

بررسی روش‌های انتقال پایگاه داده‌های رابطه‌ای به آنتولوژی

محمد رضا پورخانی^۱، بیتا شادگار^۲، علی رضا عصاره^۳

چکیده

هدف: هدف از انجام این پژوهش، معرفی روش انتقال پایگاه داده‌های رابطه‌ای به آنتولوژی، دسته-بندی منابع ورودی برای روش انتقال و بررسی مهم‌ترین روش‌های موجود در هر دسته است. **روش تحقیق:** در این پژوهش از روش تحقیق کتابخانه‌ای برای جمع‌آوری و بررسی مقالات موجود در زمینه انتقال پایگاه داده‌های رابطه‌ای به آنتولوژی استفاده شده است.

یافته‌ها: پس از بررسی روش‌های موجود برای انتقال پایگاه داده‌های رابطه‌ای به آنتولوژی، این روش‌ها براساس منبع ورودی به سه دسته تقسیم شدند. منبع ورودی برای پایگاه داده رابطه‌ای می‌تواند شمای رابطه‌ای، مدل‌های داده‌ای مفهومی و صفحات HTML باشد. ساده‌ترین منبع قابل دسترسی که نیاز به همکاری مالک پایگاه داده ندارد، صفحات HTML مرتبط با پایگاه داده است. صفحات HTML ارتباط بین فرم‌های وبسایت و پایگاه داده را برقرار می‌کنند ولی نمی‌توان آن‌ها را به‌عنوان مدل واقعی پایگاه داده رابطه‌ای در نظر گرفت، در نتیجه نمی‌توان معانی نهفته در پایگاه داده‌های رابطه‌ای را از این منبع استخراج کرد. همچنین از نظر دسترسی به داده‌ها باید خاطر نشان کرد که بازیابی همه‌ی رکوردهای پایگاه داده رابطه‌ای از طریق وبسایت امری غیرممکن است. ساخت مدل مفهومی همیشه نخستین گام در فرآیند ایجاد پایگاه داده است، بنابراین ممکن است این مدل بیان‌گر آخرین تغییرات و اصلاحات پایگاه داده نباشد. همچنین برخی ابزارهای مدل‌سازی مفهومی پایگاه داده از همه ویژگی‌های مدل‌های مفهومی پشتیبانی نمی‌کنند. بنابراین ملاحظه می‌شود که شمای رابطه‌ای بهترین منبع برای تولید آنتولوژی از پایگاه داده رابطه‌ای است، هنگامی که ساختار پایگاه داده در دسترس و مالک پایگاه داده همکاری لازم را داشته باشد. **نتیجه‌گیری:** با توجه به بررسی‌های انجام شده، بهترین منبع ورودی برای انتقال پایگاه داده‌های رابطه‌ای به آنتولوژی، شمای رابطه‌ای است.

کلیدواژه‌ها: بررسی، وب معنایی، انتقال، آنتولوژی، پایگاه داده رابطه‌ای

۱- دانشجوی کارشناسی ارشد، مهندسی کامپیوتر-نرم‌افزار، مؤسسه آموزش عالی جهاد دانشگاهی خوزستان.

Mr.pourkhani@hotmail.com

۲- استادیار، دانشکده مهندسی، گروه کامپیوتر، دانشگاه شهید چمران اهواز. Bita.shadgar@scu.ac.ir

۳- دانشیار، دانشکده مهندسی، گروه کامپیوتر، دانشگاه شهید چمران اهواز. Alireza.osareh@scu.ac.ir

