارسی با روش‌های نیمه‌خودکار و قالب داده‌های پیوندی

زهره متشکر آرانی [۱]، احمد عبدالله زاده بارفروش [۲]، حسین شیرازی [۳*]

تاریخ دریافت: ۱۳۹۷/۰٤/۰۷ تاریخ پذیرش: ۱۳۹۷/۰۹/۱۲

چکیده

منابع واژگانی، از منابع ضروری در حوزه‌ی پردازش زبان طبیعی هستند و نحوه‌ی ارائه و در دسترس قرار دادن آن‌ها نیز از اهمیت ویژه‌ای برخوردار شده است. در این مقاله، قالب ارائه شبکه‌ی واژگان به عنوان یک منبع واژگانی مهم در پردازش‌های معنایی زبان طبیعی مورد توجه قرار گرفته است، با این هدف که ارائه‌ی این دانش به نحوی باشد که بتوان از مزایای داده‌های پیوندی و شبکه‌ی معنایی بهره برد. قالب داده‌های پیوندی امکان دسترسی به دانش و پیوند آن با سایر منابع دانشی را فراهم می‌آورد. ارائه‌ی شبکه‌ی واژگان به فرمت داده‌ی پیوندی و بازنمایی آن به صورت پایگاه شناخت، امکان استفاده از فناوری‌های شبکه‌ی معنایی، مانند استنتاج را فراهم می‌آورد؛ همچنین، فرایند توسعه و استفاده از شبکه‌ی واژگان با استفاده از داده‌های پیوندی به نحو مؤثرتری امکان‌پذیر است. وجود شبکه واژگان فارسی یکپارچه در قالب داده‌های پیوندی، گامی مهم در توسعه‌ی پردازش متون فارسی برای نیل به سامانه‌های آگاهی وضعیتی بلادرنگ است. این مقاله شمای یک پایگاه شناخت را، که برای یک شبکه‌ی واژگان به زبان فارسی طراحی شده است، معرفی می‌کند. این شما به صورت داده‌ی پیوندی و قابل ارزیابی با مدل لمون [۲] است. در این مقاله، گزارشی از فرایند انتقال دانش از شبکه‌ی واژگان فارس‌نت به یک پایگاه شناخت با قالب داده‌های پیوندی ارائه می‌شود و همچنین از نظر روابط و تعداد واژه‌ها، بهبودها و توسعه‌های انجام گرفته در این شبکه‌ی واژگان جدید بیان می‌گردد.

واژگان کلیدی: شبکه واژگان، پایگاه شناخت، داده پیوندی، شبکه معنایی، استنتاج.

[۱] دانشجوی دکتری، دانشگاه صنعتی مالک اشتر zohre.arani@chmail.ir

[۲] استاد دانشگاه صنعتی امیرکبیر ahmadaku@aut.ac.ir

[۳] استاد دانشگاه صنعتی مالک اشتر shirazi@mut.ac.ir (نویسنده مسئول)

[۲] lemon

۱. مقدمه

امروزه تعداد زیادی از منابع واژگانی وجود دارند که حجم قابل توجهی از اطلاعات زبانی را در اختیار کاربران انسانی و ماشینی می‌گذارند. شبکه‌های واژگان یکی از مهمترین منابع واژگانی قابل خواندن توسط ماشین هستند. در حوزه‌ی پردازش‌های زبان طبیعی وجود یک شبکه واژگان با پوشش مناسب روی واژه‌های زبان، انجام پردازش‌های نحوی و معنایی را سرعت می‌بخشد و پردازش‌های معنایی مانند ابهام‌زدایی معنایی کلمات، برچسب‌گذاری نقش معنایی ، تحلیل نیت و... را تسهیل می‌نماید. پردازش‌های ذکرشده از نیازمندی‌های مهم در سامانه‌های پردازش متن به شمار می‌روند. بسیاری از اطلاعات مورد نیاز برای آگاهی وضعیتی در قالب متن موجود است، با توجه به این که سامانه آگاهی وضعیتی از ملزومات سامانه‌های فرماندهی و کنترل یکپارچه است، توسعه‌ی منابع واژگانی از قبیل شبکه واژگان و پیوند آن با پایگاه‌های شناخت برای نیل به یک سامانه آگاهی وضعیتی بلادرنگ و هوشمند راهگشا به نظر می‌رسد.

با توجه به گستردگی و پویایی زبان طبیعی، منابع واژگانی نیازمند به‌روزرسانی مداوم هستند و لازم است میان منابع مختلف ورت داده‌های پیوندی افزایش پیدا کرده است [٤]. همچنین مدل‌سازی‌هایی برای ایجاد قالب‌های مشخص برای منابع واژگانی ایجاد شده‌اند؛ که از آن جمله می‌توان به پروژه‌های آنتولکس [٥] و مدل لمون [٦] اشاره کرد. همچنین پروژه‌ی داده‌های باز پیوندی [٧]که همچنان در حال توسعه است، نمونه‌ی بزرگی از ابر داده‌های پیوندی است، که منابع واژگانی را در بر می‌گیرد.

از سوی دیگر، شبکه‌ی واژگان خود از ماهیتی هستان‌شناسانه برخوردار است و بسیاری از روابطی که در شبکه‌ی واژگان تعریف می‌شود، ویژگی‌هایی از قبیل تقارن و یا تراگذری دارند و یا روابطی هستند که معکوس یکدیگرند. رابطه‌ی شمول نیز یکی از روابط اصلی در شبکه‌ی واژگان است. همه این روابط در منطق توصیفی به خوبی تعریف شده‌اند و در ساختار پایگاه شناخت به آسانی قابل مدیریت هستند، و قابلیت انجام استنتاج برای آن‌ها فراهم است[٨].

واژگانی پیوند برقرار باشد. اما از آنجا که این منابع در قالب‌های مختلفی تولید شده‌اند، برقراری پیوند میان آن‌ها و سایر منابع واژگانی (از همان زبان و یا زبان‌های دیگر) و یا برقراری پیوند میان آن‌ها و پایگاه‌های شناخت با مشکلاتی روبه‌رو است.

در سال‌های اخیر تلاش‌های زیادی برای پیوند دادن منابع واژگانی صورت گرفته است. به عنوان مثال، شبکه واژگان اروپایی [١] یک شبکه واژگان برای چند زبان مختلف اروپایی (شامل هلندی، ایتالیایی، اسئانیایی، آلمانی، فرانسوی، چکسلواکی، و استونی) است. علاوه بر تولید شبکه واژگان‌های چند زبانه، پروژه‌هایی برای تولید شبکه واژگان جهانی [٢] و پیوند دادن شبکه‌های واژگان به پایگاه‌های شناخت رسمی مانند سومو [٣] انجام شده‌اند.

ظهور شبکه‌ی معنایی و شبکه‌ی چیزها به تلاش‌ها برای پیوند منابع دانشی، جهت‌گیری جدیدی داده است. داده‌های پیوندی امکان انتشار داده‌های مختلف را به صورت یک‌پارچه و مرتبط به هم فراهم می‌آورد و در سال‌های اخیر گرایش و نیاز به نمایش منابع واژگانی به ص

بنابر مطالب گفته شده، با توجه به نیازمندی‌های شبکه‌ی واژگان و پیشرفت‌های صورت‌گرفته در مدل‌های نمایش منابع واژگانی، در این مقاله، روشی برای نمایش شبکه واژگان به صورت پایگاه شناخت ارائه می‌گردد. در بخش ٢، یک معرفی از شبکه‌ی واژگان، و داده‌های پیوندی و ارائه می‌گردد و مدل لمون برای انتشار منابع واژگانی به صورت داده‌های پیوندی به اختصار شرح داده می‌شود. در بخش ٣، پایگاه شناخت طراحی شده معرفی می‌گردد و در بخش ٤، فرایند انتقال اطلاعات و توسعه‌ی آن شرح داده می‌شود. در آخر نتیجه‌گیری، در بخش ٥ ارائه می‌شود.

۲. کلیات

در این بخش مباحث مرتبط با شبکه واژگان و داده‌های پیوندی ارائه می‌گردد و پژوهش‌های مرتبط اجمالاً بررسی می‌گردد.

۲-۱. شبکه‌ی واژگان

به دلیل انعطاف‌پذیری زبان طبیعی رابطه‌ی بین واژه‌ها و معانی آن‌ها یک رابطه‌ی چند به چند است. به این معنا که ممکن است هر واژه معانی مختلفی داشته باشد و از طرف دیگر به ازای هر معنا یا مفهوم ذهنی چندین واژه مورد استفاده قرار بگیرد. شبکه‌ی واژگان یک شبکه‌ی معنایی میان واژه‌ها و معانی آن‌هاست که در آن، همه‌ی واژه‌هایی که به ازای یکی از معانی خود به یک مفهوم ذهنی مشترک اشاره می‌کنند، یک گروه هم‌معنایی[۳] را می‌سازند. بر این اساس می‌توان سه کلاس واژه، معنای واژه و گروه هم‌معنایی را در این حوزه تعریف نمود، به طوریکه هر گروه هم‌معنایی نشان‌دهنده یک مفهوم ذهنی مشترک برای اهالی زبان است. هر چند ممکن است هر فرد یک مفهوم ذهنی خاص را در ذهن خود به گونه‌ای متفاوت از دیگران تجسم کند؛ اما وجوه اشتراک آن مفهوم به اندازه‌ای اند که می‌توانند با یکدیگر به انتقال اطلاعات بپردازند.

شبکه‌ی واژگان تاکنون برای حدود صد زبان مختلف ساخته شده است[1] . برای برخی از زبان‌ها مانند زبان انگلیسی نسخه‌های مختلفی از شبکه‌ی واژگان موجود است. رایج‌ترین نسخه شبکه‌ی واژگان انگلیسی، به نام وردنت[٤] یا وردنت پرینستون[٥] است ([2] و [3]) که توسط دانشگاه پرینستون ارائه شده است. وردنت پرینستون در قالب یک پایگاه داده‌ی رابطه‌ای از اسامی، افعال، صفات و قیود زبان انگلیسی ارائه شده است. واژه‌های زبان انگلیسی در بیش از ۱۱۷هزار گروه هم‌معنایی گروه‌بندی شده‌اند و روابط دو طرفه‌ای بین واژه‌ها و گروه‌های هم‌معنایی وجود دارد. این روابط شامل روابط شمول یا زیرنوع/ ابرنوعی (مانند سرو – درخت) روابط عضویت یا جزءنامی/ کل‌نامی (مانند شاخه – درخت) رابطه‌ی تضاد (مانند کوتاه– بلند) و تعدادی رابطه‌های استنتاجی (مانند فروختن– خریدن، نشان دادن– دیدن) است. عضو بودن چندواژه در یک گروه هم‌معنایی نشان‌دهنده‌ی شباهت معنایی زیاد آن واژه‌هاست و همچنین روابطی که بین گروه‌های هم‌معنایی برقرار می‌گردد؛ ارتباط معنایی آن‌ها را نشان می‌دهد. به همین علت شبکه‌ی واژگان یک ابزار اصلی برای ابهام‌زدایی معنایی کلمات [4]و تحلیل نیت و به طور کلی پردازش زبان طبیعی محسوب می‌شود.

