معرفی چارچوبی برای مقایسه و ارزیابی سیستم های یادگیرهستان شناسی

مهرنوش شمس فرد[۱و۲] و احمد عبداله زاده بارفروش[۱]
[۱] آزمایشگاه سیستم های هوشمند، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر، تهران
[۲] دانشکده مهندسی برق و کامپیوتر، دانشگاه شهید بهشتی، تهران

Archive of SID

چکیده

در سالهای اخیر فعالیت هایی در زمینه یادگیری هستان شناسی انجام و روش ها، متدولوژی ها، ابزارها و سیستم های مختلفی نیز ارائه شده اند. این سیستم ها با داشتن وجوه مشترک بسیار از ابعاد مختلف با هم متفاوتند. در این مقاله با دسته بندی عوامل افتراق سیستم های یادگیر هستان شناسی به معرفی چارچوبی شامل شش بعد اصلی برای مقایسه و ارزیابی این سیستم ها می پردازیم. هدف از پیشنهاد این چارچوب ارائه راهنماییهایی برای محققین و کاربران هستان شناسی در جهت انتخاب صحیح یکی از ابزارها و یا سیستم های یادگیر موجود و یا ارائه راهکاری برای ساخت یک سیستم یادگیر با ویژگی های متناسب با نیازشان می باشد.

کلمات کلیدی: هستان شناسی، یادگیری ماشینی، مهندسی دانش، استخراج دانش، بازیابی هوشمند اطلاعات.

۱- مقدمه

یادگیری هستان شناسی به معنی استخراج دانش مفهومی از منابع ورودی و ساخت یک هستان شناسی بر اساس آنهاست. یادگیری هستان شناسی از روشها و الگوریتم های رشته های مختلفی چون پردازش زبان طبیعی، مهندسی دانش، یادگیری ماشینی، اکتساب دانش، استخراج اطلاعات، استنتاج خودکار و پردازش نمادین و احتمالاتی بهره می گیرد. در دو دهه اخیر فعالیت هایی در زمینه یادگیری هستان شناسی انجام و روش ها، متدولوژی ها ابزارها و سیستم های مختلفی نیز ارائه شده اند. برخی از این سیستمها، سیستم های خودمختار یادگیر هستان شناسی هستند در حالیکه برخی دیگر ابزارهای پشتیبانی ساخت هستان شناسی می باشند. در این مقاله به معرفی چارچوبی برای مقایسه و ارزیابی سیستم های یادگیر هستان شناسی می پردازیم. در این راستا ابتدا قالبی متشکل از شش بعد اصلی برای بررسی و مقایسه سیستم های یادگیر ارائه خواهیم نمود. سپس به بررسی چند سیستم یادگیر هستان شناسی به عنوان نمایندگان برجسته گروه خود در قالب معرفی شده پرداخته و در نهایت نتایج حاصله را به صورت راهکارهایی برای ساخت و یا استفاده از سیستم های یادگیر هستان شناسی ارائه خواهیم نمود.

www.SID.ir

۲- معرفی قالب مقایسه و ارزیابی سیستم های یادگیر

در این مقاله عوامل افتراق میان سیستم های یادگیر هستان شناسی به شش بعد اصلی و تعدادی زیربعد تقسیم شده است [۱]. ابعاد اصلی عبارتند از عنصر آموختنی (یادگرفته شده)، نقطه شروع، پیش پردازش لازم، نحوه یادگیری، نتیجه نهایی و روش ارزیابی. در ادامه این بخش در مورد هریک از این ابعاد، مقادیر ممکن برای آنها و نمونه هایی از سیستم های یادگیری که این مقادیر را دارند توضیحاتی داده خواهد شد.

۲-۱- عنصر یادگرفته شده (عنصر آموختنی)

در این بخش به این سوال پاسخ داده می شود که "چه نوع دانش مفهومی و چه ساختار مفهومی اکتساب می شود؟" عنصر آموختنی می تواند دانش مفهومی و هستان شناسانه به تنهایی و یا ترکیب آن با دانش لغوی باشد. اصلی ترین عناصر لغوی که سیستم های موجود یاد می گیرند کلمات و اصلیترین عناصر آموختنی هستان شناسی مفاهیم، روابط مفهومی و اصول بدیهی هستند. همچنین برخی سیستم ها فرادانشی در مورد نحوه استخراج دانش های فوق از ورودی را می آموزند.

الف) کلمات: اگرچه اکثر سیستم های یادگیر هستان شناسی مانند TEXT-TO-ONTO [۲] و DODDLE II [۳] از واژگان های از پیش تعریف شده استفاده می کنند ولی برخی سیستم ها مانند هستی [۴]، SYNDIKATE [۵] و WOLFIE [۶] کلمات جدید را نیز می آموزند. البته نحوه مواجهه با کلمات جدید و نوع دانش لغوی اکتساب شده در سیستم های مختلف متفاوت است. دانش استخراج شده در مورد کلمه ناشناس ممکن است ویژگی های ساختواژی (هستی و SYNDIKATE)، محدودیت های ترتیبی (SYNDIKATE)، طبقه نحوی (هستی و SYNDIKATE) و یا معنی کلمه (هستی و WOLFIE) باشد.

ب) مفاهیم: سیستم های یادگیر مفاهیم مانند هستی،SYNDIKATE و[۷] ممکن است آنها را از منابع ورودی اکتساب نمایند (اکتساب ترمینولوژیک یا مبتنی بر عبارات) و یا در حین فرآیندهای داخلی سیستم از روی مفاهیم دیگر بسازند (اکتساب مفهومی یا مبتنی بر معنا). همچنین برخی سیستم های یادگیر مفاهیم جدید را نمی آموزند و تنها از هستان شناسی های موجود استفاده و آنها را با نمونه هایی از رده ها و روابط موجود تکمیل و نمونه دهی می کنند مانند WEB→KB [۸].

ج) روابط مفهومی: روابط از دو بعد قابل بررسی اند:

۱) یک رابطه گره ای در هستان شناسی است و مانند سایر مفاهیم آموخته خواهد شد. در این بعد روابط به صورت معنایی و مستقل از اینکه چه مفاهیمی را بهم مربوط میکنند با ویژگی ها و جایگاهشان در هستان شناسی ارائه می شوند

۲) یک رابطه دو یا چند مفهوم را بهم مرتبط می کند و لذا بصورت زیرمجموعه ای از حاصلضرب n مفهوم آموخته می شود. در این بعد هر رابطه به شکل مبسوط و با با ذکر لیست عناصر مرتبط توسط آن آموخته می شود.