مقدمه

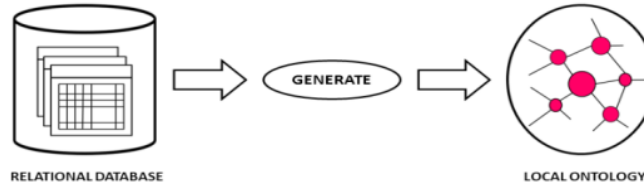
هدف وب معنایی که به عنوان وب داده‌ها^۱ نیز شناخته می‌شود، یکپارچه‌سازی داده‌های منابع گوناگون است. در سال ۲۰۰۷، هی و همکاران (He, Patel, Zhang & Chang, 2007) مشخص کردند که دسترسی به داده‌های صفحات وبی که با پایگاه‌های داده ارتباط دارند (صفحات پویای وب^۲)، حدود ۵۰۰ برابر بیش از داده‌های صفحات ایستاست و تقریباً ۷۰ درصد وب‌سایت‌ها، از پایگاه داده‌های رابطه‌ای استفاده می‌کنند. با وجود این مقدار عظیم از داده‌ها که در پایگاه داده‌های رابطه‌ای ذخیره شده است، موفقیت وب معنایی به طرز چشم‌گیری به توسعه روش‌هایی برای یکپارچه‌سازی پایگاه داده‌های رابطه‌ای با وب معنایی گره خورده است. از سوی دیگر محبوبیت آنتولوژی‌ها (آنتولوژی در وب معنایی اصطلاحات و ارتباط بین آن‌ها در دامنه‌ی مفروض را بیان می‌کند (شادگار، عصاره و هراتیان، ۱۳۸۹)). از زمان ظهور وب معنایی به سرعت در حال گسترش است، زیرا برای استفاده از وب معنایی، اطلاعات باید به فرمت قابل فهم برای ماشین نمایش داده شوند. کنسرسیوم W3C چند فرمت برای نمایش داده‌ها در وب معنایی از جمله RDF Schema، RDF و OWL پیشنهاد کرده است. از آنجایی که تولید آنتولوژی به روش کاملاً دستی کاری زمان‌گیر، پرزحمت و مستعد خطاست که نیاز به همکاری کارشناس (ان) دامنه دارد، پس بسیار مهم است که دست‌کم به صورت نیمه‌خودکار از پایگاه داده‌های رابطه‌ای، آنتولوژی استخراج شود. در این مقاله، ابتدا روش انتقال پایگاه داده‌های رابطه‌ای به آنتولوژی بیان و سپس مهم‌ترین روش‌های انتقال براساس منبع ورودی دسته‌بندی و بررسی می‌شوند. در پایان نیز نتیجه‌گیری از بررسی‌های انجام شده ارائه می‌شود.

انتقال پایگاه داده‌های رابطه‌ای به آنتولوژی

در رویکرد انتقال، از شمای پایگاه داده و داده‌های آن استفاده می‌شود تا پایگاه داده به نمایشی از وب معنایی تبدیل شود. در این رویکرد، آنتولوژی محلی وجود ندارد و با استفاده از قوانین انتقال، هر بخش از پایگاه داده‌های رابطه‌ای به معادل آن در آنتولوژی تبدیل می‌شود (شکل ۱).

1 - Web of Data

2 - Dynamic Web Pages



شکل ۱: انتقال پایگاه داده رابطه‌ای به آنتولوژی

قوانین انتقال در این رویکرد ناشی از شباهت‌های موجود بین مدل رابطه‌ای و مدل آنتولوژی و همچنین با در نظر گرفتن مدل آنتولوژی به عنوان نوعی مدل مفهومی و استفاده از شیوه‌های مهندسی معکوس است. همچنین در این رویکرد می‌توان از منابع زیر به منظور بهبود یا غنی‌سازی آنتولوژی استخراج شده استفاده کرد:

- اطلاعات کارشناس دامنه
- تجزیه و تحلیل پرس‌وجوهای کاربر
- محتوای (تاپل‌های) پایگاه داده رابطه‌ای
- مخزن دانش واژگانی مانند وردنت^۱
- صفحات HTML
- مدل‌های داده‌ای مفهومی مانند ER/EER
- اسناد XML

در رویکرد انتقال پایگاه داده رابطه‌ای به آنتولوژی، از شمای پایگاه داده رابطه‌ای دست‌کم به صورت نیمه‌خودکار آنتولوژی استخراج شده و از تولید آنتولوژی به صورت کاملاً دستی و غیرخودکار جلوگیری می‌شود. به دو دلیل باید به روش انتقال پایگاه داده‌های رابطه‌ای به آنتولوژی توجه ویژه‌ای شود. پایگاه داده رابطه‌ای و آنتولوژی مرتبط با آن ممکن است همواره دردسترس نباشد، بنابراین استخراج آنتولوژی از پایگاه داده رابطه‌ای برای به دست آوردن معانی نهفته در آن و همچنین استفاده از آنتولوژی استخراج شده برای انتشار معنایی محتوای پایگاه داده رابطه‌ای روی اینترنت به صورت سه‌تایی‌های^۲ RDF می‌تواند گامی بلند در حرکت به سوی ماشین فهم کردن