برای زبان فارسی نیز چند شبکه واژگان مختلف تهیه شده و یا در حال تهیه است. شمس فرد و همکاران [5] با رویکرد نیمه‌خودکار یک شبکه هماهنگ با وردنت پرینستون برای زبان فارسی تهیه کرده‌اند. منتظری و فیلی [6] با استفاده از ترجمه‌ی ماشینی خودکار یک شبکه واژگان از روی وردنت پرینستون ارائه کرده‌اند. تقی زاده و فیلی [7] بر روی توسعه شبکه واژگان فارسی با استفاده از ترجمه متن به انگلیسی کار کرده‌اند. بسیاری از شبکه‌های واژگانی در زبان‌های مختلف به وردنت پرینستون نگاشت دارند و وجود این پیوند امکان ترجمه در سطح واژگانی و نیز ابهام‌زدایی‌های معنایی بین زبانی را فراهم می‌سازد. همچنین وردنت پرینستون به پایگاه‌های شناخت رسمی مانند سومو[8] نیز نگاشت دارد.

۲-۲. شبکه واژگان به عنوان یک پایگاه شناخت زبانی

هر چند ارائه‌دهندگان وردنت پرینستون آن را به عنوان یک پایگاه داده واژگانی و با قالب پایگاه داده رابطه‌ای تولید و ارائه کرده‌اند، [2] در بسیاری از متون از شبکه واژگان به عنوان یک پایگاه شناخت عمومی نام برده می‌شود، زیرا با پوشش روی لغات یک زبان، پوشش گسترده‌ای در موضوعات مختلف به دست آورده است. تعاریف متنوعی از پایگاه شناخت وجود دارد، رهنما و بارفروش با ارائه‌ی مدل کوگنی‌بیس یک پایگاه شناخت را شامل پنج عنصر معرفی می‌کنند؛ که این عناصر عبارتند از: مفاهیم (شامل کلاس‌ها و نمونه‌ها) روابط سلسله‌مراتبی، روابط غیر سلسله‌مراتبی، حوزه و دامنه‌ی پایگاه شناخت[9] . در این پژوهش این تعریف، مبنای تعریف پایگاه شناخت در نظر گرفته شده است.

می‌توان گفت که شبکه‌های واژگان و پایگاه‌های شناخت فواید و کاربردهای دوجانبه‌ای دارند. شبکه واژگان علاوه بر کاربردهای شناخته‌شده‌ای که در حوزه‌های ابهام‌زدایی معنایی کلمات، بازیابی اطلاعات و طبقه‌بندی متون دارد، در تولید و توسعه‌ی پایگاه‌های شناخت نیز نقش به‌سزایی دارد. به عنوان

[۳] **SynSet (Synonym Set)**

[٤] **WordNet**

[٥] **Princeton WordNet(PWN)**

نمونه‌هایی از کاربردهای شبکه‌ی واژگان در تولید پایگاه‌های شناخت می‌توان به تطابق میان پایگاه‌های شناخت مختلف اشاره کرد. پایگاه‌های شناخت بسیاری که در حوزه‌های گوناگون وجود دارند، اغلب شامل اطلاعات هم‌پوشان هستند. برای آن که اطلاعات در میان کاربردها و حوزه‌های مختلف قابل بازاستفاده و به‌اشتراک‌گذاری باشند؛ لازم است راه‌هایی بیابیم که پایگاه شناخت‌ها را مقایسه کنیم، تطابق دهیم و یا مجتمع کنیم. برای یافتن تشابهات معنایی میان برچسب مفاهیم و ساختارها در پایگاه‌های شناخت مختلف، از شبکه‌ی واژگان استفاده می‌گردد. در [10] و [11] روش‌هایی برای تطابق پایگاه شناخت با استفاده از شبکه‌ی واژگان ارائه شده است. کاربرد دیگر شبکه‌ی واژگان مربوط به تولید پایگاه‌های شناخت است. از آنجایی که شبکه واژگان مفاهیم را شاخص‌گذاری می‌کند و در برقراری روابط بر معنای واژه‌ها و گروه‌های هم‌معنایی آن‌ها تمرکز دارد و نه شکل ظاهری واژه‌ها یک منبع بسیار مناسب برای ساخت پایگاه‌های شناخت عمومی و یا خاص‌قلمرو است. منبع یاگو [12] یک پایگاه شناخت عمومی است؛ که با استفاده از شبکه واژگان پرینستون و جعبه‌های اطلاع ویکی‌پدیا ساخته شده است.

از سوی دیگر می‌توان از شبکه‌واژگان به عنوان یک پایگاه شناخت زبانی که در تمام حوزه‌های عمومی زبان واژه‌ها و مفاهیم را پوشش می‌دهد؛ استفاده کرد. برای این منظور لازم است، تا تغییراتی در آن اعمال شود. یکی از این تغییرات این است که بین مفاهیم کلی زبانی و نمونه‌ها (اسامی خاص) تمایز قائل شویم. در شبکه واژگان پرینستون از نسخه ۲,۰ به بعد این تمایز اعمال شده است.

همچنین استفاده از قالب پایگاه شناخت برای شبکه‌ی واژگان علاوه بر آن که در انتشار و پیوند این منبع با سایر منابع مفید است، فواید دیگری نیز دارد؛ با توجه به این که روابط در شبکه‌ی واژگان، ویژگی‌هایی از قبیل تقارن، تراگذری، و معکوس بودن را دارند، می‌توان با تعریف اصول[۶] مناسب این اطلاعات را در پایگاه شناخت مدل کرد، بدون آنکه نیاز به برنامه‌نویسی و ایجاد قالب‌های خاص حوزه باشد. همچنین، با معرفی این اصول انجام استنتاج و استخراج اطلاعات بیشتر با استفاده از موتورها و ابزارهای استنتاج میسر می‌گردد. و نیز، با مشخص بودن دامنه و برد روابط و یا محدوده‌ی مجاز مقادیر آن‌ها امکان اعتبارسنجی و یافتن خطاها میسر می‌شود و همه‌ی این امکانات توسط ابزارهای پایگاه شناخت فراهم می‌شوند.

نسخه‌های مختلفی از تبدیل شبکه‌ی واژگان پرینستون به صورت داده‌ی پیوندی منتشر شده‌اند. مهم‌ترین تبدیل انجام‌گرفته توسط دانشگاه پرینستون صورت پذیرفته است [1]. تبدیل مهم دیگر را کنسرسیوم جهانی وب [۷] انجام داده است و یک مدل در قالب استاندارد زبان هستان‌شناسی شبکه (owl)[۸] و سازگار با شبکه واژگان پرینستون در دو نسخه‌ی پایه و کامل ارائه کرده است[13] . اما طرح جامعی برای تبدیل شبکه‌ی واژگان به قالب پایگاه شناخت و ارائه‌ی آن در قالب داده‌های پیوندی وجود ندارد و برای بسیاری از زبان‌ها مانند زبان فارسی چنین تبدیلی انجام نشده است.

۲-۳. داده‌های پیوندی

فناوری‌های نوین شبکه و به طور خاص داده‌های پیوندی منجر به انتشار داده‌ها بر روی وب و برقرار کردن پیوند بین آن‌ها شده است. بر خلاف گذشته که شبکه مجموعه‌ای از اسناد به هم پیوند داده شده بود شبکه نوین مجموعه‌ای از چیزها[۹] متصل به هم است. در داده‌های پیوندی منابع و پایگاه‌های داده‌ی مختلف با قالب‌های مبتنی بر آردی‌اف[۱۰] منتشر و به هم پیوند داده می‌شوند. داده‌های پیوندی چهار اصل برای انتشار دارند: ۱- بایستی با شناسه‌های یکتا[۱۱] یکتایی و هویت آنها مشخص شود. ۲- یوآرآی ها بایستی قابل احراز باشند. ۳- با استانداردهایی مانند آردی‌اف اطلاعات معنایی آن‌ها قابل بازنمایی باشد. ٤- به سایر منابع ارجاع داشته باشند.[14]

انتشار شبکه واژگان به صورت داده‌ی پیوندی مزایای زیادی دارد که عبارتند از ۱- امکان استفاده مؤثر از مجموعه‌داده‌های

---

۶ **Axiom**

۷ **W3C**

۸ **Web Ontology Language**

۹ **Web of Thing**

۱۰ **RDF（Resource Description Framework）**

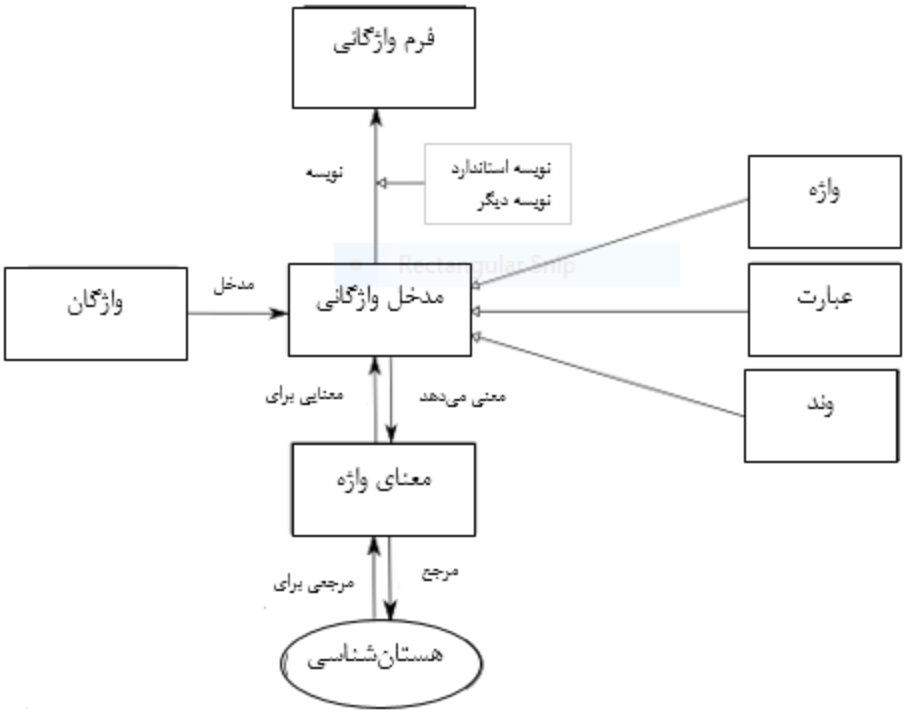
۱۱ **Unified Resource Index (URI)**

زبانی و داده‌های چندزبانی را فراهم می‌آورد. ۲- یکپارچه‌سازی اطلاعات زبانی و سایر اطلاعات دیجیتالی را ممکن می‌سازد و دیگر نیازی نیست که به ازای داده‌های هر حوزه از فرمت‌های داده‌ای و واسط‌های کاربری خاص استفاده شود. ۳- توصیف داده‌ها در فرمت آردی‌اف، امکان شاخص‌گذاری و جستجوی آن‌ها به وسیله موتورهای جستجو را نیز ممکن می‌سازد [15].

استخراج موجودیت‌های نامدار و پیوند آن‌ها به پایگاه شناخت[۱۲] و نیز شناسایی مفاهیم متن، یک پیش‌شرط برای بسیاری از کاربردهای شبکه معنایی است. بنابراین واسط بین پایگاه شناخت‌ها (که مفاهیم یک حوزه را توصیف می‌کنند) و واژه‌ها (که ویژگی‌ها یا عبارات زبانی هستند که به آن مفاهیم اشاره می‌کنند) روزبه‌روز مهم‌تر می‌شوند. پروژه‌ی آنتولکس [16] روی ایجاد یک واسط بین بازنمایی دانش در پایگاه شناخت‌ها و بازنمایی و استفاده از دانش زبانی که در منابع واژگانی (یک یا چندزبانه) وجود دارند، تمرکز دارد. و از نتایج این پروژه، مدل لمون است، که در ادامه به آن می‌پردازیم.

