یادگیری هستان‌شناسی از شمای XML

بتول لکزائی[1]، مهرنوش شمس‌فرد[2]،

[1] آزمایشگاه تکنولوژی های وب، گروه کامپیوتر، دانشکده فنی و مهندسی، دانشگاه زابل، زابل
b.lakzaei@yahoo.com

[2] آزمایشگاه پردازش زبان طبیعی، دانشکده مهندسی برق و کامپیوتر، دانشگاه شهید بهشتی، تهران
shams@sepehrs.com

چکیده

هستان‌شناسی یک توصیف صریح و صوری از یک مفهوم‌سازی اشتراکی است. ساخت دستی هستان‌شناسی‌ها امری بسیار پرهزینه، وقت‌گیر و مستعد خطا است. خودکارسازی فرآیند ساخت هستان‌شناسی یا به عبارت دیگر یادگیری هستان‌شناسی راه حل رفع این مشکلات است. یکی از منابع یادگیری هستان شناسی می تواند داده های نیمه ساخت یافته مانند اسناد XML باشد.
در این مقاله روشی نوین برای استخراج خودکار یک هستان‌شناسی به زبان OWL از شمای XML ارائه شده است. در مقایسه با روش‌های موجود، این روش علاوه بر تحلیل انواع ساده و پیچیده، عناصر، ویژگی‌ها، گروه عناصر و گروه ویژگی‌های موجود در شما، به تحلیل محدودیت‌های تعریف شده برای انواع ساده، تعریف کلید و تعریف کلید خارجی نیز می‌پردازد و در نتیجه مفاهیم و روابط و اصول بدیهی بیشتر و بهتری کشف می‌کند. این روش با یک هستان‌شناسی مرجع که به طور دستی در قلمروی مورد نظر ساخته شده است، مورد مقایسه قرار گرفته و معیارهای فراخوان، دقت و معیار-F به منظور ارزیابی کیفیت هستان‌شناسی تولید شده، محاسبه شده است. مقایسه این معیارها، بیانگر این است که روش پیشنهادی، نسبت به سایر روش‌های موجود عملکرد بهتری دارد.

کلمات کلیدی

هستان‌شناسی، یادگیری هستان‌شناسی، XML، XSD.

۱- مقدمه

وب جهان‌گستر[1]،یک منبع پهناور و رو به رشد از اطلاعات و سرویس‌هایی است که توسط افراد و برنامه‌های کاربردی مختلف به اشتراک گذاشته می‌شوند. درواقع، وب درحال تکامل از یک فضای اطلاعاتی و ارتباطاتی گسترده به سمت مخزنی انبوه از دانش وسرویس‌های مختلف می‌باشد و هستان‌شناسی[2] یکی ازعواملی است که امکان چنین تکاملی را فراهم می‌کند [1]. هستان شناسی نقش مهمی در پشتیبانی از تبادل و به اشتراک‌گذاری اطلاعات، از طریق فراهم کردن امکان تعامل‌پذیری[3] معنایی در وب، ایفا می‌کند.

انجمن بازنمایی دانش، هستان شناسی را به عنوان یک توصیف صریح و صوری از یک مفهوم‌سازی اشتراکی، تعریف می‌کند. در این تعریف، "مفهوم-سازی" به معنای یک مدل انتزاعی از پدیده‌های جهان و مفاهیم مرتبط با این پدیده‌ها است. منظور از "صریح" این است که مفاهیم مورد استفاده در این مدل و محدودیت‌های آن‌ها به طور صریح تعریف می‌شوند. "صوری بودن" این توصیف به معنای مدل‌پذیری هستان‌شناسی با هدف قابل فهم بودن آن برای ماشین می‌باشد. در نهایت، منظور از "اشتراکی" بودن، این است که هستان‌شناسی باید شامل دانش مشترکی باشد که مورد پذیرش جوامع علمی مختلف می‌باشد [1].