1 - Word Net

2 - Triples

محتوای وب کنونی باشد. همچنین آنتولوژی‌های مرجع^۱ از قبل تهیه شده و در دسترس هستند، ولی آنتولوژی‌های محلی^۲ که مهم و پر تعداد هستند، ممکن است در دسترس نباشند و باید با استفاده از روش انتقال این نوع از آنتولوژی‌ها را دست‌کم به صورت نیمه خودکار تولید کرد. همچنین لازم به ذکر است که ساخت آنتولوژی با استفاده از تجزیه و تحلیل شمای رابطه‌ای، ممکن است از نظر تمامیت^۳ و صحت^۴ اطلاعات ورودی دارای محدودیت‌هایی باشد، از جمله:

- بسیاری از طراحان پایگاه داده‌ها، با بهینه‌سازی و کاهش درجه (سطح) نرمالیتی^۵ شمای رابطه‌ای، قوانین طراحی خوب را نقض می‌کنند.
 - معمولاً اطلاعات کامل درباره وابستگی‌های تابعی^۶ در شمای رابطه‌ای وجود ندارد.
 - از آنجایی که شمای رابطه‌ای از همه‌ی ساخت‌های شمای مفهومی پشتیبانی نمی‌کند، در نتیجه برخی معانی موجود در شمای مفهومی مانند وراثت، هنگام ترجمه‌ی شمای مفهومی به شمای رابطه‌ای از دست خواهند رفت.
 - اغلب اسامی روابط و صفات در شمای رابطه‌ای به شکل مختصر هستند، در نتیجه استنتاج معانی داده‌ها از این اسامی بسیار دشوار یا حتی غیرممکن است.
- نکات مهمی که در رویکرد انتقال پایگاه داده رابطه‌ای به آنتولوژی باید مورد توجه باشند، عبارتند از:

- کیفیت آنتولوژی تولید شده وابسته به کیفیت و پیچیدگی مدل‌سازی مفهومی است که شمای پایگاه داده رابطه‌ای از آن به وجود آمده است.
- دلیل این که هیچ روش قانون‌مندی برای مدل‌سازی یک آنتولوژی وجود ندارد، آن است که این موضوع بسیار وابسته به دامنه‌ی مفروض یا کارشناس (های) آن دامنه است. به همین دلیل هنگامی که قرار است از پایگاه داده رابطه‌ای، آنتولوژی استخراج

1 - Global Ontology

2 - Local Ontology

3 - Completeness

4 - Correctness

5 - De-Normalized

6 - Functional Dependency

شود، پایگاه داده رابطه‌ای باید از نظر پوشش مفاهیم دامنه‌ی مفروض و همچنین وابستگی شمای رابطه‌ای به سطوح مختلف نرمال تجزیه و تحلیل شود.

- به همان شکل که یک پایگاه داده رابطه‌ای را می‌توان به سطوح مختلف نرمال‌سازی کرد، آنتولوژی که از هر یک از این سطوح نرمال به دست می‌آید، با توجه به دامنه مورد نظر می‌تواند اشکال مختلفی داشته باشد.
- ممکن است نیاز باشد که آنتولوژی استخراج شده با آنتولوژی مرجع یا دیگر آنتولوژی‌های محلی از آن دامنه یکپارچه‌سازی شود، البته اگر هدف ایجاد تعامل در مقیاس وسیع باشد.
- سیستم‌های انتقال به دو دسته‌ی مستقل از دامنه و وابسته به دامنه تقسیم می‌شوند. با استفاده از سیستم‌هایی که مستقل از دامنه هستند، می‌توان از پایگاه‌های داده در دامنه‌های مختلف آنتولوژی استخراج کرد ولی در سیستم‌های وابسته به دامنه چنین امکانی مُیسر نیست.
- قوانین انتقال هر دو سیستم باید کاملاً بدون ابهام باشند، به عبارت دیگر خوانندگان از قانون دو برداشت متفاوت نکنند. به علاوه قوانین در سیستم‌های مستقل از دامنه باید همه-منظوره باشند، یعنی قانون در همه‌ی دامنه‌ها برقرار باشد.
- قوانین سیستم انتقال با استفاده از روش‌های توصیف رسمی (استفاده از منطق مرتبه اول)، توصیف به زبان طبیعی (انگلیسی یا فارسی) و مثال بیان می‌شوند. در اغلب روش‌ها از یک یا ترکیبی از این روش‌ها استفاده شده است. لازم به ذکر است که در سیستم‌های مستقل از دامنه برای این که ویژگی تعمیم قوانین رعایت شود، حتماً باید قوانین سیستم انتقال به روش توصیف رسمی فرمول‌نویسی شوند.