شکل ۱ نمایش داده شده است. این مدل شامل چهار مؤلفه‌ی اصلی است که در ادامه به توضیح آن‌ها می‌پردازیم.

مدل لمون، مدلی برای مدل‌سازی منابع واژگانی و فرهنگ‌های قابل خواندن توسط ماشین و پیوند دادن آن‌ها به وب معنایی و ابر داده‌های پیوندی است. این مدل طوری طراحی شده است که بر مبنای آردی‌اف باشد تا از فناوری‌های وب معنایی موجود استفاده کند (از جمله زبان‌های پرسش‌و پاسخ مثل اسپارکوئل[۱۳]، استانداردهای منطق توصیفی مثل زبان اننتولوژی شبکه، و قالب‌های بیان قواعد مثل آرآی اف، و ...) و بتواند اطلاعات واژگانی مربوط به پایگاه‌های شناخت را بازنمایی کند. اجزای این مدل [17] در



شکل ۱. مدل لمون برای ارائه‌ی منابع واژگانی[17]

[۱۲] **Named Entity Recognition and Disambiguation**

[۱۳] **SPARQL**

- **مدخل واژگانی:** یک مدخل واژگانی ممکن است یک واژه، یک عبارت چندواژه‌ای و یا حتی یک وند (پیشوند یا پسوند) باشد و یک واحد زبانی است که ویژگی‌های عمومی مانند برچسب ادات سخن را به ازای همه معانی و فرم‌هایش (نویسه‌هایش) می‌پذیرد.
- **فرم واژگانی:** فرم‌ها اشکال مختلف یک مدخل واژگانی هستند که ممکن است بر اثر تغییرات ریخت‌شناسی یا نوشتاری (مثل املاهای مختلف یا کدینگ‌های مختلف کاراکترهای یک واژه) و یا حتی رسانه‌های مختلف (مثل بیان گفتاری یک واژه) به وجود آیند.
- **معنای واژگان:** معنای واژه کاربرد یک واژه در یک معنای خاص است و پیوندی میان مدخل واژگانی و مدخل هستان‌شناسی برقرار می‌کند. معنای واژگانی مثل جایگاهی است که در آن می‌توان بسیاری از ویژگی‌های کاربردشناسانه واژه را حاشیه‌نویسی نمود.
- **مرجع:** یک موجودیت در پایگاه شناخت است، که می‌توان آن را به وسیله‌ی مدخل واژگانی تفسیر نمود و یا نمایش داد.

۳. پایگاه شناخت طراحی شده برای شبکه‌ی واژگان

یک پایگاه شناخت شامل بخش‌های مختلفی از جمله کلاس‌ها (مفاهیم و یا نوع‌ها)، روابط، ویژگی‌ها، قوانین و مصداق‌ها (نمونه‌های کلاس‌ها) است. روابط و ویژگی‌ها روی کلاس‌ها تعریف می‌شوند و روی مصداق‌ها مقدار می‌گیرند.

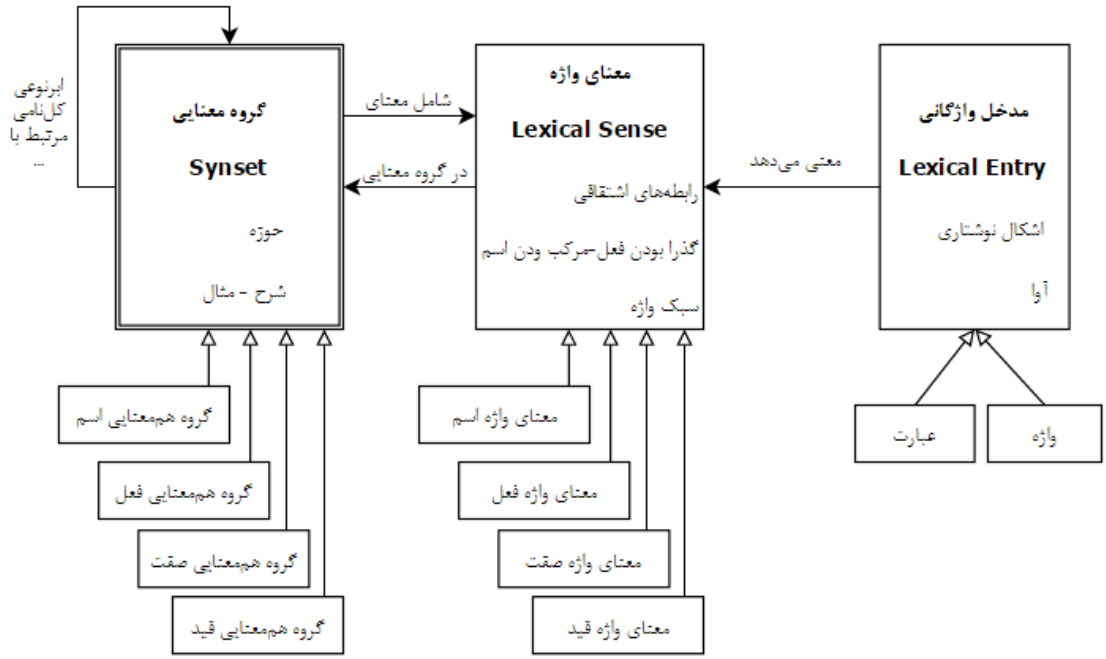
بنابراین یک تقسیم‌بندی دیگر این است که پایگاه شناخت را به دو قسمت بخش اصطلاحات[۱۴] و بخش ادعاها[۱۵] تقسیم می‌کنند. بخش اصطلاحات شامل شمای پایگاه شناخت است، و در آن کلاس‌های اصلی پایگاه شناخت و روابط، ویژگی‌ها و اصول میان آن‌ها تعریف می‌شود. اما بخش ادعاها، شامل مصداق‌ها یا نمونه‌هاست.

در طراحی شما برای شبکه‌ی واژگان سؤال اولیه این است که آیا گروه‌های معنایی که در بردارنده‌ی مفاهیم زبان هستند، بایستی به صورت مفهوم و یا به صورت مصداق مدل‌سازی شوند. هر چند وجود روابط سلسله‌مراتبی میان گروه‌های هم‌معنایی (مانند روابط زیرنوع/ ابرنوعی یا روابط جزنام/ کل‌نامی) و نیز گرفتن نمونه‌های خاص در جهان واقع به نظر می‌رسد که گروه‌های هم‌معنایی می‌توانند کلاس‌های پایگاه شناخت باشند، اما با نگاهی به هدف شبکه‌ی واژگان در می‌یابیم که می‌توان گروه‌های هم‌معنایی را به صورت مصادیق کلاس‌های گروه هم‌معنایی در نظر گرفت و روابط سلسله‌مراتبی را میان آن‌ها برقرار کرد.

در این پژوهش قالب طراحی شده برای شبکه‌ی واژگان شامل سه کلاس اصلی واژه، معنای واژه و گروه هم‌معنایی است، در یک شبکه‌ی واژگان روابط از پیش تعیین‌شده‌ای وجود دارد. این روابط ممکن است بین واژه‌ها و یا بین گروه‌های معنایی برقرار باشد و نیز ممکن است درون مقوله‌ای یا میان مقوله‌ای باشد، که در ادامه به آن‌ها بیشتر خواهیم پرداخت. به طور کلی مدل پیشنهادی مشابه شکل ۲ است:

[۱۴] **Terminology-Box(T-box)**

[۱۵] **Assertion-Box (A-box)**



شکل ۲. مدل پیشنهادی پایگاه شناخت برای شبکه واژگان

ساختار کلاس‌ها

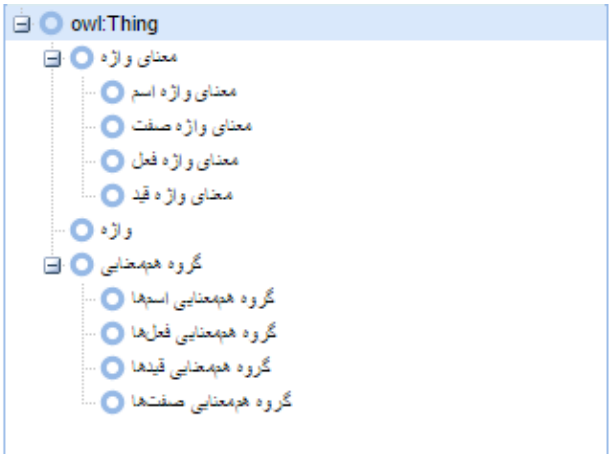
همانگونه که در شکل ۳، مشاهده می‌شود، همه کلاس‌ها در یک ساختار درختی جای می‌گیرند. ریشه این ساختار درختی در همه هستان‌شناسی‌ها شیء[۱۶] است. هر چند می‌توان نام آن را تغییر داد در شبکه واژگان ترجیح داده شده است، این کلاس به صورت پیش‌فرض باقی بماند. در شبکه واژگان کلاس‌های سطح یک (سطح بعد از ریشه) عبارتند از: واژه، معنای واژه و گروه معنایی.

بنابراین سه کلاس مختلف وجود دارد، و هر کلاس نمونه‌های خود را دارد. هر یک از مداخل معنای واژه و گروه معنایی نیز به چهار دسته مدخل بر اساس مقوله‌های نحوی (اسم، فعل، صفت و قید) تقسیم می‌شوند. این مداخل، با یکدیگر و بین خودشان ارتباطات مشخص دارند.

در مدل پیشنهادی اطلاعات مربوط به فرم واژه در کلاس واژه، اطلاعات نحوی، صرفی و اشتقاقی در کلاس معنای واژه؛ و اطلاعات و روابط معنایی در کلاس گروه معنایی

[۱۶] Thing

ذخیره می‌شوند. تمامی واژه‌های زبان، معانی و گروه‌های هم‌معنایی تشکیل شده از آن‌ها مصادیق پایگاه شناخت را تشکیل می‌دهند. بعد از طراحی هستان‌شاسی تعداد کلاس‌ها ثابت است و تنها به تعداد نمونه‌های هر کلاس اضافه می‌شود.



شکل ۳. ساختار سلسله‌مراتبی کلاس‌ها

کلاس واژه

در کلاس واژه اطلاعات نویسه‌های مختلف و آوای کلمه ثبت می‌گردد و واژه به معنا یا معناهای مختلف آن با رابطه‌ی «معنی_می‌دهد» متصل می‌شود. همانطور که در بخش ۲ گفته شد، در مدل لمون نویسه‌های مختلف یک واژه باید به درستی در مدل در نظر گرفته شوند. بسیاری از کلمات زبان فارسی نویسه‌های مختلفی دارند، (برای مثال، جابه‌جایی/ جابجایی، توفان/ طوفان و پارادوکس/ پارادکس) تمامی این نویسه‌ها در مدخل واژه ثبت می‌گردد؛ و یک نویسه‌ی معیار (بر اساس نظر زبان‌شناسان) از میان آن‌ها انتخاب می‌گردد.