در یک دسته‌بندی کلی روش‌های ساخت هستان‌شناسی به دو دسته روش-های دستی و روش‌های (نیمه)/خودکار تقسیم می‌شوند. در روش‌های دستی کلیه مراحل فرآیند توسعه هستان‌شناسی به طور دستی و با دخالت مستقیم مهندس هستان‌شناسی انجام می‌شود. ساخت دستی هستان‌شناسی‌ها برای انواع مختلف قلمروها و کاربردها، عملی پرهزینه، زمان‌بر و مستعد خطا است. در روش‌های

Archive of SID

www.SID.ir

(نیمه)/خودکار ساخت هستان‌شناسی همه یا اکثر گام‌های فرآیند ساخت، به طور خودکار و بدون دخالت انسان انجام می‌شود. ازروش‌های (نیمه)/خودکار اغلب تحت عنوان یادگیری هستان‌شناسی[۴] یاد می‌شود. به دلیل مشکلات ذکر شده در روش‌های دستی ساخت هستان‌شناسی، یادگیری هستان‌شناسی بسیار مورد توجه محققان و مهندسان هستان‌شناسی می‌باشد، چرا که خودکارسازی عمل ساخت هستان‌شناسی نه تنها هزینه ساخت را کاهش می‌دهد، بلکه هستان‌شناسی تولید شده به این روش، انطباق بهتری با حوزه کاربردی مورد نظرخواهد داشت [2].

روش‌های یادگیری هستان‌شناسی بر اساس نوع منابع ورودی، به سه دسته تقسیم می‌شوند:

- یادگیری هستان‌شناسی از منابع غیرساخت‌یافته[۵] مانند متون زبان طبیعی و یا متون موجود در وب.
- یادگیری هستان‌شناسی از منابع نیمه ساخت‌یافته[۶] مانند مستندات HTML و XML و فرهنگ‌های لغات.
- یادگیری هستان‌شناسی از منابع ساخت‌یافته[۷] مانند پایگاه‌های داده و هستان‌شناسی‌های موجود.

زبان XMLیک زبان نشانه‌گذاری است که با تعریف مجموعه‌ای از قواعد، اسناد را به فرمتی کدگذاری می‌کند که هم برای انسان و هم برای ماشین قابل خواندن باشد. درواقع XML، یک فرمت داده متنی با پشتیبانی قوی از یونیکد برای زبان‌های مختلف موجود در جهان است. با وجود این‌که در هنگام طراحی XML تمرکز عمده روی داده‌های اسناد بود اما امروزه به طور گسترده شاهد به کارگیری این زبان برای نمایش ساختارهای داده‌ای دلخواه مانند وب سرویس‌ها نیز هستیم [3]. بنابراین می‌توان گفت بخش مهمی از داده‌ها در قالب مستندات XML ذخیره می‌شوند. از این رو در این مقاله به بررسی روش‌های موجود برای استخراج خودکار هستان‌شناسی از اسناد مبتنی بر XML و ارائه روشی نوین به منظور بهبود روش‌های پیشین پرداخته‌ایم.

۲- مطالعه روش‌های موجود

تا کنون روش‌های متعددی به منظور استخراج خودکار یک هستان‌شناسی از اسناد مبتنی بر XML پیشنهاد شده‌اند. برخی از این روش‌ها، هستان‌شناسی را به طور مستقیم از اسناد XML استخراج می‌کنند و برخی دیگر از شمای XML به منظور استخراج خودکار هستان‌شناسی استفاده می‌کنند. در ادامه این بخش به معرفی برخی از مهم‌ترین این روش‌ها می‌پردازیم.

یکی از روش‌های یادگیری هستان‌شناسی، روشی است که Cruz و همکارانش [4] برای ادغام اسناد ناهمگن XML با استفاده از یک معماری واسطه۸ مبتنی بر هستان‌شناسی پیشنهاد کردند. این فرآیند شامل دو گام است: انتقال شما۹ و ادغام هستان‌شناسی۱۰. بخش مورد نظر ما، گام اول می‌باشد که در آن، هر یک از اسناد XSD ورودی با استفاده از زبان RDF تبدیل به یک هستان‌شناسی محلی RDF می‌شوند. عمل تبدیل XSD به RDF به این صورت انجام می‌شود که عناصر نوع پیچیده تبدیل به rdfs:Class می‌شوند و ویژگی‌ها و عناصر نوع ساده به rdfs:property و روابط عنصر-زیرعنصر با استفاده تعریف سازنده‌ای جدید به نام rdfx:contain تبدیل به روابط بین کلاسی می‌شوند. در این روش، هستان‌شناسی تولید شده از لحاظ غنای معنایی بسیار ضعیف است و از جمله دلایل آن، استفاده از RDF برای نمایش هستان‌شناسی و استفاده از سازنده‌ای تعریف شده توسط نویسنده (rdfx:contain) که به طور معمول توسط RDF پشتیبانی نمی‌شود، ناتوانی در استخراج حتی روابط سلسله مراتبی، ناتوانی در استخراج اصول بدیهی حتی در حد چندی ویژگی‌های هستان‌شناسی و ... می‌باشد.