در برخی سیستم ها مانند هستی هردو بعد روابط آموخته می شوند در حالیکه در سایر سیستم ها خود روابط (بعد اول) از پیش تعریف شده است و سیستم تنها مفاهیم مرتبط توسط این روابط (بعد دوم) را می آموزد. همچنین روابط ممکن است به صورت درون-هستانی (رابطه میان عناصر یا ساختارهای یک هستان شناسی) و یا میان-هستانی (ارتباط میان ساختارهای دو یا چند هستان شناسی) باشند. اکثر کارهای انجام شده روابط درون-هستانی را می آموزند ولی برخی سیستم ها مانند [۹] نیز بر روی آموختن دسته دوم کار می کنند. اصلی ترین دسته بندی روابط مفهومی آنها را به روابط طبقه ای و غیر طبقه ای تقسیم می کند:

Archive of SID

www.SID.ir

ج-۱) روابط طبقه ای: این روابط با استفاده از کلی سازی و جزئی سازی طبقه بندی را ایجاد و امکان اعمال ارث بری ساده یا چندگانه را فراهم می کنند. رابطه شمول معنایی (رابطهISA) بعنوان مهمترین رابطه طبقه ای یادگرفته شده، توسط سیستم هایی چون هستی، SYNDIKATE، DODDLE II و [۱۰-۱۵] استخراج می شود.

ج-۲) روابط غیر طبقه ای: روابط مفهومی غیر طبقه ای به هرگونه رابطه مفهومی بجز روابط ISA اطلاق می شوند مانند روابط جزء-کل، تضاد، ترادف، ویژگی چیزی بودن، مالکیت، علیت و غیره. این روابط توسط سیستم هایی چون هستی، TEXT-TO-ONTO و [۱۲و۱۶] آموخته می شوند. همچنین اکتساب دانش در مورد طبقات نحوی یا نقشهای موضوعی خاص نیز در همین دسته جای دارد مانند یادگیری فاعل و مفعول افعال در [۱۷]، اکتشاف قالب افعال در ASIUM [۱۸]، استنتاج معنی افعال در [۱۹]، دسته بندی صفات [۲۰] و اسامی در SVETLAN' [۲۱] و تشخیص اسم ها [۲۲].

د) اصول بدیهی: اصول بدیهی برای مدلسازی جملاتی که همیشه درست هستند بکار می روند. از جمله کاربردهای این اصول می توان به اعمال محدودیت ها بر دانش های موجود، اعتبارسنجی آنها و استنتاج دانشهای جدید اشاره نمود. نمونه ای از یادگیری این اصول در هستی دیده می شود که اصول بدیهی صریح موجود در جملات شرطی و جملات سوردار متون ورودی را در شکل محدود بطور خودکار می آموزد.

ه) فرادانش یادگیری: در کنار سیستم هایی که دانش مفهومی را می آموزند، سیستم هایی نیز وجود دارند که نحوه استخراج این دانش را یاد می گیرند. آنها فرادانش لازم مانند قوانین استخراج نمونه ها، قواعد و فیلدهای خاص از وب (WEB→KB) یا الگوهای استخراج دانش از متن [۷و۲۳] یا قوانین همگونی در پیکره های زبانی را اکتشاف می کنند.

۲-۲- نقطه شروع

این بعد به جواب این سوال توجه دارد که " از کجا اکتساب دانش را شروع کنیم و از چه منبعی یاد بگیریم؟" سیستم های یادگیر هستان شناسی از دانش زمینه ای خود استفاده و دانش های جدید را براساس ورودیشان آموخته و یا دانشهای موجود را بهنگام می کنند. در این راستا پارامترهای زیر در سیستم های مختلف متفاوت است.

۲-۲-۱- دانش پیش زمینه (هستان شناسی مبنا)

دانش پیش زمینه مورد نیاز هم از جهت نوع و هم از جهت حجم در سیستم های مختلف متفاوت است. این دانش ممکن است از نوع دانش زبانی (دستور زبان، دانش لغوی، الگوها و ...) و یا از نوع دانش مفهومی (هستان شناسی مبنا) باشد. در برخی پروژه ها از واژگان از پیش تعریف شده ای برای پردازش متن استفاده می شود مانند استفاده از WordNet (Euro) در TEXT-TO-ONTO، DODDLE II، SYNDIKATE و [۱۲و۲۴و۲۵]. اندازه و محدوده پوشش هستان شناسی مبنا عامل متمایز کننده دیگری در میان سیستم های یادگیر است. این عامل در محدوده ای میان تقریبا تهی (هسته کوچکی از دانش اولیه) مانند هستی تا استفاده از یک هستان شناسی کلی بزرگ حاوی دانش عرفی مانند Cyc [۲۶] متغیر است. در این محدوده می توان به استفاده از هسته های کوچک هستان شناسی که توسط کاربر تعیین می شود [۲۷]، تعداد کمی کلمه اولیه که نمایشگر مفاهیم سطوح بالایی هستان شناسی هستند در InfoSleuth [۲۸] و هستان شناسی های موجود در قلمروهای خاص (مانندSYNDIKATE) اشاره نمود.

۲-۲-۲-منبع ورودی

سیستم های یادگیر هستان شناسی دانش موردنیازشان را از منابع ورودی که از نظر نوع و زبان با هم متفاوتند تامین می کنند.

Archive of SID

www.SID.ir

الف) نوع ورودی: ورودی سیستم یادگیرممکن است (۱) *داده های ساخت یافته* مانند ساختار جداول پایگاه داده، هستان شناسی های موجود دیگر [۹]، پایگاه های دانش [۱۰] و شبکه های معنایی لغوی یا واژگان های معنایی مانند WordNet، (۲) *داده های نیمه ساخت یافته* مانند فرهنگ های لغات، مستندات HTML و XML و DTD ها (مورد استفاده در TEXT-TO-ONTO, WEB→KB) و یا (۳) *داده های غیر ساخت یافته* مانند متون زبان طبیعی در پیکره های زبانی (هستی، SVETLAN'، SYNDIKATE، و [۱۳]) و یا متون موجود در وب [۲۹] باشد.

ب) زبان ورودی: ورودی ممکن است متونی از زبان های طبیعی مانند انگلیسی (DODDLE II و [۲۴و ۲۵])، فرانسه ([۱۱] و ASIUM; SVETLAN')، آلمانی (SYNDIKATE; TEXT-TO-ONTO)، فارسی (هستی) و غیره و یا داده های بازتمایی شده در زبان های مصنوعی مانند XML (TEXT-TO-ONTO) یا RDF [۱۴] باشد.