کلمات زبان با به کار برده شدن در جملات، برخی تفاوت‌های ریخت‌شناسی نیز پیدا می‌کنند. این تغییرات به دو دسته‌ی تغییرات قاعده‌مند و استثنائات تقسیم می‌شوند. به عنوان مثال تغییرات رایج افعال (در صیغه‌های زمانی و شخص و شمار) و یا جمع‌های با علامت «ها» و «ان» در زبان فارسی نمونه‌هایی از تغییرات قاعده‌مند هستند. بنابراین نیازی نیست که این شکل‌های مختلف در منبع زبانی ذخیره شوند و بهتر است که با اجرای کد، ریشه‌گیری صورت بپذیرد. اما استثنائات زبانی مانند جمع‌های مکسر بهتر است در منبع واژگانی ثبت شوند. رابطه‌ی میان یک کلمه‌ی جمع مکسر و معانی آن در قسمت معنای واژه مشخص می‌شود، زیرا برای مثال، کلمه‌ی «مواد» به ازای برخی از معانی کلمه «ماده» جمع مکسر آن است، اما به ازای معنای «ماده در مقابل نر» جمع مکسر آن نیست. این اطلاعات نیز به صورت پیوندهای نوع‌دار در شبکه‌ی واژگان ثبت می‌گردند.

کلاس معنای واژه

اطلاعات نحوی در قسمت معنای واژه قرار می‌گیرند، گذرا بودن یا نبودن افعال و یا شمارش‌پذیری اسامی ویژگی‌هایی هستند که به ترتیب برای معنای واژه فعلی و معنای واژه‌ی اسمی تعریف می‌شوند. یکی دیگر از اطلاعاتی که در قسمت معنای واژه ثبت می‌گردد سبک واژه است. یک واژه ممکن است، به ازای برخی از معانی خود قدیمی و نامتداول باشد و به ازای معنای دیگری به عنوان یک کلمه‌ی رایج در نظر گرفته شود. فرهنگستان زبان فارسی نیز برای معادل‌سازی برخی از کلمات بیگانه از این روش استفاده می‌کند. به عنوان مثال، کلمه‌ی هم‌نشست که در گذشته به معنای هم‌نشین بوده است، امروزه به عنوان معادل سمپوزیوم انتخاب شده است.

هر معنای واژه دقیقاً در یک گروه هم‌معنایی قرار می‌گیرد و یک رابطه‌ی نوع‌دار میان معنای واژه و گروه هم‌معنایی برقرار می‌گردد. یکی از اصول تعریف شده در این پایگاه شناخت این است، که معنای واژه اسمی باید در گروه هم‌معنایی اسمی قرار بگیرد و به همین ترتیب معناهای واژه‌ی فعلی، صفتی و قیدی نیز باید در گروه هم‌معنایی متناظر قرار بگیرند. یکی از اصول دیگر این است که برای برقراری رابطه‌ی بین معنای واژه و گروه هم‌معنایی دو رابطه معکوس یکدیگر تعریف شده‌اند؛ یعنی یک

معنای واژه با یک رابطه‌ی نوع‌دار در گروه هم‌معنایی عضویت دارد و از طرف دیگر گروه هم‌معنایی طی یک رابطه شامل آن معنای واژه است.

کلاس گروه هم‌معنایی

عضویت چند معنای واژه در یک گروه هم‌معنایی نشان‌دهنده‌ی شباهت معنایی زیاد آن واژه‌هاست بنابراین بین آنها به صورت ضمنی رابطه‌ی هم‌معنایی برقرار می‌گردد. همه‌ی مصداق‌های گروه‌های هم‌معنایی شامل شرح و مثال هستند. برخی از روابط تنها میان نمونه‌های گروه‌های هم‌معنایی یک مقوله‌ی خاص برقرارند. بر این اساس در هر دسته از گروه‌های هم‌معنایی به روابط خاص آن مقوله پرداخته می‌شود. مثلاً برای نمونه‌های گروه‌های هم‌معنایی اسمی، روابط شمول (ابرنوعی و زیرنوعی) و روابط عضویت (جزنامی و کل‌نامی) و رابطه‌ی نمونه‌ای برقرار می‌گردد؛ در بین نمونه‌های گروه‌های هم‌معنایی فعلی، روابط علت و معلولی، و در بین گروه‌های هم‌معنایی صفات روابط تشابه و تضاد اهمیت دارد. در ساختار شبکه واژگان دو رابطه‌ی «در نتیجه‌ی» و «نتیجه می‌دهد» برای روابط علت و معلولی استفاده می‌شود.

رابطه‌ی تضاد و تشابه میان گروه‌های هم‌معنایی صفات به این صورت است، که صفاتی که در دو قطب مخالف وجود دارند، با یکدیگرند متضاد هستند؛ و سایر صفات به صورت اقماری با صفات قطبی رابطه‌ی تشابه دارند. به عنوان مثال صفت‌های تر و خشک با یکدیگر متضاد هستند اما صفاتی مانند نیمه‌تر و مرطوب مشابه با صفت تر در نظر گرفته می‌شوند.

پایگاه شناخت طراحی شده در روابط سلسله‌مراتبی بین گروه‌های معنایی که به یک مفهوم عمومی اشاره می‌کنند، و گروه‌های معنایی که به یک مفهوم خاص اشاره می‌کنند، تمایز قائل می‌شود. رابطه‌ی نوعی–از (Is-a) زمانی میان دو گروه هم‌معنایی برقرار می‌گردد که هر دو مفاهیم عمومی باشند، اما یک مفهوم کلی‌تر از مفهوم دیگر باشد. به عنوان مثال کلان‌شهر نوعی–از شهر است، اما آبادان نمونه‌ای–از شهر می‌باشد.

از اصول دیگر این پایگاه شناخت این است که کلیه‌ی روابط سلسله مراتبی، تراگذری هستند. به عنوان مثال زمانی که گفته می‌شود؛ تایر بخشی–از چرخ است و چرخ بخشی–از ماشین است؛ تایر نیز به عنوان بخشی– از ماشین قابل استنتاج است. همچنین اگر بگوییم؛ اصفهان نمونه‌ای–از کلان‌شهر است و کلان‌شهر نوعی–از شهر باشد، اصفهان به عنوان نمونه–ای از شهر نیز قابل استنتاج است.

پس از انتخاب هر کلاس می‌توان اطلاعات مربوط به آن را در پنجره‌ای مشابه شکل ٤ مشاهده کرد. Display Name نام کلاس است که در سلسله‌مراتب درختی نمایش داده می‌شود.

همه مداخل در شبکه واژگان دارای یک شناسه‌ی یکتا به نام آی‌آرآی[17] هستند. این شناسه مبتنی بر قواعد پایگاه شناخت، مشابه با آدرس اینترنتی و به شکل زیر برای هر یک از مداخل

[17] **IRI (Internationalized Resource Identifier)**

ایجاد شده است: نام مدخل + آدرس پایه. آدرس پایه عبارت است از: http://www.mitrc.ir/mobina# و نام هر مدخل همانگونه که در بالا برای هر یک از مداخل واژه، معنای واژه و گروه معنایی توضیح داده شده، ایجاد می‌شود.

در شکل ٤ آی‌آرآی مربوط به کلاس واژه عبارت است از: http://www.mitrc.ir/mobina#Word به قسمت ابتدای این شناسه تا علامت مربع آدرس پایه[18] و به بقیه‌ی آن نام محلی[19] گفته می‌شود.

در ابزار استفاده‌شده، با استفاده از مربع‌هایی که روی آنها علامت x وجود دارد، می‌توان هر یک از اطلاعات وارد شده را حذف نمود.

در قسمت Annotation توضیحات این کلاس آمده است. قرار دادن توضیحات برای کلاس‌ها، نمونه‌ها و سایر عناصر هستان‌شناسی اجباری نیست، اما راهنمای خوبی برای کسانی است که ممکن است، در آینده بخواهند از این هستان‌شناسی استفاده کنند.

در قسمت بعد یک قاعده از نوع محدودیت[20] برای کلاس واژه تعریف شده است، که از اصول پایگاه شناخت شمرده می‌شود، این قاعده این است که اگر نمونه‌ای از مدخل واژه در طرف اول رابطه‌ی معنی می‌دهد قرار بگیرد؛ در طرف دوم آن باید نمونه‌ای از کلاس معنای واژه (یا فرزندان آن) قرار بگیرد. (رابطه‌ی معنی می‌دهد در بخش ۱-۳ توضیح داده شده است.)

در ادامه توضیح می‌دهیم که نمونه‌های هر یک از کلاس‌ها (یا همان مداخل) چه هستند و چه اطلاعاتی در آن‌ها جای می‌گیرد.

اطلاعات مدخل واژه

در مدخل واژه، اطلاعات مربوط به موارد زیر با رعایت نکاتی که در ذیل هر یک می‌آید، وارد می‌شوند:

الف) شکل نوشتاری

هر واژه یک شکل نوشتاری استاندارد دارد و ممکن است چند شکل نوشتاری دیگر نیز داشته باشد. شکل نوشتاری استاندارد در قسمت نام مدخل (rdfs:label) و سایر اشکال نوشتاری در قسمت نویسه‌ی دیگر نوشته می‌شوند.

مثال ۱:

- شکل نوشتاری استاندارد: میکروسکوپ
- نویسه‌ی دیگر: میکرسکوپ
- نویسه‌ی دیگر: میکروسکپ
- نویسه‌ی دیگر: میکرسکپ

مثال ۲:

- شکل نوشتاری استاندارد: جابه‌جایی
- نویسه‌ی دیگر: جابجایی

استفاده از کاراکتر فاصله به جای کاراکتر نیم‌فاصله به عنوان نویسه‌ی دیگر نوشته نمی‌شود زیرا این مسئله به صورت نرم‌افزاری قابل محاسبه است و این اطلاعات، افزونه[21] تلقی می‌شوند. مثلاً واژه‌ی «ساختمان‌سازی» نویسه‌ی دیگری ندارد اما به صورت نرم‌افزاری می‌توان واژه‌های ساختمان سازی و ساختمانسازی را به عنوان نویسه‌های دیگر آن تشخیص داد.

در شکلِ استانداردِ نوشتاری، اصولِ جدانویسی یا پیوسته‌نویسی بر اساس نظرات فرهنگستان رعایت می‌شود. به عنوان مثال، زمانی که دو بخش اسمی، یا یک بخش اسمی و یک بخش فعلی، با هم ترکیب شوند و واژه‌ی جدید بسازند؛ اگر حرف آخر واژه‌ی اول در پیوسته‌نویسی به حرف اول واژه‌ی دوم بچسبد از نیم‌فاصله استفاده می‌شود، در غیر این صورت به صورت پیوسته نوشته می‌شود. به عنوان مثال، در واژه‌های ساختمان‌سازی و لوله‌کشی از نیم‌فاصله استفاده شده است؛ اما در اتوکشی و خودروسازی از نیم‌فاصله استفاده نمی‌شود.

زمانی که چند اسم به صورت ترکیب اضافی، یک مدخل را می‌سازند؛ (یعنی بین آنها کسره‌ی غیرنوشتاری وجود دارد) و زمانی که یک فعل شامل بخش اسمی است؛ از کاراکتر فاصله استفاده می‌شود: به عنوان مثال، مجلس شورای اسلامی- قانون اساسی- اتو کردن- انصراف دادن.