یکی دیگر از روش‌های یادگیری هستان‌شناسی از منابع نیمه ساخت‌یافته، روش ارائه شده توسط Battle [5] می‌باشد. در این روش نگاشتی مستقیم بین سند XML و مدل RDF انجام می‌شود و فرض می‌شود که شمای XML به عنوان راهنمای این نگاشت در دسترس می‌باشد. این روش نیز مانند روش قبلی غنای معنایی ندارد و تنها بخش کوچکی از دانش معنایی نهفته در اسناد XML را کشف می‌کند.

روش پیشنهادی Ferdinand و همکارانش [6] روشی دیگر برای ساخت یک هستان‌شناسی OWL از شمای XML و نیز تبدیل اسناد XML به گراف‌های RDF می‌باشد. فرآیند نگاشت XSD به هستان‌شناسی OWL بر مبنای مجموعه‌ای از قواعد از پیش تعریف شده انجام می‌شود و کلاس‌ها، ویژگی‌های نوع داده، ویژگی‌های شیئی و روابط سلسله مراتبی موجود در هستان‌شناسی به طور خودکار استخراج می‌شوند. اصول بدیهی نیز، محدود به تعیین چندی ویژگی‌های شیئی یا ویژگی‌های نوع داده، تعریف محدودیت owl:AllValuesFrom برای ویژگی‌های شیئی یا ویژگی‌های نوع داده، تبدیل xs:sequence و xs:all به owl:IntersectionOf و تبدیل xs:choice به ترکیبی از owl:complementOf، owl:intersectionOf و owl:unionOfمی‌باشد.

یکی دیگر از روش‌های استخراج هستان‌شناسی از شمای XML روشی است که توسط Bohring و Auer [7] ارائه شده است. در این روش هیچ قانونی برای تحلیل گروه عناصر، گروه ویژگی‌ها و استخراج وراثت تعریف نشده است. اصول بدیهی استخراج شده در این روش تنها محدود به استخراج چندی ویژگی‌های هستان‌شناسی و تبدیل xs:sequence و xs:all به owl:IntersectionOf و تبدیل xs:choice به ترکیبی از owl:complementOf، owl:intersectionOf و owl:unionOf می‌باشد.

Tsinaraki و Christodoulakis [8] ابزاری به نام XS2OWL ارائه کردند که امکان تبدیل خودکار شمای XML به یک هستان‌شناسی OWL و نیز تبدیل مجدد هستاشناسی OWL به شمای XML را می‌دهد. این ابزار نیز از روشی مشابه روش‌های دیگر برای استخراج خودکار هستان‌شناسی استفاده می‌کند اما در این ابزار هیچ قانونی برای استخراج خودکار رابطه سلسله‌مراتبی تعریف نشده است.

ابزار X2OWL توسط Ghawi و Cullot [9] برای ساخت خودکار یک هستان‌شناسی OWL از یک منبع داده XML پیشنهاد شد. روش پیاده سازی شده در این ابزار، بر مبنای شمای XML است که به تولید خودکار ساختار یک هستان‌شناسی (شامل کلاس‌ها، ویژگی‌های شیئی، ویژگی‌های نوع داده، روابط سلسله‌مراتبی)، نگاشت بین موجودیت‌های منبع داده XML و عناصر هستان‌شناسی تولید شده، نگاشت میان پرسش‌های OWL و XMLمی‌پردازد. در این روش، هیچ اصل بدیهی استخراج نمی‌شود.

روش پیشنهاد شده توسط Yahia و همکارانش [10] یکی دیگر از روش‌های یادگیری هستان‌شناسی از اسناد XMLاست. این روش از چهار گام اساسی تشکیل شده شده است:

- در مرحله اول سند XML به عنوان ورودی دریافت می‌شود و با استفاده از Trang [11] تبدیل به شمای XML می‌شود.
- شمای XML حاصل از مرحله قبل، با استفاده ازXSOM [12] تحلیل می‌شود.

Archive of SID

www.SID.ir

- خروجی XSOM به عنوان ورودی JUNG [13] مورد استفاده قرار می‌گیرد. چارچوب JUNG برای دستکاری‌های مبتنی بر گراف[11] به کار می‌رود و از روی شمای XML، گراف شمای XML[12](XSG) را تولید می‌کند.
- در نهایت این گراف XSGبه عنوان ورودی به ابزار Jena[14]داده می‌شود و از آن برای تولید عناصر هستان‌شناسی OWLاستفاده می‌شود.