۲-۳- پیش پردازش لازم

این بعد به این سوال پاسخ می دهد که " آیا پیش پردازشی برای تبدیل ورودی به ساختار مناسبی برای شروع عمل یادگیری لازم است؟" متداولترین پیش پردازش استفاده شده در سیستم های یادگیر هستان شناسی از متن، پیش پردازش زبانی است. این پیش پردازش می تواند به صورت درک عمیق متن جهت کشف روابط خاص (SYNDIKATE و هستی) و یا پردازش سطحی آن برای استخراج روابط کلی میان مفاهیم (TEXT-TO-ONTO و InfoSleuth) باشد. از آنجا که درک عمیق معمولا مشکل تر است و سرعت ساخت هستان شناسی را کاهش می دهد لذا اکثر سیستم ها از تکنیک های پردازش سطحی متن مانند نمونه برداری، تعیین اجزاء کلام، تحلیل نحوی و غیره به عنوان پیش پردازش جهت استخراج ساختارهای موردنیازشان از متن استفاده می کنندهمچنین پیمانه هایی وجود دارند که برای استخراج ساختارهای خاص از ورودی بکار می روند. این ساختارها توسط سیستم یادگیر جهت ساخت هستان شناسی استفاده می شوند. از جمله می توان به SVETLAN' برای استخراج طبقه بندی اشاره نمود.

۲-۴- نحوه یادگیری

این بعد به این سوال پاسخ می دهد که "ازچه روش ها، تکنیک ها و الگوریتمهایی برای استخراج دانش استفاده می شود؟" روش های استخراج دانش در محدوده میان روشهای با دانش ضعیف مانند تکنیک های آماری تا روش های غنی از دانش مانند استدلال منطقی گسترده اند. در این محدوده می توان به رهیافت ها، روش ها و تکنیک های مختلف، برای استخراج دانش مفهومی و ساخت هستان شناسی به صورت با / بی نظارت و یا به صورت بر / برون خط اشاره نمود.

۲-۴-۱-رهیافت یادگیری

رهیافت یادگیری هستان شناسی می تواند آماری یا نمادین باشد. به عنوان روش های نمادین در اینجا به روش های منطقی، زبان پایه و مبتنی بر الگوها اشاره می نماییم. روش های مکاشفه ای نیز ممکن است برای تسریع و تسهیل در عملکرد هر یک از این روش ها و رهیافتها مورد استفاده قرار گیرند. همچنین برخی سیستم ها از رهیافتهای ترکیبی استفاده می کنند که در آن دو یا چند روش مختلف برای استخراج گونه های متفاوت دانش مفهومی بکار می روند. این سیستم ها از مزایای کلیه روش ها استفاده و محدودیت های هریک را با دیگری تا حدودی جبران می کنند.

- **روش های آماری** – در این روش ها تحلیل آماری بر روی داده های جمع آوری شده از ورودی انجام می شود. روش های آماری ممکن است بر روی لغات تکی و یا دسته ای از لغات در کنار هم عمل کنند. این روش ها از نظر اندازه دسته ها، تابع توزیع آنها و تحلیل آماری که بر آنها اعمال می کنند با هم متفاوتند. مدلهای مبتنی بر لغات تک یا unigram به دنباله ای که

Archive of SID

www.SID.ir

لغت در آن واقع شده توجهی ندارند و فرض می کنند حضور هر لغت در یک سند مستقل از سایر لغات واقع شده در آن سند طبقه لغت را مشخص می کند (مانند روش WEB→KB برای طبقه بندی صفحات وب). روشهای آماری دیگر اغلب به دسته های کلمات توجه دارند. ایده اصلی مشترک در این روش ها این است که مشخصه معنایی یک کلمه در توزیع آن در بافتار های مختلف منکس می شود و لذا معنی یک کلمه بر اساس کلماتی که همراه آن واقع می شوند و فرکانس این هم وقوعی ها بازنمایی می گردد. وقوع دو یا چند کلمه در یک واحد خوش- تعریف از اطلاعات (مثل یک جمله یا یک سند) هم-مکانی نامیده می شود. یادگیری بر اساس هم وقوعی و هم مکانی متداولترین روش در یادگیری آماری هستان شناسی است و در سیستم هایی چون TEXT-TO-ONTO، DODDLE II و [۱۳] بکار می رود.