اگر چند واژه مختلف یک شکل نوشتاری یکسان و شکل گفتاری مختلف داشته باشند، برای آنکه نام مداخل یکی نباشد؛ از قالب زیر برای نام مدخل استفاده می‌شود.

<شکل نوشتاری>_<Ava>

در قسمت بعد در مورد نحوه‌ی ورود اطلاعات آوا توضیح داده شده است. به عنوان مثال، شکل نوشتاری «خلق» دارای دو مدخل واژه‌ای زیر است:

خلق_xalq

خلق_xolq

[18] **base address**
[19] **local name**
[20] **Restriction**
[21] **redundant**

شکل ۵. مثالی از یک نمونه از کلاس واژه

علاوه بر یکی از کلاس‌های بالا یک نمونه می‌تواند عضو کلاس موقت نیز باشد. که پس از اتمام مراحل توسعه آن کلاس موقت حذف می‌شود و مشکلی پیش نمی‌آید. بنابراین یک نمونه ممکن است عضو بیش از یک کلاس باشد. شکل نوشتاری استاندارد در قسمت rdfs:label وارد شده است. در مورد سایر اطلاعات در ادامه توضیحاتی می‌آید.

جدول ۱علائم متناظر با صامت‌ها و مصوت‌ها برای نوشتن شکل گفتاری

صامت یا مصوت	علامت نوشتاری	صامت یا مصوت	علامت نوشتاری	صامت یا مصوت	علامت نوشتاری
فتحه	a	ج	J	غ-ق	q
کسره	e	چ	C	ف	f
ضمه	o	ح-ه	H	ک	k
آ	A	خ	X	گ	g

ب) شکل گفتاری

شکل گفتاری کلمه در قسمت آوا با حروف لاتین نوشته می‌شود. در جدول ۱ حروف لاتین متناظر با هر یک از صامت‌ها و مصوت‌های زبان فارسی آورده شده است:

او	u	د	D	ل	l
ای	i	ذ-ز-ض-ظ	z	م	m
ب	b	ر	r	ن	n
پ	p	ژ	Z	و	v
ت-ط	t	ش	S	ی	y
ث-س-ص	s	ع-ء	‘		

به عنوان مثال معتاد: mo'tAd – ایران: irAn‘– یخ: yax– موسیقی: musiqi – پله: pelle– افغانستان: afqAnestAn‘ – پژمرده شدن: paZmorde Sodan

– اگر یک واژه دو شکل گفتاری داشته باشد، ترجیح برآن است که شکل رایج آن وارد شود هر چند محدودیتی برای آنکه بیش از یک شکل گفتاری وارد شود در نظر گرفته نشده است و می‌توان همه اشکال گفتاری را وارد نمود.

ج) اطلاعات تکرار واژه

در بسیاری از کاربردها اطلاع از پرکاربرد بودن یک واژه دارای اهمیت بسیاری است. میزان تکرار واژه‌ها با استفاده از یک پیکره‌ی خبری به صورت خودکار محاسبه شده‌اند. در صورت نیاز به به‌روزرسانی این بخش بایستی از یک پردازشگر قوی برای بخش‌بندی واژه‌ها استفاده نمود، و با شمردن تعداد تکرار هر بخش آن را در قسمت frequency وارد نمود.

د) اطلاعات پیوند به معنای واژه

هر واژه معانی مختلفی دارد به عنوان مثال واژه خلق_xalq دارای دو معنی است (اولی آفرینش و دومی همگان، عموم مردم) گاهی ممکن است این معانی مقوله‌های نحوی مختلفی داشته باشند. اما ما برای آن‌ها تنها یک مدخل واژه در نظر می‌گیریم و به ازای هر معنی به مدخل معنای واژه مرتبط پیوند داده می‌شود.[۲۲] نام این رابطه <<معنی می‌دهد>> می‌باشد.

اطلاعات مربوط به قسمت‌های الف، ب و ج ویژگی‌های یک مدخل بودند و از نوع **ویژگی**[۲۳] بودند که یک مقدار مشخص می‌گرفتند، اما اطلاعات پیوند به معنای واژه در قالب یک **رابطه‌ی** بین دو مدخل است و از جنس رابطه [۲۴] است.

ه) اطلاعات پیوند به نسخه قبلی

اطلاعات مربوط به نسخه، برای هر مدخل به وسیله‌ی ویژگی‌های حاشیه‌نویسی[۲۵] ثبت می‌گردد. برای این منظور ویژگی «شناسه واژه در فارس‌نت» در نظر گرفته شده است و در صورتی که واژه مورد نظر در فارس‌نت ۲,۰ وجود داشته باشد، این ویژگی با شناسه آن واژه[۲۶] مقدار گرفته است. از آنجا که برخی واژه‌ها به دلیل شکل‌های نوشتاری مختلف (به دلیل اشتباه) یا به دلیل مقوله‌های نحوی متفاوت (به دلیل طراحی) در فارس‌نت تکرار شده بودند؛ تعداد ۲۱۱۹ واژه در شبکه واژگان اکنون بیش از یک شناسه فارس‌نت دارند و ۱۱٤ مورد از آنها ۳ شناسه‌ی فارس‌نت دارند.

اطلاعات مدخل معنای واژه

در مدخل مربوط به معنای واژه نامگذاری مداخل معنای واژه به صورت زیر است:

<نام واژه>-n

به عنوان مثال واژه‌ی شیرینی دارای دو معنا در مقوله نحوی اسم است. در یکی از معناها نام یک خوراکی و در معنای دیگر نام یک طعم است. نام‌گذاری معانی واژه به صورت زیر هستند:

- شیرینی-۰
- شیرینی-۱

[۲۲] در نسخه‌ی فارس‌نت ۲,۰ واژه‌ها به تفکیک آن که در چه مقوله‌های نحوی باشند جداسازی شده‌اند؛ بنابراین واژه‌هایی که در مقوله‌های نحوی مختلف باشند، تکرار می‌شوند. اما در نسخه ارائه‌شده تفکیک از معنای واژه انجام می‌پذیرد، بنابراین از این تکرار غیرضروری اجتناب می‌گردد.

[۲۳] **DataProperty**

[۲٤] **ObjectProperty**

[۲۵] **annotationProperty**

[۲٦] **WID**

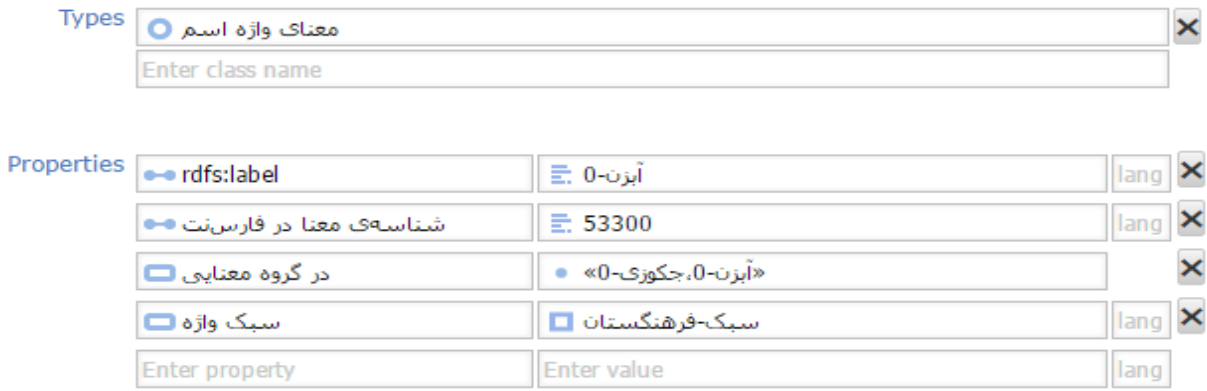
شکل ۶. مثالی از یک نمونه از معنای واژه

الف) اطلاعات مربوط به سبک واژه

از نظر ژانر متنی یک واژه ممکن است قدیمی، ادبی و یا گفتاری باشد. این مسئله بایستی در معنای واژه مشخص شود. سبک واژه شش مقدار مختلف می‌گیرد: سبک-محاوره، سبک-ادبی، سبک-کنایی، سبک-قدیمی، سبک-بیگانه و سبک-فرهنگستان.

به عنوان مثال‌هایی برای هر یک از سبک‌ها در ادامه واژه‌هایی را نام می‌بریم. واژه‌های منجم‌باشی ، اجتمال و چوشیدن هر یک، معنای واژه‌ای قدیمی دارند. تلاش اصلی بر این بوده است که واژه‌هایی که در فارسی امروز کاربرد دارند، در شبکه واژگان وارد شوند اما اگر یک معنای واژه‌ی قدیمی در یک گروه قرار داده شود، با استفاده از سبک واژه و دادن مقدار سبک-قدیمی مشخص می‌کنیم که این واژه در فارسی امروز کاربرد چندانی ندارد. برخی از واژه‌ها نسبت به دیگر واژه‌های هم‌معنی خود، ادبی‌تر تلقی می‌شوند. به عنوان مثال، واژه‌ی حلاوت با یکی از معانی واژه‌ی شیرینی هم‌معناست، اما نسبت به آن ادبی‌تر است؛ در بخش معنای واژه برای آن‌ها مقدار سبک ادبی مشخص می‌گردد.

واژه‌هایی وجود دارند که تنها در ادبیات گفتاری و محاوره‌ای کاربرد دارند. این گونه معانی واژه‌ها با سبک محاوره مشخص می‌گردند. همچنین زمانی که عبارت یک معنای دور و یک معنای نزدیک داشته باشد و منظور گوینده یا نویسنده معنای دور آن باشد، سبک کنایی مشخص می‌شود. برای مثال، اختر شمردن و شاخ به شاخ شدن دارای سبک کنایی هستند.



شکل ۷مثالی برای معنای واژه‌ی اسم که توسط فرهنگستان مصوب شده است.

معنای واژه‌هایی که فرهنگستان زبان فارسی آن‌ها را تصویب نموده است، با سبک فرهنگستان مشخص می‌شوند. گاهی فرهنگستان یک واژه‌ی قدیمی را برای یک مفهوم مدرن معادل‌سازی می‌کند. به عنوان مثال واژه‌ی آبزن در گذشته به نوعی حمام شبیه وان اطلاق می‌شده است. در حال حاضر فرهنگستان واژه‌ی آبزن را واژه‌ی برابر با جکوزی مصوب کرده است. استفاده از واژه‌هایی که از فرهنگ بیگانه به زبان فارسی راه پیدا کرده‌اند، برای پاس‌داشت زبان فارسی توصیه نمی‌شود. واژه‌هایی مانند آپدیت، جکوزی و فوبیا واژه‌هایی با سبک بیگانه هستند.

ب) اطلاعات نحوی

به ازای هر یک از مقوله‌های نحوی اطلاعات خاصی در بخش معنای واژه در نظر گرفته می‌شود. شکل ۸، نمونه‌ای از اطلاعات نحوی مربوط به یک معنای واژه فعل را نشان می‌دهد. «نوع مشارکت» یک ویژگی برای معنای واژه فعل و مختص افعال چندبخشی است که مقادیر زیر را می‌گیرد:

- Propositional_verb: وقتی یک بخش از فعل حرف اضافه و بخش دیگر فعلی باشد.
- Noun_verb: وقتی یک بخش از فعل اسم و بخش دیگر فعلی باشد..
- adjective_auxiliary: وقتی یک صفت در فعل وجود داشته باشد، مثل موافق بودن.
- Moakkad: وقتی یک حرف اضافه و یک اسم در کنار بخش فعلی وجود داشته باشد، مثل از دست رفتن.
- Moteghabel:مثل مهمان کردن، سهیم کردن و شریک کردن

شکل ۸. اطلاعات نحوی برای معنای واژه فعل

ج) عضویت در گروه معنایی

هر معنای واژه در یک و فقط یک گروه معنایی از همان مقوله‌ی نحوی عضو است.