در این روش هیچ اصل بدیهی استخراج نمی‌شود و این یکی از مهم‌ترین اشکالات وارد بر این روش می‌باشد.

Bendini و همکارانش [15] چارچوبی به نام Janus ارائه کرده‌اند که هدف آن تولید خودکار یک هستان‌شناسی از منبع عظیمی از شماهای XML است. آن‌ها مجموعه ای از الگوهای مختلف تعریف کرده‌اند که امکان تبدیل مستقیم و خودکار شمای XML به یک هستان‌شناسی OWL را فراهم می‌کند. تمرکز این روش، بر روی یک نمایش منطقی پیشرفته[13] از مولفه‌های شمای XML است و امکان کاوش در منابع شمای XML را برای استخراج دانش به منظور ساخت یک هستان‌شناسی فراهم می‌کند. این روش نیز با تعریف قواعد مختلف امکان تبدیل انواع پیچیده، گروه ویژگی‌ها و گروه عناصر به کلاس‌های هستان‌شناسی، تبدیل عناصر نوع ساده و ویژگی‌ها به ویژگی نوع داده، تبدیل عناصر نوع پیچیده به ویژگی شیئی، استخراج محدودیت‌های چندی و ... را فراهم می‌کند.

روشی که توسط Thuy و همکارانش [۱۶] ارائه شده، یکی دیگر از روشهای یادگیری هستان‌شناسی از شمای XML است. در این روش تمرکز اصلی بر روی یافتن عناصر مشابه در یک سند XML است. در واقع ایده این روش این است که گاهی در برخی اسناد XML ممکن است یک عنصر یا صفت چند بار تکرار شود و در همه یا اکثر موارد به یک معنا باشد. بنابراین در این روش، روشی برای محاسبه درصد تشابه میان دو عنصر یا صفت با نام یکسان تعیین شده و در صورتی که میزان تشابه از یک مقدار آستانه بالاتر باشد، این عناصر به یک مفهوم یا رابطه واحد تبدیل می‌شوند. در این روش، تلاشی برای بهبود روشهای استخراج ناصر هستان‌شناسی نشده است و در این زمینه مشابه سایر روش‌های ذکر شده و در برخی موارد ضعیف‌تر عمل می‌کند.

در بخش بعدی روش جدیدی برای یادگیری هستان‌شناسی از شمای XML پیشنهاد می‌شود که هدف آن پوشش هر چه بهتر و کاملتر ضعف‌های موجود در این روش‌ها می‌باشد.

۳- روش پیشنهادی برای یادگیری هستان‌شناسی از شمای XML

در این بخش روشی نوین به منظور استخراج خودکار هستان‌شناسی از شمای XML معرفی می‌کنیم. مبنای روش پیشنهاد شده در این بخش شمای XML می‌باشد اما در صورت عدم وجود شما در یک قلمروی خاص، می‌توان با استفاده از ابزارهای موجود مانند Trang [11] ابتدا از روی سند XML، شمای XML را استخراج کرد و سپس از این شما به عنوان ورودی الگوریتم یادگیری هستان‌شناسی استفاده کرد.

۳-۱- قواعد تبدیل شمای XML به هستان‌شناسی

در این بخش قواعد استخراج خودکار دانش هستان‌شناسانه را در پنج دسته مختلف شرح می‌دهیم: قواعد استخراج کلاس‌ها (مفاهیم)، قواعد استخراج ویژگی‌های نوع داده، قواعد استخراج ویژگی‌های شیئی، قواعد استخراج وراثت، و قواعد استخراج اصول بدیهی. در هردسته ترکیبی از قواعد معرفی شده در مراجع دیگر و قواعد جدید وضع شده در این مقاله قرار دارند که برای جلوگیری از تطویل مطلب تنها به توضیح قواعد جدید پرداخته و برای قواعد موجود قبلی به ذکر مرجع بسنده می کنیم.

۳-۱-۱- قواعد استخراج کلاس‌ها

قواعد C1، C2، C3 و C4 مشابه قواعد تعریف شده توسط Ghawi و Cullot [9] به منظور استخراج کلاس می‌باشد.