- **روشهای منطقی** – در برخی سیستم ها روشهای منطقی مانند برنامه سازی منطق استقرائی (ILP)، خوشه بندی بر اساس منطق درجه اول، یادگیری قواعد منطق درجه اول (WEB→KB) و یادگیری گزاره ای [۳۰] برای استخراج دانش مفهومی از ورودی بکار می روند. روش های یادگیری مبتنی بر منطق ممکن است دانش جدید را با قیاس (هستی) و یا استقراء (WEB→KB و [۳۰]) بدست آورند و با گزاره ها، منطق درجه اول یا درجات بالاتر نمایش دهند
- **روش های زبانی** – روش های زبانی مانند تحلیل نحوی (ASIUM)، تحلیل ساختواژی-نحوی [۲۰]، تجزیه الگوهای لغوی-نحوی [۲۳]، پردازش معنایی (هستی) و درک متن (SYNDIKATE) از جمله روش های موجود برای استخراج دانش مفهومی از متون زبان طبیعی می باشند. این روش ها عموماً وابسته به زبان هستند و معمولا برای انجام پیش پردازش بر روی متن ورودی جهت استخراج ساختارهای دانش لازم برای ساخت هستان شناسی بکار می روند.
- **روش های مبتنی بر الگو** – . در روش های مبتنی بر الگو، ورودی (معمولا متن) بدنبال الگو یا کلمه کلیدی خاص که نشانگر رابطه مفهومی خاصی است جستجو می شود. این الگوها دارای انواع مختلفی هستند. آنها ممکن است عمومی و مستقل از قلمرو یا کاربرد خاص باشند مانند الگوهای مورد استفاده توسط هستی، هیرست [۳۱] و [۱۵] و یا تخصصی و در مورد قلمرو یا کاربرد خاص باشند مانند الگوهای پیشنهادی در [۳۲] برای استخراج دانش از متون مربوط به طراحی شبکه های الکتریکی. از سوی دیگر الگوها ممکن است به صورت دستی تعریف شده باشند مانند الگوهای موجود در هستی و [۱۵و۱۶] یا به طور (نیمه) خودکار کشف شوند مانند سیستم های PROMETHEE [۲۳]، AutoSlog-TG [۳۳] و CRYSTAL [۷]. همچنین الگوها می توانند برای استخراج عناصر مختلف هستان شناسی مثل روابط طبقه ای یا غیر طبقه ای و یا اصول بدیهی بکار می روند. به عنوان نمونه الگوهای لغوی- نحوی هیرست برای استخراج روابط شمول معنایی از متن و الگوهای لغوی-نحوی و معنایی پیشنهادی در هستی جهت استخراج روابط شمول معنایی، جزء-کل، نقش های موضوعی، ویژگی-مقدار، مالکیت و غیره و همچنین برای ساخت اصول بدیهی از جملات شرطی زبان بکار میروند.
- **روش های مکاشفه ای** – روش های مکاشفه ای ممکن است در کنار هریک از روش های دیگر بکار گرفته شوند. به عبارت دیگر روش های مبتنی بر مکاشفه روش های مستقلی نیستند بلکه برای پشتیبانی و تسهیل روش های دیگر بکار می روند. به عنوان نمونه هایی از کاربرد این روشها می توان به سیستم های TEXT-TO-ONTO (برای افزایش معیار توجه در روابط وابستگی زبانی)، هستی (جهت کاهش فضای فرضیات)، InfoSleuth (برای جادادن مفاهیم جدید در مکان مناسبی در هستان شناسی) و [۱۶] (برای انتخاب وابستگی کاندید) اشاره نمود. .
- **روش های ترکیبی** – اکثر سیستم هایی که بیش از یک نوع عنصر آموختنی دارند از روش های ترکیبی برای یادگیری هستان شناسی استفاده می کنند. برای مثال TEXT-TO-ONTO از ترکیبی از قوانین همگونی، تحلیل صوری مفاهیم و تکنیک های خوشه بندی استفاده می کند، WEB→KB یادگیری قواعد منطق درجه اول را با یادگیری بیزین ترکیب می

Archive of SID

www.SID.ir

کند، هستی ترکیبی از روش های منطقی، زبان پایه، مبتنی بر الگو و تحلیل معنایی را بر ورودی اعمال می کند و [۲۵] از ترکیب روش های آماری و معنایی برای خوشه بندی کلمات و مستندات بهره می برد.

۲-۴-۲-عمل یادگیری

روش های یادگیری ممکن است بر اساس تکنیک مورد استفاده و عملی که انجام میدهند گروه بندی شوند. از جمله این گروه ها می توان به روش های دسته بندی [۱۰و۳۰]، خوشه بندی (هستی و ASIUM)، یادگیری قواعد (WEB→KB و [۷])، تحلیل صوری مفاهیم (TEXT-TO-ONTO) و نمونه دهی هستان شناسی (WEB→KBو [۲۷]) اشاره نمود. این تکنیک ها ممکن است با هر یک از رهیافت های فوق اجرا شوند. مثلا سیستم ASIUM خوشه بندی مفهومی را با استفاده از تجزیه نحوی انجام می دهد در حالیکه [۲۴] برای این عمل از روش های آماری و [۲۵] از ترکیب روش های آماری و معنایی بهره می برد.

۲-۴-۳-درجه خودکارسازی

مرحله اکتساب دانش می تواند بصورت دستی، نیمه خودکار و یا خودکار انجام شود. سیستمهای غیر دستی برای اکتساب دانش مفهومی از روشها و ابزارهای خودکار (هستی و [۲۴])، نیمه خودکار (TEXT-TO-ONTO و [۱۰]) و با همکاری (هستی و ASIUM) استفاده می کنند. در سیستم های نیمه خودکار و با همکاری، نقش کاربر و میزان دخالت او در ساخت هستان شناسی متفاوت است. این دخالت می تواند به صورت پیشنهاد هستان شناسی اولیه [۲۷]، اعتبارسنجی و تغییر تصمیمات سیستم (هستی و [۲۷])، انتخاب الگوهای روابط [۱۰]، کنترل سطح تجرید و مواجهه با نویز (ASIUM)، برچسب زنی به مفاهیم جدید (هستی و ASIUM) و تعیین وزن های عناصر هستان شناسی (هستی) باشد.

۲-۵- نتیجه نهایی

این بعد به نتیجه فرآیند یادگیری توجه دارد و به این سوال پاسخ می دهد که " چه چیزی ساخته خواهد شد و ویژگی های آن چه خواهند بود؟" قدم اول برای پاسخگویی به این سوال تمایز قایل شدن میان یادگیری هستان شناسی و پشتیبانی برای ساخت هستان شناسی است. اکثر سیستم های مطالعه شده کمابیش عناصر هستان شناسی را یاد می گیرند (سیستم های خودمختار) ولی برخی از آنها نیز فقط با ایجاد محیط مناسب یا فراهم کردن اطلاعات لازم در ساختارهای مناسب، کاربر، فرد خبره یا سیستم دیگر را برای یادگیری هستان شناسی پشتیبانی می کنند. بعد حاضر برای سیستم های خودمختار که هستان شناسی می سازند ویژگی های هستان شناسی حاصله را بررسی می کند. از میان ویژگی های مختلفی که برای هستان شناسی ها معرفی شده اند، در اینجا به متمایز کننده ترین موارد در میان خروجی های سیستم های یادگیر هستان شناسی می پردازیم.

۲-۵-۱-نوع نتیجه حاصل

نوع هستان شناسی از ویژگی های متعددی مانند سطح پوشش، قلمرو و کاربرد، نوع محتویات و درجه جزئیات تشکیل شده است. سطح پوشش اشاره به عمومی (مثل Cyc) یا تخصصی بودن (مانند DODDLE II) هستان شناسی دارد. منظور از قلمرو، زمینه خاصی است که یک هستان شناسی در مورد آن تهیه شده است مانند اعمال TEXT-TO-ONTO در قلمروی بیمه و کاربرد ASIUM در قلمروی تروریسم. همچنین کاربردی که هستان شناسی را بکار می گیرد می تواند استخراج اطلاعات، پاسخگویی به سوالات، دلالی اطلاعات، موتورهای جستجو و پردازش معنایی زبان طبیعی باشند. از نظر محتوا نیز هستان شناسی ممکن است هستان شناسی بازنمایی (مثل Frame Ontology)، هستان شناسی زبان طبیعی (مثل WordNet) و یا هستان شناسی خاص یک قلمرو باشد (طبقه بندی های دیگری از انواع هستان شناسی در [۴] آمده اند).