د) ارجاع به معنای واژه دیگر

سه رابطه <<مفرد>>، <<مصدر>> و <<رجوع به>> برای ارجاع معنای واژه به یک معنای واژه دیگر طراحی شده‌اند. در این صورت معنای واژه در یک گروه معنایی قرار نمی‌گیرد بلکه به یک معنای واژه‌ی دیگر ارجاع داده می‌شود:

- رابطه «مفرد»: کلماتی که به صورت جمع هستند با این رابطه به معنای واژه‌ی مفرد ارجاع داده می‌شوند. برای مثال، کلماتی مانند ذخایر- سوابق - ستارگان- صفحات و ...
- رابطه «رجوع به»: کلماتی که احتمالا یک کلمه کامل نیستند، به کلمه کامل خود ارجاع داده می‌شوند. به عنوان مثال واژه هرج به هرج و مرج- واژه ی خاطرنشان به خاطرنشان کردن و واژه‌ی خصوص به

در خصوص ارجاع داده می‌شوند. به عنوان مثالی از نوع دیگر واژه‌ی جایگه به برخی معانی جایگاه ارجاع داده شده است.

- رابطه «مصدر»: گاهی یکی از معناهای یک واژه به صورت فعلی است. برای مثال واژه‌ی گشت علاوه بر معنای واژه‌ی اسم معنای واژه فعل هم دارد،
- بنابراین به معانی واژه‌ی فعل گشتن ارجاع داده می‌شود.

اطلاعات مدخل گروه معنایی

معانی واژه‌ای که با یکدیگر هم‌معنی باشند یک گروه را تشکیل می‌دهند که به این گروه‌ها،گروه هم‌معنایی یا برای سادگی گروه معنایی می‌گوییم. چهار نوع گروه معنایی وجود دارد که عبارتند از: گروه معنایی اسم‌ها، گروه معنایی فعل‌ها، گروه معنایی صفت‌ها و گروه معنایی قیدها. در هر چهار نوع گروه معنایی اطلاعات معانی عضو آن گروه ثبت می‌شود و نام مدخل گروه معنایی از کنار هم قرار گرفتن این معانی واژه به شکل زیر به دست می‌آید:

«معنای واژه اول،معنای واژه دوم،...،معنای واژه آخر»

همه معانی واژه که یک گروه معنایی را تشکیل می‌دهند، باید از جنس یک مقوله نحوی مشترک باشند؛ برای مثال، همه معانی واژه اسم باشند، یا همه معانی واژه فعل باشند. در ادامه ابتدا مراحل و اطلاعات مشترک برای همه گروه‌های معنایی بیان می‌شود، سپس برای هر گروه معنایی اطلاعات خاص به تفکیک بیان می‌شوند.

۱. در هر گروه معنایی جدید، معانی که در این گروه عضویت دارند مشخص می‌شود و از طریق این معانی بین گروه معنایی و واژه‌ها ارتباط ایجاد می‌شود.
۲. مفهوم آن گروه معنایی توسط ویژگی *شرح* تعریف می‌شود.
۳. یک *مثال* از کاربرد واژه‌های این گروه معنایی در یک جمله ساده‌ی زبان فارسی آورده می‌شود.
۴. روابط خاص آن مقوله تکمیل می‌شوند. برای مثال، برای گروه معنایی اسمی، یک گروه معنایی ابرنوع و گروه‌های معنایی زیرنوع مشخص می‌شوند. اگر گروه معنایی یک مصداق باشد، رابطه‌ی نمونه‌ای_از برای آن تکمیل می‌شود یا در صورت نیاز روابط شمول و عضویت تکمیل می‌شوند.
۵. گروه‌های معنایی مرتبط با این گروه معنایی مشخص می‌شوند.

در ادامه اطلاعات خاص مربوط به گروه‌های معنایی اسم، فعل، صفت و قید را جداگانه مشخص می‌کنیم.

الف) اطلاعات مربوط به مدخل گروه معنایی اسم

- ابرنوعی و زیرنوعی: هر گروه معنایی اسمی یک ابرنوع از نوع گروه معنایی اسمی دارد.
- مرتبط با
- جزنامی و کل‌نامی
- شرح و مثال
- معادل انگلیسی
- حوزه: گاهی برخی از واژه‌ها در برخی از حوزه‌ها دارای معنای تخصصی هستند و یا تنها در یک حوزه خاص به کار می‌روند. در این صورت مشخص کردن حوزه در کاربردهایی مانند دسته‌بندی بسیار مفید است. به عنوان مثال گروه معنایی «خوانده-0» در حوزه‌ی گروه معنایی «حقوق-1» تعریف می‌شود. و «قانون-4» نوعی ساز زهی است و به حوزه‌ی «موسیقی-1» ارتباط داده شده است.
- شناسه گروه معنایی فارس‌نت: گروه‌های معنایی که در فارس‌نت ۲,۰ نیز وجود داشته‌اند با این شناسه به آنجا پیوند خورده‌اند.

دو نوع رابطه‌ی سلسله‌مراتبی در گروه هم‌معنایی اسمی وجود دارد: روابط زیرنوع/ ابرنوعی و روابط جزنام/ کل‌نامی. این روابط سلسله‌مراتب اصلی را در شبکه‌واژگان می‌سازند. روابط ابرنوعی برای گروه هم‌معنایی فعل نیز معنی‌دار است و حالات خاص انجام یک کار زیرنوعی از حالت کلی انجام آن هستند. برای مثال چرت زدن نوع خاصی از خوابیدن و یا زمزمه

کردن نوع خاصی از سخن گفتن است. به عنوان یک اصل در پایگاه شناخت تمامی روابط عضویت دو طرفه هستند و در همه حالات یک رابطه‌ی کل‌نامی عکس یک رابطه‌ی جزنامی تعریف می‌شود. روابط عضویت (جزنامی و کل‌نامی) در گروه هم‌معنایی اسمی شامل سه نوع رابطه‌ی زیر است:

- رابطه‌ی عضوی از: در رابطه‌ی «عضوی–از» اعضای مشابه یک کل را می‌سازند. برای مثال چند انسان عضو یک گروه هستند، چند ناو عضو یک ناوگان هستند یا تعدادی گوسفند عضوی از یک گله هستند.
- رابطه‌ی بخشی از: در رابطه‌ی «بخشی–از»، بخش‌های مختلف یک کل را می‌سازند. برای مثال یقه، آستین، جیب، دکمه و.... بخش‌هایی از پیراهن هستند. انگشت بخشی از دست است. دست (از نوک انگشت تا مچ)، مچ، ساق، آرنج و بازو بخش‌هایی از دست هستند.
- رابطه‌ی سهمی از: در رابطه‌ی «سهمی–از» قسمتی از یک کل مدنظر است که نه از نظر ماهیت بلکه از نظر مقدار با آن کلیت تفاوت دارد. برای مثال یک تکه از یک کیک سهمی از یک کیک است و یک دانگ از خانه سهمی از خانه است. در زبان فارسی واژه‌هایی که برای سهمی از یک کلیت، یک نام جدید بسازند، اندک هستند و معمولاً این مفاهیم با توضیح بیان می‌شوند.

ب) اطلاعات مربوط به مدخل گروه معنایی فعل

- ابرنوعی و زیرنوعی: یک گروه معنایی فعلی ممکن است و یک ابرنوع از نوع گروه معنایی فعلی داشته باشد.
- مرتبط با.
- شرح و مثال.
- معادل انگلیسی.
- حوزه (مشابه گروه معنایی اسمی).
- شناسه گروه معنایی فارس‌نت (مشابه گروه معنایی اسمی).

ج) اطلاعات مربوط به مدخل گروه معنایی صفت

- متضاد با: صفاتی که در دو سر یک طیف قرار دارند. به عنوان مثال «بدنیت–0، سیاه‌دل–0، تیره‌دل–0، بدخواه–1» با گروه معنایی «دولت‌خواه–0، نیک‌خواه–0، خیراندیش–0، خیرخواه–0، خوش‌نیت–0» متضاد هستند.
- مشابه با: صفاتی که هر چند از نظر معنایی متفاوت هستند، اما به یکی از سرهای طیف نزدیک‌ترند: به عنوان مثال گروه معنایی «کودک‌کش–0» با گروه معنایی «سفاک–0،سنگ‌دل–0،بی‌رحم–0،شقی–0،قسی القلب–0،خونریز–0» مشابه است.
- ویژگی برای
- شرح و مثال (مشابه گروه معنایی اسمی)
- معادل انگلیسی (مشابه گروه معنایی اسمی)
- حوزه (مشابه گروه معنایی اسمی)
- شناسه گروه معنایی فارس‌نت (مشابه گروه معنایی اسمی)

د) اطلاعات مربوط به مدخل گروه معنایی قید

- شرح و مثال (مشابه گروه معنایی اسمی)
- معادل انگلیسی (مشابه گروه معنایی اسمی)
- حوزه (مشابه گروه معنایی اسمی)
- شناسه گروه معنایی فارس‌نت (مشابه گروه معنایی اسمی)

در جدول ۲۲ تمامی روابط و ویژگی‌های پایگاه شناخت طراحی شده فهرست شده‌اند. دامنه و برد روابط نیز مشخص شده است