قاعده**C5**: هرگاه یک نوع یا عنصر پیچیده شامل یک عنصر ساده یا یک ویژگی باشد که نوع آن، یک نوع ساده تعریف شده با استفاده از برچسب <xs:simpleType> باشد و مقادیر مجاز برای این نوع ساده با استفاده از برچسب <xs:enumeration> محدود شده باشد، آن‌گاه به ازای هر مقدار مجاز (یا به عبارتی به ازای هر برچسب <xs:enumeration>) یک کلاس جدید تعریف می‌شود که این کلاس زیرکلاسی است از کلاس متناظر با نوع یا عنصر پیچیده مورد نظر (قاعده**H5**).

۳-۱-۲- قواعد استخراج ویژگی‌های نوع داده

قواعد Dt1، Dt2 و Dt3 مشابه قواعد تعریف شده توسط Ghawi و Cullot [9]به منظور استخراج ویژگی‌های نوع داده می‌باشد.

قاعده**Dt4**: وقتی در تعریف یک نوع ساده از برچسب <xs:union> استفاده می‌شود به ازای هر نوع تعیین شده در memberTypes یک ویژگی نوع داده تعریف می‌شود که دامنه آن برابر است با کلاس متناظر با والد این نوع ساده و برد آن برابر است با آن نوع تعیین شده در memberTypes.

۳-۱-۳- قواعد استخراج ویژگی‌های نوع شیئی

قواعد Obj1، Obj2 و Obj3 مشابه قواعد تعریف شده توسط Bendini و همکارانش [15] می‌باشد.

قاعده**Obj4**: هر کلید خارجی تعریف شده در XSD (با استفاده از برچسب <xs:keyref>) منجر به تعریف دو ویژگی شیئی می‌شود. در ویژگی اول، دامنه برابر است با کلاس متناظر با عنصر تعریف شده در ویژگی xpath برای برچسب <xs:selector>درون این کلید و برد این ویژگی برابر است با عنصر متناظر با مرجع این کلید خارجی. ویژگی دوم، معکوس ویژگی اول می‌باشد.

۳-۱-۴- قواعد استخراج وراثت

قواعد H1، H2، H3 و H4 مطابق با قواعد تعریف شده توسط Bendini و همکارانش [15] می‌باشد.

قاعده**H5**: هر کلاس جدید تعریف شده با استفاده از قاعده C5 زیرکلاسی است از کلاس متناظر با نوع یا عنصر پیچیده مورد نظر.

۳-۱-۵- قواعد استخراج اصول بدیهی

قواعد A1 تا A8 ترکیبی از قواعد تعریف شده توسط Bendini و همکارانش [15] و قواعد تعریف شده توسط Ferdinand و همکارانش [6] می‌باشد.

Archive of SID

www.SID.ir

قاعده A9: در صورتی که تنها ارزش مجاز برای یک ویژگی یا عنصری از نوع ساده با استفاده از fixed تعیین شده باشد، آن‌گاه یک محدودیت <owl:hasValue> روی ویژگی حاصل از آن عنصر یا ویژگی تعریف می‌شود.

قاعده A10: وقتی یک نوع ساده با استفاده از برچسب <xs:simpleType> تعریف می‌شود اگر محدودیتی با استفاده از برچسب‌های <xs:pattern>، <xs:length>، <xs:fractionDigits>، <xs:maxExclusive>، <xs:maxInclusive>، <xs:maxLength>، <xs:minExclusive>، <xs:minInclusive>، <xs:minLength>، <xs:totalDigits>، <xs:whiteSpace> روی آن نوع ساده تعریف شود، آن‌گاه این محدودیت روی تمام ویژگی‌های نوع داده‌ای که بر مبنای این نوع ساده تعریف می‌شوند نیز اعمال می‌شود.

قاعده A11: وقتی یک نوع ساده با استفاده از برچسب <xs:list> تعریف می‌شود، بدین معنا است که هر ویژگی یا عنصری که بر مبنای این نوع ساده تعریف می‌شود، حاوی لیستی از مقادیر است که نوع این مقادیر در ویژگی itemType این برچسب تعیین می‌شود. همان‌طور که در قوانین قبلی گفته شد، عناصر ساده یا ویژگی‌ها منجر به تعریف یک ویژگی نوع داده می‌شوند. بنابراین عنصر یا ویژگی تعریف شده بر اساس این نوع ساده، همانند یک صفت چندمقداری برای یک رابطه در شمای رابطه‌ای می‌باشد. بنابراین ارزش چندی حداکثر برای این ویژگی نوع داده، نباید یک باشد.

قاعده A12: هر گاه با استفاده از برچسب <xs:unique> برای یک نوع پیچیده مقادیر یک عنصر یا ویژگی منحصر بفرد تعریف شوند، باید با استفاده از <rdf:Description> این محدودیت را در هستان‌شناسی نشان دهیم.