Archive of SID

www.SID.ir

۲-۵-۲-ساختار و هم بندی

سیستم های موجود یادگیر هستان شناسی بر روی ساختارها و هم بندی (توپولوژی) های مختلف عمل می کنند. سلسله مراتب محدود مفاهیم که در آن هر گره تنها یک پدر دارد [۳۴]، سلسله مراتب هرمی که در آن هیچ دو اتصالی یکدیگر را قطع نمی کنند (ASIUM)، گراف های جهت دار (TEXT-TO-ONTO) و ترکیبی از گراف و اصول بدیهی (هستی) نمونه هایی از ساختارهای مورد استفاده هستند.

۲-۵-۳-زبان بازنمایی

دانش هستان شناسانه می تواند به زبان های بازنمایی مختلف از جمله زبان های مبتنی بر منطق، مبتنی بر قاب، شبکه های معنایی، زبان های ترکیبی و زبانهای مبتنی بر وب نمایش داده شود. از زبان های مبتنی بر منطق می توان به KIF (در هستی)، CycL (در Cyc)، منطق توصیفی [۱۱] و زبانهای خانواده KL-ONE (در SYNDIKATE) اشاره نمود. زبانهای مبتنی بر قاب مانند OKBC در InfoSleuth مورد استفاده قرار گرفته اند و شبکه های معنایی از جمله گراف های مفهومی در [۳۵] استفاده شده اند. زبانهای ترکیبی از مزایا و توانایی های چند روش مبنایی بهره می برند مانند ترکیب منطق و قاب در زبانهایی مثل Ontolingua و F-Logic، ترکیب منطق، قواعد تولید و قاب در زبان Loom و ترکیب شبکه های معنایی و قاب در زبان TMR. زبانهای مبتنی بر وب بیشتر برای تبادل هستان شناسی ها بر روی وب و اینترنت استفاده می شوند و ممکن است با زبان های دیگر ترکیب شوند مانند بسط RDF بر اساس F-Logic (در TEXT-TO-ONTO).

۲-۶- روش ارزیابی

برای ارزیابی سیستم های یادگیر هستان شناسی استفاده از دو روش (۱) ارزیابی روش یادگیری و (۲) ارزیابی نتیجه حاصله پیشنهاد شده است. از آنجا که محاسبه درستی تکنیک های مورد استفاده در یادگیری هستان شناسی کار ساده ای نیست روش اول که بر اساس اندازه گیری میزان صحت روش یادگیری است زیاد مورد توجه نیست. روشی که بیشتر مورد استفاده قرار می گیرد ارزیابی سیستم یادگیر با ارزیابی (پاره ای) هستان شناسی تولید شده است. این ارزیابی ممکن است با مقایسه دو یا چند هستان شناسی مدلسازی شده در یک قلمروی خاص صورت پذیرد و یا از طریق کاربردی که هستان شناسی تولید شده در آن مورد استفاده قرار می گیرد انجام شود. در روش اخیر هر سیستم یادگیر محیط تست خاص خود را دارند و عمل ارزیابی را با مقایسه حاصل کار با نتایج مورد نظر فرد خبره و یا با محاسبه معیارهای ارزیابی متداول در استخراج اطلاعات مثل دقت و توجه انجام می دهد.

۳- مقایسه سیستم های منتخب در چارچوب پیشنهادی

در این بخش هفت سیستم یادگیر هستان شناسی انتخاب و در چارچوب پیشنهادی مورد بررسی قرار گرفته اند. در انتخاب این سیستم ها از میان حدود پنجاه سیستم مطالعه شده، معیار جدید بودن (ارائه شده از سال ۱۹۹۷ به بعد)، متفاوت بودن در ویژگی ها و مستند بودن در نظر گرفته شده است. جدول ۱ ویژگی های سیستم های منتخب در قالب معرفی شده را به ترتیب الفبایی نشان می دهد.

Archive of SID

www.SID.ir

جدول (۱) مقایسه سیستم های یادگیر

سیستم یادگیر	عنصر آموختنی	نقطه شروع			پیش پردازش	خروجی			ارزیابی و آزمون
		دانش زمینه	نوع ورودی	زبان		نوع	توپولوژی	زبان بازنمایی	
ASIUM	قاب زیرطبقه بندی افعال + سلسله مراتب	دانش زبان شناسی	غیر ساخت یافته (پیکره زبانی)	فرانسه	تحلیل نحوی (توسط Sylex)	خاص قلمرو	سلسله مراتب هرمی (نوع خاص DAG)	قاب های زیرطبقه بندی	تست شده برای قلمروهای آشپزی، ثبت اختراع و تروریسم
DODDLE II	روابط طبقه ای	واژگان معنایی WordNet	متون قلمروی خاص	انگلیسی		سلسله مراتب (درخت)			ارزیابی تجربی در زمینه قانون
	روابط غیر طبقه ای					زوج مفاهیم، الگوهای توصیف مفاهیم			
HASTI	کلمات، مفاهیم، روابط طبقه ای و غیر طبقه ای، اصول بدیهی	تقریبا تهی (یک هسته کوچک ابتدایی)	متون غیر ساخت یافته زبان طبیعی	فارسی	تحلیل ساختواژی-نحوی (توسط PeTex)	واژگان و هستان شناسی: عمومی یا تخصصی براساس ورودی	گراف جهت دار + اصول بدیهی	زیرمجموعه ای از KIF	آزموده شده بر متون عمومی و تخصصی
SVETLAN'	کلاس اسامی		داده های ساخت یافته (قلمروی معنایی با واحدهای موضوعی از SEGCOHLEX) + ورودی غیر ساخت یافته SEGAPSITH (مقالات روزنامه)	فرانسه	تحلیل نحوی (توسط Sylex) + تحلیل موضوعی (توسط SEGAPSITH)	Structured Domains که نشانگر کلاسهای کوچک اسامی هستند	سلسله مراتب (درخت)	فرمت خاصی از structured domains: v-> r-> n_1, n_2, …	آزموده شده بر مستندات یک ماه آژانس فرانس پرس
SYNDIKATE	کلمات، مفاهیم، روابط طبقه ای و غیر طبقه ای	واژگان و هستان شناسی مبنا – عمومی و تخصصی	متون غیر ساخت یافته زبان طبیعی	آلمانی	تجزیه افزایشی وابستگی ها (توسط PARSETALK)	پایگاه دانش متن + واژگان و هستان شناسی بهنگام شده	گراف جهت دار	منطق توصیفی از خانواده KL-One	ارزیابی جداگانه اجزاء در قلمروهای تکنولوژی اطلاعات و پزشکی
TEXT-TO-ONTO	مفاهیم، روابط طبقه ای و غیر طبقه ای	پایگاه داده لغوی + هستان شناسی تخصصی	غیر ساخت یافته (زبان طبیعی، مستندات وب)، نیمه ساخت یافته (XML, DTD) و ساخت یافته (شماهای DB، هستان شناسی های دیگر)	آلمانی، XML, HTML, DTD	پردازش منابع + پردازش سطحی متن (توسط SMES)		گراف جهتدار	بسط RDF بر اساس F-Logic قابل انتقال بر DAML-OIL، OIL و …	ارزیابی تجربی با تست کردن برای قلمروها، ورودی ها و روش های یادگیری مختلف
WEB→KB	نمونه های کلاس ها و روابط	هستان شناسی که نمونه های کلاس هایش آموخته می شوند	یک هستان شناسی + مثال های یادگیری از نمونه ها	HTML		مشابه هستان شناسی ورودی			ارزیابی شده توسط یک روش اعتبارسنجی ضربدری چهارلا بر روی پایگاه دانش دانشکده های کامپیوتر
	قواعد تشخیص نمونه ها					قوانین منطق درجه اول			