جدول ۲روابط تعریف‌شده در پایگاه شناخت طراحی شده برای شبکه‌ی واژگان

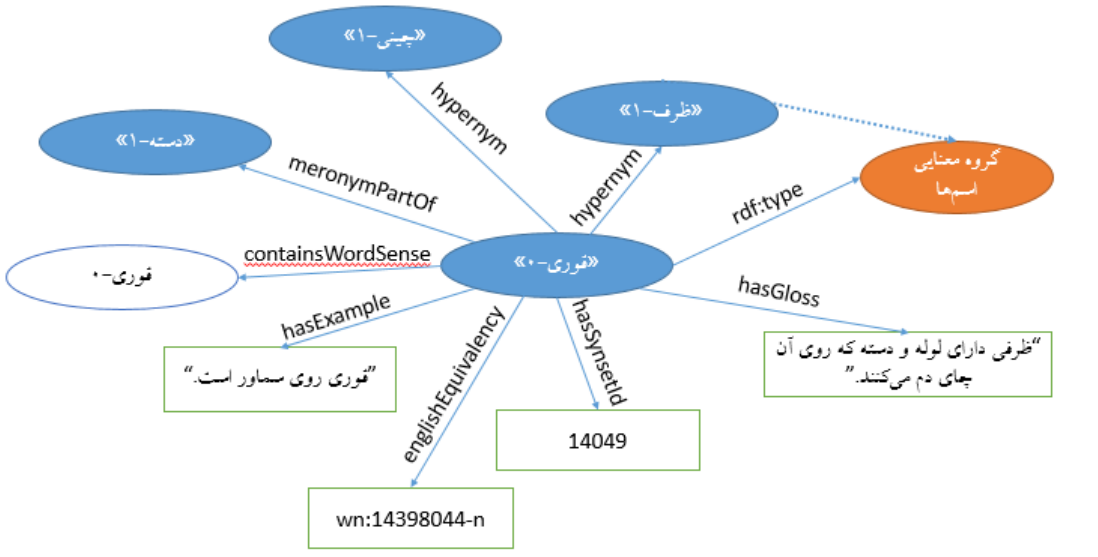
نام رابطه یا ویژگی	دامنه	برد
شرح	گروه هم‌معنایی	رشته کاراکتری
مثال	گروه هم‌معنایی	رشته کاراکتری
شامل معنای	گروه هم‌معنایی	معنای واژه
مرتبط با	گروه هم‌معنایی	گروه هم‌معنایی
برابر با	گروه هم‌معنایی	URI وردنت پرینستون
تقریبا برابر با	گروه هم‌معنایی	URI وردنت پرینستون
معنی می‌دهد	واژه	معنای واژه
در گروه هم‌معنایی	معنای واژه	گروه هم‌معنایی
شامل معنای	گروه هم‌معنایی	معنای واژه
آوا	واژه	کاراکترهای انگلیسی
معنی می‌دهد	واژه	معنای واژه
مقوله نحوی واژه	واژه	{اسم، فعل، صفت، قید}
نویسه‌ی دیگر	واژه	رشته کاراکتری
شمارش‌پذیری	معنای واژه اسم	{true،false }
مفرد واژه	معنای واژه اسم	معنای واژه اسم
گذرا	معنای واژه فعل	{true،false }
بن ماضی	معنای واژه فعل	رشته کاراکتر
بن مضارع	معنای واژه فعل	رشته کاراکتر
نوع مشارکت	معنای واژه فعل	{noun_verb، ropositional_verb و....}
نوع صفت	معنای واژه صفت	{ساده، مرکب}
نوع قید	معنای واژه قید	{قید جمله، ....}
مشتق از	معنای واژه	معنای واژه
در گروه هم‌معنایی	معنای واژه	گروه هم‌معنایی
ابرنوع	گروه هم‌معنایی اسم (فعل)	گروه هم‌معنایی اسم(فعل)
زیرنوع	گروه هم‌معنایی اسم(فعل)	گروه هم‌معنایی اسم(فعل)
کل نام – عضوی از	گروه هم‌معنایی اسم‌ها	گروه هم‌معنایی اسم‌ها
جز نام – عضوی از	گروه هم‌معنایی اسم‌ها	گروه هم‌معنایی اسم‌ها
کل نام – بخشی از	گروه هم‌معنایی اسم‌ها	گروه هم‌معنایی اسم‌ها
جز نام – بخشی از	گروه هم‌معنایی اسم‌ها	گروه هم‌معنایی اسم‌ها
کل نام – سهمی از	گروه هم‌معنایی اسم‌ها	گروه هم‌معنایی اسم‌ها
جز نام – سهمی از	گروه هم‌معنایی اسم‌ها	گروه هم‌معنایی اسم‌ها
استلزام	گروه هم‌معنایی فعل‌ها	گروه هم‌معنایی فعل‌ها
نتیجه می‌دهد	گروه هم‌معنایی فعل‌ها	گروه هم‌معنایی فعل‌ها
متضاد با	گروه هم‌معنایی صفت‌ها	گروه هم‌معنایی صفت‌ها
مشابه با	گروه هم‌معنایی صفت‌ها	گروه هم‌معنایی صفت‌ها
ویژگی برای	گروه هم‌معنایی صفت‌ها	گروه هم‌معنایی اسم‌ها
حوزه	گروه هم‌معنایی	گروه هم‌معنایی
سبک واژه	معنای واژه	{قدیمی، ادبی، فرهنگستان، بیگانه و...}

همه اطلاعات مربوط به پایگاه شناخت در یک فایل واحد با قالب اکس‌ام‌ال و منطبق با استاندارد آردی‌اف ذخیره می‌شود. این فایل حتی به صورت دستی نیز قابل ویرایش است. برای منظم‌شدن فایل سه تایی‌هایی مربوط به یک نمونه در کنار هم قرار می‌گیرند و طرف راست رابطه تنها یک بار تکرار می‌شود. شکل ۹. (الف) بخشی از فایل اکس‌ام‌ال/آردی‌اف مربوط به یک گروه هم‌معنایی را نشان می‌دهد. در قسمت (ب) گراف این بخش از اطلاعات به تصویر کشیده شده است.

```
<!-- http://www.mitrc.ir/mobina#«0-قوری» -->

<owl:NamedIndividual rdf:about="http://www.mitrc.ir/mobina#«0-قوری»">
    <rdf:type rdf:resource="http://www.mitrc.ir/mobina#گروه_معنایی_اسم_ها"/>
    <Hypernym rdf:resource="http://www.mitrc.ir/mobina#«1-ظرف»"/>
    <Hypernym rdf:resource="http://www.mitrc.ir/mobina#«1-چینی»"/>
    <MeronymPartOf rdf:resource="http://www.mitrc.ir/mobina#«8-دسته»"/>
    <containWordSense rdf:resource="http://www.mitrc.ir/mobina#0-قوری"/>
    <englishEquivalency rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
            http://wordnet-rdf.princeton.edu/wn2/14398044-n</englishEquivalency>
    <hasExample>قوری روی سماور است</hasExample>
    <hasGloss>ظرفی دارای لوله و دسته که در آن چای دم میکنند</hasGloss>
    <www:hasSynsetId>14049</www:hasSynsetId>
    <rdfs:label>«0-قوری»</rdfs:label>
</owl:NamedIndividual>
```

(الف)



(ب)

شکل ۹. یک زیرگراف از شبکه واژگان در قالب **آردی‌اف** که بخشی از اطلاعات گروه معنایی «قوری-۰» را نشان می‌دهد. (الف) گراف اکس‌ام‌ال/آردی‌اف. (ب) بصری‌سازی گراف

٤. انتقال اطلاعات به شمای طراحی شده و توسعه‌ی پایگاه شناخت

برای انتقال اطلاعات از فارس‌نت ۲,۰ به پایگاه شناخت طراحی شده روال‌های خودکاری در زبان جاوا و با استفاده از کتابخانه جنا[۲۷] پیاده‌سازی شدند و به وسیله‌ی آن‌ها تمام محتوا انتقال داده شد. بعد از آن از ابزار وب‌پروتژ[۲۸] برای ویرایش و توسعه‌ی پایگاه شناخت استفاده گردید. ورودی روال‌های خودکار برای انتقال محتوا پنج فایل اکس‌ام‌ال[۲۹] بوده است، این فایل‌ها عبارتند از: فایل words.xml شامل مداخل واژه و معنای واژه؛ فایل Synset.xml شامل مداخل گروه هم‌معنایی؛ فایل sensesRelations.xml شامل ارتباط بین معانی واژه‌ها و گروه‌های معنایی (یعنی ارتباطات بین دو فایل اول)؛ فایل گروه هم‌معنایی Relations.xml شامل ارتباطات بین گروه هم‌معنایی در فارس‌نت و همچنین نگاشت آن‌ها به شاخص‌های وردنت

[۲۷] **jena**

[۲۸] **WebProtege**

[۲۹] **XML**

پرینستون و فایل verbFrames.xml شامل ۵۶۹ قاب برای ۵۶۹ فعل ساده زبان فارسی.

انجام روال‌های خودکار به این صورت بوده است که ابتدا برای این فایل‌های اکس‌ام‌ال فایل‌های تعریف نوع داده[۳۰] تهیه شد و پارسرهایی برای خواندن محتوای فایل‌های موردنظر طراحی و پیاده‌سازی شدند. ساختار پایگاه شناخت مقصد با توجه بر اساس شمای طراحی شده و مدیریت کردن محدودیت‌های فایل‌های مبدأ طراحی شد. اطلاعات مورد نظر با بهبودها و تکمیل تدریجی ساختار پایگاه شناخت هدف به فرمت اودبلیوال انتقال داده شدند. سپس با مقایسه‌های پیوسته، فایل‌های اکس‌ام‌ال و پایگاه شناخت تولیدشده به لحاظ کمیت و کیفیت تطبیق داده شدند. پس از اطمینان از انتقال محتواهای اصلی کار روی پایگاه شناخت تبدیل شده آغاز شد.

وجود برخی تفاوت‌های ساختاری در پایگاه شناخت طراحی‌شده و محتواهای موجود باعث لزوم ایجاد تغییراتی شد که نیاز به ویرایش‌هایی داشتند. این روابط عبارت بودند از رابطه‌ی نویسه‌ی دیگر در مدخل واژه، روابط میان کلمات مفرد و جمع مکسر و رابطه‌ی نمونه‌ای-از. همچنین یک تفاوت ساختاری دیگر در مدخل واژه وجود داشت؛ این که در مدل‌سازی شبکه واژگان در نسخه قبلی و به وسیله پایگاه داده به ازای هر مقوله‌ی نحوی جدول جداگانه‌ای وجود دارد و به همین علت اگر یک کلمه در دو مقوله نحوی وجود داشته باشد دو مدخل یا شناسه‌ی واژه دارد. این نحوه تعریف مدخل در فارس‌نت و وردنت پرینستون که از پایگاه داده استفاده می‌کردند برقرار است. اما در قالب پایگاه شناخت واژه یک مدخل دارد و معانی مختلف در مقوله‌های نحوی مختلف از طریق برقراری رابطه مدل‌سازی می‌شوند.

پس از اعمال تغییرات مورد نظر در فرایند توسعه حدود ۱۳ هزار واژه‌ی پرکاربرد جدید نیز به پایگاه شناخت اضافه شده است. در بخش بعدی در این زمینه توضیحاتی ارائه می‌گردد.

.توسعه و ارزیابی شبکه واژگان

بخش‌هایی مانند تهیه داده‌ی ورودی، انجام برخی ارزیابی‌ها و تولید روابط پیشنهادی به صورت خودکار صورت می‌گیرد و وارد کردن اطلاعات نهایی توسط شخص خبره انجام می‌پذیرد. به عنوان مثال در بخش شبکه‌ی واژگان تهیه‌ی داده‌ها و کشف لغاتی که در فارس نت وجود ندارد، واژه‌های زبان فارسی که در متون خبری مورد استفاده قرار می‌گیرند، در فارس نت مورد سؤال قرار می‌گیرند و اگر وجود نداشته باشند به عنوان پیشنهاد برای اضافه شدن به سیستم به خبره معرفی می‌شوند. در بخش اطلاعات زبانی اطلاعاتی مانند برچسب‌های ادات سخن ممکن برای یک واژه یا روابط مرتبط_با توسط برنامه‌ی کامپیوتری به زبان‌شناس پیشنهاد داده می‌شوند.

بسیاری از خودکارسازی‌ها با کمک سایر منابع زبانی موجود مانند دی‌بی‌پدیا [18] و بیبل‌نت[19] و [20] بوده است.

دو نسخه از نرم افزار پروتژ [۳۱] (نسخه رومیزی و نسخه تحت شبکه) در انجام این پروژه مورد استفاده قرار گرفته است. تفاوت‌های عمده این دو نسخه به اختصار در این است که نسخه تحت شبکه امکان کار گروهی را در یک واسط کاربری ساده‌شده فراهم می‌کند؛ در حالی که نسخه رومیزی از نظر تعریف قواعد و طراحی هستان‌شناسی، و نیز اضافه کردن افزونه‌هایی برای بصری‌سازی و پرس‌وجو کردن با استفاده از زبان‌های پرسش از هستان‌شناسی امکانات بیشتری دارد.