قاعده A13: هر گاه با استفاده از <xs:key> برای یک عنصر یک کلید تعریف می‌شود، رفتار این کلید مشابه کلید اصلی در یک شمای رابطه‌ای می‌باشد. عنصری که این کلید برای آن تعریف شده است در ویژگی xpath در برچسب <xs:selector> و صفتی که نقش کلید را دارد در ویژگی xpath در برچسب <xs:field> تعیین می‌شود. بر اساس تعریف این کلید، محدودیتی تعریف می‌شود بدین صورت که ارزش چندی (<owl:cardinality>) برای ویژگی متناظر با این کلید برابر با یک قرار می‌گیرد. این ویژگی به صورت وظیفه‌ای (functional) تعریف می‌شود. در نهایت چون مقادیر کلید باید منحصر بفرد باشند محدودیتی نیز با استفاده از <rdf:Description> برای نشان دادن منحصر بفرد بودن مقادیر این کلید تعریف می‌شود.

۴- ارزیابی

در این بخش روش ارائه شده در این مقاله و روش‌های پیشین را مورد ارزیابی قرار می‌دهیم. هر یک از معیارهای دقت[۱۴]، فراخوان[۱۵]، و معیار-F[۱۶] برای تک تک روش‌های موجود و نیز روش پیشنهادی محاسبه و مورد مقایسه قرار گرفته است. نتایج حاصل از این مقایسه در جدول ۱ نشان داده شده است.

به دلیل نبود داده‌های تست استاندارد برای ارزیابی روش‌های یادگیری هستان‌شناسی از شمای XML و در دسترس نبودن داده‌های تست روش‌های موجود، از سه شمای XML مختلف برای ارزیابی روش ارائه شده استفاده کرده‌ایم. شمای اول مربوط به قلمروی یک محیط دانشگاهی است و هستان‌شناسی مرجع آن شامل ۳۰ مفهوم، ۵۰ ویژگی نوع داده، ۲۸ ویژگی نوع شیئی، ۲۰ رابطه سلسله مراتبی و ۱۰۲ اصل بدیهی می‌باشد. شمای دوم، شمای یک پایگاه داده مرتبط به یک قلمروی مربوط به سیستم هتل‌داری است. هستان‌شناسی مرجع مورد استفاده برای این قلمرو شامل ۲۳ مفهوم، ۲۷ ویژگی نوع داده، ۱۹ ویژگی شیئی، ۱۸ رابطه سلسله مراتبی و ۷۲ اصل بدیهی می‌باشد. در نهایت شمای سوم، شمای مربوط به یک شبیه‌سازی پویای مولکولی می‌باشد که در وب سایت http://www.quantum-simulation.org/ قابل دسترسی است. هستان-شناسی مرجع در این قلمرو شامل ۲۷ مفهوم، ۵۳ ویژگی نوع داده، ۱۴ ویژگی شیئی، ۱۵ رابطه سلسله مراتبی و ۱۶۲ اصل بدیهی است. در هر مورد هستان-شناسی مرجع به صورت دستی ساخته شده است و سپس به عنوان معیار ارزیابی روش‌های مختلف مورد استفاده قرار گرفته است. پس از پیاده‌سازی و اجرای هر یک از روش‌های موجود (که کدهای پیاده‌سازی آن‌ها در دسترس نمی‌باشد و پیاده‌سازی آن‌ها در این مقاله با توجه به قواعد تعریف شده در آن‌ها انجام شده است)، معیارهای مختلف در مورد عناصر مختلف هستان‌شناسی محاسبه و مقایسه شده است. مقایسه در هر مورد بر اساس مفاهیم، روابط، و اصول بدیهی موجود در هستان‌شنای مرجع انجام شده است، یعنی اگر در یک روش، عنصری استخراج شود که در هستان‌شناسی مرجع وجود ندارد، یا با عناصر موجود در هستان‌شناسی مرجع تداخل دارد، یک عنصر غلط محسوب می‌شود؛ و تنها عناصر استخراج شده-ای که در هستان‌شناسی مرجع وجود داشته باشند به عنوان عناصر صحیح در نظر گرفته می‌شود.