Archive of SID

www.SID.ir

۴- نتیجه گیری

در این بخش به بررسی نتایج حاصل از استفاده از چارچوب پیشنهادی و بررسی سیستم های مختلف در این چارچوب پرداخته و راهکارهایی برای محققین و کاربرانی که مایل به ساخت و یا استفاده از سیستم های یادگیر هستان شناسی هستند، ارائه خواهیم نمود.

۴-۱- انتخاب مقادیر مناسب برای ابعاد قالب ارزیابی

در این بخش تفاوت ها و نقاط قوت و ضعف مقادیر مختلف انتخابی برای ابعاد شرح داده شده بررسی و راهکارهایی برای گزینش ویژگی های مناسب برای ساخت یک سیستم یادگیر هستان شناسی و یا انتخاب یکی از سیستم های موجود ارائه شده است.

برای ساخت هستان شناسی مناسب یک کاربرد (در صورتی که هستان شناسی مناسبی موجود نباشد) ابتدا باید بدانیم چه داریم (منابع موجود و در دسترس) و چه می خواهیم داشته باشیم (هستان شناسی نهایی) . سپس باید مسیر مناسبی از نقطه شروع به نتیجه مطلوب بیابیم. این مسیر می تواند استفاده از سیستم های یادگیر موجود و یا ایجاد یک سیستم یادگیر جدید باشد.

۴-۱-۱-نقطه شروع

انتخاب نقطه شروع بستگی به در دسترس بودن و لزوم دانش پیش زمینه و منابع ورودی دارد. در محیط هایی که دانش پیش زمینه در دسترس نیست، مثلا قلمروهایی که هستان شناسی تخصصی برای آنها موجود نیست یا زبانهایی که واژگان معنایی برای آنها تولید نشده، لازم است عمل ساخت هستان شناسی را از پایه انجام دهیم. ساخت هستان شناسی از پایه (با حداقل دانش اولیه) دارای این مزیت است که مشکلات هماهنگی و یکپارچه سازی منابع (هستان شناسی های) موجود را نداریم و گلوگاه اکتساب (دستی) دانش اولیه برای ایجاد پایگاه دانش زمینه را نیز با خودکارسازی آن دور زده ایم. در این سیستم آنچه را براساس کاربرد لازم دارد اکتساب و از جمع آوری زوائد و دانش های نامرتبط پرهیز می کند. بعلاوه کاهش دانش اولیه منجر به کاهش میزان وابستگی سیستم و دانش موجود در آن به عقاید سازنده آن می شود. به این ترتیب حاصل کار انعطاف پذیرتر و منطبق تر بر کاربرد و منظور خاصی است که برای آن طراحی شده است. ایراد اصلی این روش فرآیند طولانی تر اکتساب دانش و مشکلات آن در رفع ابهام ناشی از کمبود دانش اولیه است.

لزوم و نوع دانش اولیه مورد نیاز نیز براساس نوع هستان شناسی مطلوب و روش یادگیری موردنظر تعیین می شود. برای مثال ساخت هستان شناسی های عمومی و خاص منظوره نیاز به انواع متفاوتی از دانش پیش زمینه دارند: هستان شناسی های عمومی معمولا مبتنی بر منابع لغوی عمومی مثل فرهنگ های لغت ساخته می شوند در حالیکه برای ساخت هستان شناسی های خاص منظوره زبان محدودتر بوده و روش ها عمدتا مبتنی بر محاسبه توزیع رده ها هستند. نوع ورودی بستگی شدید به کاربرد موردنظر دارد. اگر بخواهیم هستان شناسی برای وب معنایی بسازیم لازم است ورودی های نیمه ساخت یافته و غیر ساخت یافته بپذیریم درحالیکه مثلا برای یادگیری دانش هستان شناسی از اطلاعات یک پایگاه داده باید قادر به پردازش داده های ساخت یافته باشیم.

Archive of SID

www.SID.ir

۴-۱-۲-هستان شناسی حاصله

کاربردهای مختلف نیاز به انواع مختلف هستان شناسی دارند. هستان شناسی ها هم از نظر محتوا و هم از نظر اعمالی که پشتیبانی می کنند با هم متفاوتند. برای مثال کاربردهای حل مسائل علمی، تخیلی یا تجاری نیازمند هستان شناسی های کوچک، باریک و عمیق با جزئیات خاص کد شده در قالب اصول بدیهی هستند در حالیکه سیستم های بازیابی اطلاعات (مثل موتورهای جستجو) ممکن است به هستان شناسی های بزرگ و سطحی که سلسله مراتبی از مفاهیم با تعداد اندکی روابط میان آنها بدون اصول بدیهی را در بر میگیرند (مثل WordNet) نیاز داشته باشند تا کارآیی جستجوی مبتنی بر کلیدواژه ها را افزایش دهند [۳۵]. کاربردهای پردازش زبان طبیعی از انواع مختلف هستان شناسی بر اساس عمق عملیات پردازش زبان و استدلالی که انجام می دهند بهره می برند. کاربردهایی که نیاز به درک عمیق زبان و انجام استنتاجات مختلف دارند مثل برنامه سازی خودکار، برخی برنامه های مشاوره و تصمیم گیری با واسط زبان طبیعی و برخی سیستم های پاسخگویی به سوالات، نیازمند هستان شناسی های عمیق، با پیش زمینه غنی و قضایای بدبهی و دارای قابلیت استدلال خودکار هستند درحالیکه برای برخی کاربردهای دیگر پردازش زبان مثل سیستم های ترجمه ماشینی تجاری که برای تولید پیش نویس سریعی از متون در زبان مقصد بکار می روند واژگان های معنایی مثل وردنت کفایت می کنند.