برای بهبود پوشش واژگان و ارزیابی شبکه واژگان از دو پیکره‌ی خبری با حجم یک میلیون و ده میلیون خبر استفاده شد. این پیکره‌ها از خزش سایت‌های خبری و استخراج اخبار و سپس انجام پیش‌پردازش به دست آمده‌اند. تعداد ده‌هزار خبر برای انجام ارزیابی کنار گذاشته شدند و بقیه برای بهبود پوشش شبکه واژگان مورد استفاده قرار گرفتند. یافتن خودکار واژه‌های غایب این پیکره‌ها نیازمند ابزارهای پردازشی قوی برای زبان فارسی مانند واژه‌یاب، ریشه‌یاب و ... است؛ و به وسیله‌ای این ابزارها فایل‌های آماری از کلمات غایب تهیه شدند. بسیاری از کلمات غایب از تکرار بالای اشکالات فاصله‌گذاری بین کلمات

[۳۰] **Data Type Definition (DTD)**

[۳۱] هر دو نسخه این نرم افزار را می توان از سایت http://protege.stanford.edu بارگیری و استفاده نمود.

ایجاد می‌شوند. بنابراین در کارهای آتی توصیه می‌شود از یک ابزار خطایاب نیز در کنار ابزارهای پردازشی استفاده شود.

ارزیابی این پروژه از دو جنبه‌ی پوشش واژه‌های مورد استفاده در اخبار و صحت اطلاعات واردشده انجام شد. برای انجام ارزیابی از جنبه پوشش روی واژگان مورد استفاده در اخبار، ده هزار خبر بخش ارزیابی از دو پیکره‌ی خبری گفته شده استفاده شد. سپس با بررسی کلمات غایب ملاحظه شد که کلمات غایب شامل خطاهای نگارشی، اشکالات توکنایزر و یا اسامی خاص هستند.

دقت اطلاعات واردشده، بالاتر از ۹۹ درصد ارزیابی شده است؛ که به دلیل بازبینی دو مرحله‌ای توسط فرد خبره می‌باشد.

حدود ۳۰۰۰ مدخل واژه‌ی تکراری تشخیص داده شده و حذف شده‌اند و با توجه به این که هر کدام از این مداخل تکراری حداقل یک معنی تکراری داشته‌اند. تعداد ۳۰۰۰ مدخل معنای واژه‌ی تکراری نیز حذف شده‌اند. در جدول ۳ آمار مربوط به هر یک از مداخل شبکه واژگان و میزان رشد آن نسبت به نسخه قبلی آورده شده است.

آمار نهایی مربوط به گراف شبکه واژگان در جدول ٤ آورده شده است. هسته اصلی بخش اصطلاحات شبکه واژگان و گراف کامل شامل بخش اصطلاحات و بخش ادعاهای شبکه واژگان است. این آمار نشان می‌دهد که قالب داده‌های پیوندی توسعه شبکه واژگان تا تسهیل نموده است و همچنین به دلیل اعتبارسنجی‌های خودکار این داده‌ها کیفیت بالاتری دارند، زیرا می‌توان بدون صرف هزینه‌های نرم‌افزاری و با امکانات موجود بسیاری از خطاها را تشخیص داده و از وقوع آن‌ها جلوگیری کرد و یا آن‌ها را برطرف نمود.

جدول ۳.. آمار هر یک از مداخل شبکه واژگان و میزان رشد آن نسبت به نسخه قبلی

نام مدخل	تعداد در نسخه قبلی	تعداد در نسخه جدید	میزان افزایش	در صد افزایش
کلاس واژه	**۲۷۰۰۰**	**٤۸۸۵۰**	**۲۱۸۵۰**	**۱٫۸**
کلاس معنای واژه اسم	۲٦۱۰۹	۳٦۷۷۲	۱۰٦٦۳	
کلاس معنای واژه صفت	٦۰۳۳	۱۰۰٤۰	٤۰۰۷	
کلاس معنای واژه فعل	٤٤۸۰	۸٦٥٦	٤۱۷٦	
کلاس معنای واژه قید	٤۷۸	۲۳۸٥	۱۹۰۷	
کلاس معنای واژه	**۳٤۲٥۰**	**٥۷۸٥۳**	**۲۰۷٥۳**	**۱٫۷**
کلاس گروه معنایی اسم	۱۱۹٥۷	۲۱۰۷۰	۹۱۱۳	۱٫۸
کلاس گروه معنایی صفت	٤۲٦۲	٥۲۹۲	۱۰۳۰	۱٫۲
کلاس گروه معنایی فعل	۳۳۰۱	۳٥۷۱	۲۷۰	۱٫۱
کلاس گروه معنایی قید	۹۲٥	۹۹٦	۷۱	۱٫۱
کلاس گروه معنایی	**۲۰٤۳٥**	**۳۰۹۲۹**	**۱۰٤۹٤**	**۱٫٥**

جدول ٤. آمار داده‌های مربوط به گراف شبکه واژگان

مشخصات گراف کامل		مشخصات هسته اصلی مدل	
تعداد گره‌های گراف	۱۳۷۷۸۴	تعداد کلاس‌ها	۱۰
تعداد لبه‌های گراف (روابط)	۳۹۰٥۳٥	تعریف حاشیه‌نویسی	٤
تعداد ویژگی‌ها	۷۰۲۷۰	تعریف ویژگی‌ها	۱۸
تعداد نمونه‌ها	۱۳۷٥۲۳	تعریف روابط	۳۰

۵. نتیجه‌گیری

در این مقاله روشی برای ارائه‌ی شبکه‌ی واژگان به صورت داده‌های پیوندی و با فرمت پایگاه شناخت مطرح شد و فرآیند انتقال دانش شبکه‌ی واژگان فارسی به پایگاه شناخت طراحی شده و توسعه‌های انجام گرفته در شبکه‌ی واژگان جدید گزارش داده شد. این بازنمایی گامی برای پیوند دادن این منبع واژگانی به ابر داده‌های پیوندی و یکپارچه‌سازی این دانش با سایر دانش‌هاست.

استفاده از داده‌های پیوندی موجب می‌شود ابزارهایی که از شبکه واژگان استفاده می‌کنند، با رشد و توسعه‌ی آن، سازگاری خود را حفظ کنند. از سوی دیگر این قالب امکان رشد و توسعه‌ی شبکه واژگان ار تسهیل می‌کند. همچنین این تبدیل امکان استفاده از مزایای پایگاه شناخت‌ها در زمینه‌هایی مانند استنتاج و تولید دانش افزونه، اعتبارسنجی، نمایش داده‌ها، و توسعه‌ی بهتر داده‌های شبکه‌ی واژگان را فراهم آورده است.

٦. مراجع

[1] J. McCrae, C. Fellbaum and P. Cimiano, "Publishing and Linking WordNet using lemon and RDF," in *Proceedings of the 3rd Workshop on Linked Data in Linguistics*, Reykjavik, Iceland, 2014.

[2] G. A. Miller, "WordNet: a lexical database for English," in *Communications of the ACM 38.11*, 1995.

[3] C. Fellbaum, WordNet, Springer, 2010.

[4] D. S. Chaplot and R. Salakhutdinov, "Knowledge-basedWord Sense Disambiguation using Topic Models," *arXiv preprint arXiv:1801.01900 (2018).,* 2018.

[5] M. Shamsfard, A. Hesabi, H. Fadaei, N. Mansoory, A. Famian, S. Bagherbeigi, E. Fekri, M. Monshizadeh and S. Assi, "Semi Automatic Development of FarsNet; The Persian WordNet," in *Proceedings of 5th Global WordNet Conference*, Mumbai, India, 2010.

[6] M. Montazery and H. Fail, "Automatic Persian wordnet construction," in *Proceedings of the 23rd International Conference on Computational Linguistics: Posters.*, 2010.

[7] N. Taghizadeh and H. Faili, "Automatic wordnet development for low-resource languages using cross-lingual WSD," *Journal of Artificial Intelligence Research*, vol. 56, pp. 61-87, 2016.

[8] I. Niles and A. Pease, "Mapping WordNet to the SUMO ontology," in *Proceedings of the ieee international knowledge engineering conference*, 2003.

[9] A. Rahnama and A. Abdollahzadeh Barforoush, "Cognibase: a new representation model to support ontology development," in *IADIS International Conference Information Systems (IS 2011)*, Avila, Spain, 2011.

[10] F. Lin and K. Sandkuhl, "A survey of exploiting wordnet in ontology matching," in *Artificial Intelligence in Theory and Practice*, US, 2008.

[11] J. Kwak and H.-S. Yong, "Ontology Matching Based on Hypernym," *International journal of Web & Semantic Technology (IJWesT),* vol. 1, no. 2, 2010.

[12] F. M. Suchanek, G. Kasneci and G. Weikum, "Yago: A large ontology from wikipedia and wordnet.," in *Web Semantics: Science, Services and Agents on the World Wide Web*, 2008.

[13] M. Van Assem, A. Gangemi and G. Schreiber, "Conversion of WordNet to a standard RDF/OWL representation," in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy, 2006.

[14] C. Bizer, T. Heath and T. Berners-Lee, "Linked Data - The Story So Far," *Semantic Services, Interoperability and Web Applications: Emerging Concepts,* pp. 205-227, 2009.

[15] J. Gracia, D. Vila-Suero, J. P. McCrae, T. Flati, C. Baron and M. Dojchinovski, "Language Resources and Linked Data: A Practical Perspective," *Knowledge Engineering and Knowledge Management.,* pp. 3-17, 2014.

[16] M. Fiorelli, M. T. Pazienza and A. Stellato, "LIME: towards a metadata module for ontolex," in *2nd Workshop on Linked Data in Linguistics*, Pisa, Italy, 2013.

[17] J. McCrae, D. Spohr and P. Cimiano, "Linking lexical resources and ontologies on the semantic web with lemon," *The semantic web: research and applications,* pp. 245-259, 2011.

[18] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak and Z. Ives, "Dbpedia: A nucleus for a web of open data.," in *The Semantic Web*, 2007.

[19] R. Navigli and S. P. Ponzetto, "BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network," *Artificial Intelligence ,* vol. 193, pp. 217-250., 2012.

[20] C. D. Bovi and R. Navigli, "Multilingual semantic dictionaries for natural language processing: The case of BabelNet," *Encyclopedia with Semantic Computing and Robotic Intelligence,* vol. 1, no. 1, p. 1630015, 2017.

[21] P. Vossen, "Introduction to eurowordnet," in *EuroWordNet: A multilingual database with lexical semantic networks.*, Netherlands, 1998.

[22] A. Pease, C. Fellbaum and P. Vossen, "Building the global WordNet grid.," in *CIL18* , 2008.

[23] C. Chiarcos, S. Nordhoff and S. Hellmann, "Linking Localisation and Language Resources," in *Linked Data in Linguistics*, Heidelberg, Springer, 2012, pp. 161-179.

[24] L. Yu, "Linked open data.," in *A Developer's Guide to the Semantic Web, pp.* , Berlin Heidelberg, 2011.

[25] I. Horrocks, "Owl: A description logic based ontology language," *Logic Programming,* pp. 1-4, 2005.