نتایج مقایسه روش ارائه شده در این مقاله با سایر روش‌های یادگیری هستان‌شناسی از اسناد مبتنی بر XML نشان می‌دهد که روش پیشنهاد شده تقریبا در تمامی زمینه‌ها کامل‌تر از روش‌های موجود است و نسبت به آن‌ها عملکرد بهتری دارد. اصلی‌ترین دلیل موفقیت این روش را می‌توان توانایی آن در تحلیل کلید اصلی، کلید خارجی، تعریف خصوصیت منحصر بفرد بودن مقادیر یک عنصر یا ویژگی، انواع محدودیت‌های تعریف شده روی یک نوع ساده و محدودیت enumeration و استخراج مفاهیم، روابط و اصول بدیهی حاصل از این ساختارهای موجود در شمای رابطه‌ای دانست. ما در این روش با تعریف قوانین مختلف به تحلیل این ساختارها و استخراج خودکار عناصر هستان‌شناسانه از آن‌ها پرداخته‌ایم و این امر منجر به بهبود کیفیت هستان‌شناسی خروجی حاصل از این روش، نسبت به سایر روش‌های موجود شده است.

۵- نتیجه گیری و کارهای آتی

محبوبیت هستان‌شناسی به عنوان یک فناوری جدید که پتانسیل عظیمی برای بهبود روش‌های سازماندهی، مدیریت و درک اطلاعات دارد، روز به روز در حال افزایش است. یکی از مهم‌ترین چالش‌های موجود در هستان‌شناسی‌ها روش ساخت آن‌ها می‌باشد. به دلیل مشکلات موجود در روش‌های دستی، در سال‌های گذشته توجه محققین روی خودکارسازی فرآیند ساخت هستان‌شناسی معطوف شده است.

در این مقاله روشی نوین برای یادگیری هستان‌شناسی از شمای XML پیشنهاد شده است. در این روش شمای XML به عنوان ورودی دریافت می‌شود و پس از اعمال قواعد مختلف به منظور استخراج عناصر مختلف هستان‌شناسی، یک هستان‌شناسی به زبان OWL تولید می‌شود. این روش با روش‌های موجود مقایسه شد و نتایج حاصل از این مقایسه نمایانگر عملکرد بهتر این روش در خودکارسازی فرآیند ساخت هستان‌شناسی می‌باشد.

با وجود عملکرد بهتر روش ارائه شده در این مقاله، یکی از مشکلات موجود در این روش و روش‌های پیشین، نام‌های معمولا بی‌معنا برای مفاهیم یا روابط استخراج شده است. چرا که این اسامی با توجه به ساختاری که مفهوم یا رابطه از

Archive of SID

www.SID.ir

آن استخراج شده است، یا با توجه به نام دامنه و برد و عامل ایجاد ارتباط بین آن-ها، انتخاب می‌شود. می‌توان در کارهای آتی با تعریف مجموعه‌ای از لغات معنادار و ایجاد ارتباط بین این لغات و ساختارهای مختلف شمای XML اسامی حداقل معنادارتری برای مفاهیم و روابط هستان‌شناسی انتخاب کرد.

جدول ۱: مقایسه روش‌های یادگیری هستان‌شناسی بر اساس معیارهای دقت، فراخوان و معیار-F

شماره منبع	Precision			Recall			F-measure		
	شمای ۱	شمای ۲	شمای ۳	شمای ۱	شمای ۲	شمای ۳	شمای ۱	شمای ۲	شمای ۳
روش پیشنهادی	۰.۹۵۹۴	۱	۱	۰.۹۶۴۳	۱	۱	۰.۹۶۱۸	۱	۱
[۴]	۰.۸۵۳۶	۰.۸۴	۱	۰.۱۷۸۶	۰.۲۰۳۹	۰.۱۷۷۸	۰.۲۹۵۴	۰.۳۲۸۱	۰.۳۰۱۹
[۵]	۰.۸۸۱۳	۰.۸۷۱۰	۱	۰.۲۶۵۳	۰.۲۶۲۱	۰.۲۶۳۰	۰.۴۰۷۸	۰.۴۰۲۹	۰.۴۱۶۵
[۶]	۰.۷۷۴۴	۰.۷۹۱۰	۰.۶۸۱۱	۰.۵۲۵۵	۰.۵۱۴۶	۰.۴۶۶۷	۰.۶۲۶۱	۰.۶۲۳۵	۰.۵۵۳۹
[۷]	۰.۸۸۷۱	۰.۸۸۲۳	۱	۰.۲۸۰۶	۰.۲۹۱۳	۰.۳۱۴۸	۰.۴۲۶۳	۰.۴۳۸۰	۰.۴۷۸۸
[۸]	۰.۸۶۷۹	۰.۸۶۶۷	۱	۰.۲۳۴۷	۰.۲۵۲۴	۰.۳۱۱۱	۰.۳۶۹۵	۰.۳۹۰۹	۰.۴۷۴۶
[۹]	۰.۸۷۰۴	۰.۸۶۲۱	۱	۰.۲۳۹۸	۰.۲۴۲۷	۰.۲۲۲۲	۰.۳۷۶۰	۰.۳۷۸۸	۰.۳۶۳۶
[۱۰]	۰.۸۷۰۴	۰.۸۶۲۱	۱	۰.۲۳۹۸	۰.۲۴۲۷	۰.۲۲۲۲	۰.۳۷۶۰	۰.۳۷۸۸	۰.۳۶۳۶
[۱۵]	۰.۸۸۱۳	۰.۸۸۵۷	۱	۰.۲۶۵۳	۰.۳۰۱۰	۰.۴۹۲۶	۰.۴۰۷۸	۱.۱۸۶۷	۰.۶۶۰۰
[۱۶]	۰.۸۹۲۳	۰.۸۹۷۴	۱	۰.۲۹۵۹	۰.۳۲۷۱	۰.۲۹۷۸	۰.۴۴۴۴	۰.۴۷۹۴	۰.۴۵۸۹