۴-۱-۳-فرآیند یادگیری

الف) درجه خودکارسازی - اگرچه در حال حاضر سیستم های یادگیری تمام خودکاری برای استخراج دانش هستان شناسانه موجودند ولی همه آنها در شرایط خیلی محدود و با کارآیی پائین تر نسبت به سیستم های نیمه خودکار یا همکارانه عمل می کنند. به عبارت دیگر سیستم های همکاری نتایج قابل قبول تری ارائه می کنند چراکه برخی تصمیمات در مورد تعابیر مختلف در حین فرآیند یادگیری به عهده کاربر گذاشته می شود. این برتری در مقایسه عملی میان یادگیری خودکار و همکارانه که در برخی سیستم ها (هستی و [۲۳]) انجام شده نیز مشهود است. گرچه این حقیقت عموما برای همه سیستم های یادگیری برقرار است ولی در مورد یادگیری از متون زبان طبیعی شاخص تر است. در این مورد مشکل ناشی از طبیعت مبهم زبان و نارسایی ابزارهای موجود (مثل پارسرها) می باشد. برای مثال در خوشه بندی مبتنی بر شباهت توزیع شدگی کلاس های ساخته شده دقیقا منطبق بر کلاس های معنایی نیستند و لذا نتایج ممکن است لیستی از کلماتی باشند که در حالت همکارانه کاربر تشخیص می دهد یک کلاس معنایی می سازند یا خیر و نام کلاس و اعضاء آن کدامند. خودکارسازی کامل این فرآیند نیاز به بهره گیری از روش های پیچیده ترکیبی دارد که بتوانند ابهامات را رفع و مکمل کاستی های روش های تکی باشند.

ب) روش یادگیری - در انتخاب روش یادگیری لازم است ابتدا میان رهیافت های آماری و نمادین انتخاب نماییم. روش های آماری روش های از نظر دانش ضعیف هستند در حالیکه روش های نمادین عمدتا غنی از دانش اند [۸]. لذا برای سیستم هایی که تحلیل معنایی یا عملیات استنتاجی صورت نمی گیرد (مثل سیستم های بازیابی اطلاعات) روش های آماری مناسبند. روش های آماری محاسبه پذیرتر، عمومی تر و اندازه پذیرترند و پیاده سازی آنها ساده تر و سریعتر است درحالیکه روش های نمادی دقیقتر و قابل اعتمادترند و نتایج قابل قبول تری ارائه می کنند. روش های آماری معمولا کلی هستند و می توانند برای قلمروها و یا زبانهای مختلف بکار روند ولی روش های نمادی مانند روش های زبان-پایه یا مبتنی بر الگو نیاز به تطبیق بیشتری دارند. روشهای آماری نیاز به منابع و دانش اولیه کمتری نسبت به روش های نمادی دارند، آنها از دانش پیش زمینه استفاده نمی کنند. ایراد اصلی روش های آماری این است که داده ها در آنها معمولا پراکنده اند بخصوص وقتی بخواهیم مفاهیم کلی و عمومی را از متون فنی و تخصصی یا مفاهیم تخصصی را از متون عمومی یا نامرتبط استخراج نماییم. ویژگی

Archive of SID

www.SID.ir

دیگر روش های آماری این است که معمولا به صورت برون خط اجراء می شوند و لذا سیستم های یادگیری که از این روش ها استفاده می کنند کل ورودی (مثلا کل متن تجزیه شده) را یکجا پردازش می کنند.

در روش های نمادی عمدتا یادگیری به صورت افزایشی و با بهره گیری از تکنیک های استدلالی و استفاده از دانش معنایی انجام می شود. در این روش ها معمولا عمق پردازش بیشتر از روش های آماری و وسعت و پهنای آن کمتر است. به عبارت دیگر روش های نمادی غالبا دارای نتایج دقیقتر با سطح پوشش کمتر هستند. این مسئله در سیستمهایی مثل WEB→KB که از هردو روش استفاده می کنند مشهود است.

در میان روش های نمادی، روشهای منطقی بیشتر از سایرین وابسته به دانش اند. این روش ها برای سیستم هایی که هدفشان درک عمیق، استدلال و یا استخراج فرا قواعد (قواعد استخراج دانش) است، مناسبند. دسته دیگری از روش های نمادی که نیاز به دانش اولیه دارند روش های مبتنی بر الگو هستند. این روش ها متداولترین روش های نمادی جهت استخراج روابط مفهومی از متون هستند.این روش ها گرچه کارآیی خوبی در قلمروهای خاص برای استخراج روابط خاص دارند، ولی محدود و غیر منعطف می باشند. به عبارت دیگر آنها روابط محدودی را استخراج می کنند و هزینه تطبیق آنها برای استخراج روابط جدید، در قلمروی جدید یا استخراج از متون زبان جدید بالاست.

مراجع

[1] Shamsfard M., and Barforoush, A. A.,The State of the art in Ontology Learning: A framework for Comparison, *Journal of knowledge Engineering Review*, to be appeared, 2004.

[2] Maedche, A., and Staab, S., (a), Ontology learning for the Semantic Web, *IEEE journal on Intelligent Systems, Vol. 16*, No. 2, 2001, 72-79.

[3 Yamaguchi, T., Acquiring Conceptual Relations from domain-Specific Texts, *Proceedings of the IJCAI 2001 Workshop on Ontology Learning (OL'2001)*, Seattle, USA, 2001

[۴] شمس فرد م.، *طراحی مدل یادگیر هستان شناسی: نمونه سازی در یک سیستم درک متن فارسی*، رساله دکتری، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر، تهران، ۱۳۸۱.

[5] Hahn U., Romacker, The SYNDIKATE Text Knowledge Base Generator, *Proceedings of the 1st International Conference on Human Language Technology Research,* San Diego, USA, 2001.