مراجع

[1] Gruber, T.R., "A translation approach to portable ontologies", Knowledge Acquisition, vol. 5, no. 2, pp. 199-220, 1993.

[2] Shamsfard, M., and Barforoush, A., "The state of the art in ontology learning: A framework for comparison", The Knowledge Engineering Review, vol. 18, no. 4, pp. 293-316, 2003.

[3] XML, 2013, Available: http://en.wikipedia.org/wiki/XML

[4] Cruz, I. F., Xiao, H., and Hsu, F., "An Ontology-based Framework for XML Semantic Integration", Proc. of the International Database Engineering and Applications Symposium, pp. 217–226, 2004.

[5] Battle, S., "Round-tripping between XML and RDF", International Semantic Web Conference(ISWC), Hiroshima, Japan, Springer, 2004.

[6] Ferdinand, M., Zirpins, C., and Trastour, D., "Lifting XML Schema to OWL", Web Engineering - 4th International Conference, Munich, Germany, pp. 354–358, 2004.

[7] Bohring, H. and Auer, S., "Mapping XML to OWL Ontologies",Leipziger*Informatik-Tage*, vol. 72, pp. 147–156, 2005.

[8] Tsinaraki C., Christodoulakis, S., "XS2OWL: A Formal Model and a System for enabling XML Schema Applications to interoperate with OWL-DL Domain Knowledge and Semantic Web Tools", Proc. of the DELOS Conference, Tirrenia, Italy, pp. 124-136, 2007.

[9] Ghawi, R., and Cullot, N., "Building Ontologies from XML Data Sources", 1st International Workshop on Modelling and Visualization of XML and Semantic Web Data (MoViX '09), held in conjunction with DEXA'09 (Linz, Austria, pp. 480-484, 2009.

[10] Yahia, N.,Mokhtar, S.A., and Ahmed, A., "Automatic Generation of OWL Ontology from XML Data Source", International Journal of Computer Science Issues, vol. 9, no. 2, pp. 77-83, 2012.

[11] Trang, 20012, Available: http://www.thaiopensource.com/relaxng/trang.html

[12] XSOM, 2012, Available: https://xsom.dev.java.net/

[13] JUNG, 2012, Availabl: http://jung.sourceforge.net/

[14] Apache Jena, 2012, Available: http://jena.apache.org/

[15] Bedini, I., Gardarin, G., and Nguyen, B., "Transforming XML Schema to OWL Using Patterns", 5th IEEE International Conference on Semantic Computing (ICSC), pp. 102-109, 2011.

[16] Thuy, P. T. T., Lee, Y. , and Lee, S., "A Semantic Approach for Transforming XML Data into RDF Ontology", Wireless Personal Communications, Vol. 73, Issue 4, pp. 1387-1402, 2013.

زیرنویس‌ها

[1] World Wide Web
[2] ontology
[3] interoperability
[4] Ontology Learning
[5] un-structured resources
[6] semi structured resources
[7] structured resources
[8] Mediation
[9] schema transformation
[10] ontology merging
[11] graph-based manipulation
[12] XML-Schema Graph
[13] Advanced logical representation
[14] precision
[15] recall
[16] F-measure

Archive of SID

www.SID.ir