[6] Thompson, C. A., Mooney, R. J., Automatic Construction of Semantic Lexicons for Learning Natural Language Interfaces, *Proc. of 16th National Conf. on Artificial Intelligence (AAAI'99)*, 1999, 487-493.

[7] Soderland, S., Fisher, D., Aseltine, J., Lehnert, W., Issues in Inductive Learning of Domain-Specific Text Extraction Rules, *Proc. of the IJCAI 95 Workshop on Approaches to Learning for NLP*, 1995.

[8]Craven, M., DiPasquo, D., Freitag, D., McCallum, A., Mitchell, T., Nigam, K., Slattery, S., Learning to construct knowledge bases from the World Wide Web, *Artificial Intelligence*, 118, 2000, 69-113.

[9] Williams, A. B., Tsatsoulis, C., An Instance-based Approach for Identifying Candidate Ontology Relations within a Multi-Agent System, *Proc. of OL'2000*, Berlin, Germany, 2000.

[10] Suryanto H., Compton, P., Learning classification taxonomies from a classification knowledge based system, *Proceedings of the ECAI 2000 Workshop on Ontology Learning (OL'2000)*, 2000.

[11] Todirascu, A., de Beuvron F., Galea, D., Rousselot, F., Using Description Logics for Ontology Extraction, *Proceedings of the ECAI 2000 Workshop on Ontology Learning (OL'2000)*, 2000.

[12] Agirre, E., Ansa, O., Hovy, E., and Martínez, D., Enriching very large ontologies using the WWW, *Proceedings of the ECAI 2000 Workshop on Ontology Learning (OL'2000)*, 2000.

[13] Heyer G., Läuter, M., Quasthoff, U., Wittig, T., Wolff, C., Learning Relations using Collocations, *Proceedings of the IJCAI 2001 Workshop on Ontology Learning (OL'2001)*, Seattle, USA, 2001.

Archive of SID

www.SID.ir

[14] Delteil, A., Faron-Zucker, C., Dieng, R., Learning ontologies from RDF annotations, *Proceedings of the IJCAI 2001 Workshop on Ontology Learning (OL'2001)*, Seattle, USA, 2001
[15] Sundblad, H., Automatic Acquisition of Hyponyms and Meronyms from Question Corpora, *OLT'2002*.
[16] Gamallo, P., Gonzalez, M., Agustini, A., Lopes, G., and de Lima, V. S., Mapping Syntactic Dependencies onto Semantic Relations, *OLT'2002*, 2002.
[17] Pereira, F., Tishby N., and Lee, L., Distributional Clustering of English Words, *Proc. of the 31st annual meeting of the Association for Computational Linguistics (ACL'93)*, 183-190, 1993.
[18] Faure, D., Nedellec C., and Rouveirol, C., Acquisition of Semantic Knowledge using Machine Learning Methods: The System ASIUM, *Technical Report number ICS-TR-88-16*, Laboratoire de Recherche en Informatique, Inference and Learning Group, Universite Paris Sud, 1998.
[19] Wiemer-Hastings, P., Graesser, A., Wiemer-Hastings, K., Inferring the meaning of verbs from Context, *Proc. of the 20th Annual Conference of the Cognitive Science Society*, NJ, USA, 1998.
[20] Assadi, H., Knowledge Acquisition from Texts: Using an Automatic Clustering Method Based on Noun-Modifier Relationship, *Proc. of 35th Annual Meeting of the ACL*, Madrid, Spain, 1997.
[21] Chalendar, G., Grau, B., SVETLAN' A System to Classify Nouns in Context, *Proceedings of the ECAI 2000 Workshop on Ontology Learning (OL'2000)*, 2000
[22] Bikel, D. A., Schwartz, R., and Weischedel, R., An Algorithm that Learns What's in a Name, *Machine Learning*, 34, 1999, 211-231.
[23] Finkelstein-Landau, M., Morin, E., Extracting Semantic Relationships between Terms: Supervised vs. Unsupervised Methods, *Proceedings of International workshop on Ontological Engineering on the Global Information Infrastructure,* Dagstuhl-Castle, Germany, 1999,71-80.
[24] Wagner A., Enriching a Lexical semantic Net with selectional Preferences by Means of Statistical Corpus Analysis, *Proc. of the Workshop on Ontology Learning*, Berlin, 2000.
[25] Termier, A., Rousset, M. C., Sebag, M., Combining Statistics and Semantics for Word and Document Clustering, *Proc. of the IJCAI 2001 Workshop on Ontology Learning,* Seattle, 2001
[26] Lenat D. B., and Guha, R. V., *Building Large Knowledge Based Systems, Representation and inference in the Cyc Project*, Readings, (MA: Addison Wesley, 1990).
[27] Brewster, C., Ciravegna F., and Wilks, Y., Knowledge Acquisition for Knowledge Management, *Proceedings of the IJCAI 2001 Workshop on Ontology Learning,* USA, 2001.
[28] Hwang, C.H., Incompletely and imprecisely speaking: Using dynamic ontologies for representing and retrieving information. *Proc. of the 6th Int'l Workshop KRDB'99*, Linköping, July, 1999.
[29] Omelayenko, B., Learning of ontologies for the Web: the analysis of existent approaches, *Proceedings of the International workshop on Web Dynamics*, London, 2001.
[30] Bowers, A. F., Giraud-Carrier, C., Lloyd, J. W., Classification of individuals with Complex Structure, *Proc. of the 17th Int'l Conf. on Machine Learning*, 2000, 81-88.
[31] Hearst, M.A., Automatic Acquisition of Hyponyms from Large Text Corpora, *Proceedings of the 14th International Conference on Computational Linguistics*, Nantes, France, July, 1992.
[32] Assadi, H., Construction of a regional ontology from Text and its use within a Documentary System, *Proceedings of Formal ontologies in Information Systems (FOIS'98),* Italy, 1999.
[33] Riloff, E., Automatically Generating Extraction Patterns from Untagged Text. *Proceedings of the 13th Conference on Artificial Intelligence*, AAAI Press/ MIT Press, 1996, 1044-1049.
[34] Emde W., and Wettschereck, D., Relational instance Based Learning, *Proceedings of 13th International Conference on Machine Learning (ICML'96)*, 1996, 122-130.
[35] Sowa, J. F., *Knowledge Representation: Logical, Philosophical and Computational Foundations*, Brooks/Cole, 2000.

Archive of SID

www.SID.ir