روش‌های انتقال پایگاه داده‌های رابطه‌ای به آنتولوژی

از آنجایی که آنتولوژی شباهت‌های بسیاری با مدل‌های داده‌ای مفهومی دارد، لذا می‌توان از روش‌های مهندسی معکوس برای تولید آنتولوژی از پایگاه داده رابطه‌ای استفاده کرد. بنابراین خروجی روش‌های مهندسی معکوس پایگاه داده رابطه‌ای، آنتولوژی است. روش‌های مهندسی معکوس به کار رفته در این زمینه را می‌توان براساس نوع ورودی‌شان به سه دسته تقسیم‌بندی کرد:

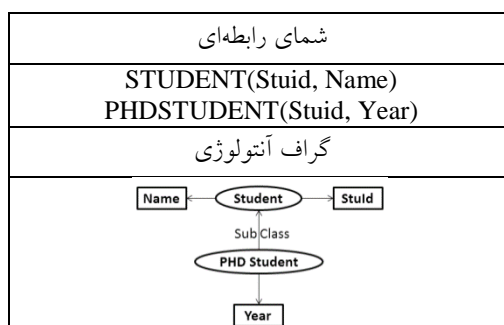
- مبتنی بر تجزیه و تحلیل شمای رابطه‌ای

- مبتنی بر تجزیه و تحلیل صفحات HTML
 - مبتنی بر تجزیه و تحلیل مدل‌های داده‌ای مفهومی مانند ER/EER
- در ادامه، این مقاله مروری بر این سه دسته روش دارد که به تفصیل بررسی خواهند شد.

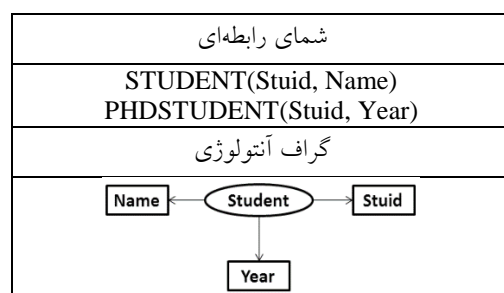
روش‌های مبتنی بر تجزیه و تحلیل شمای رابطه‌ای

در این گروه، روش‌ها از شمای رابطه‌ای به عنوان منبع ورودی برای پایگاه داده رابطه‌ای استفاده می‌کنند. هر پایگاه داده‌ی رابطه‌ای ساختار جداولش را به زبان پرس و جوی ساخت یافته یا به-اختصار SQL^۱ نگهداری می‌کند.

استویانوویچ و همکاران (Stojanovic & Volz, 2002) نخستین روش را با هدف استخراج معانی دامنه از شمای رابطه‌ای و ترجمه آن به نمایشی از وب معنایی منتشر کردند. این روش قوانینی برای انتقال شمای رابطه‌ای به Frame Logic و سپس RDF Schema فراهم می‌کند. استویانوویچ و همکاران (Stojanovic et al., 2002)، ۱۰ قانون برای نگاشت جداول، صفات و قیود به مفاهیم، ویژگی‌ها و اصول تعریف کردند. در میان این قوانین، قانونی برای شناسایی و حذف جداولی که در نتیجه‌ی ارتباطات چندبه‌چند ایجاد شده‌اند، وجود دارد. همچنین قانونی نیز برای شناسایی سلسله-مراتب ISA بین جداول پدر و فرزند، هنگامی که این جداول از طریق کلید خارجی به یکدیگر مرتبط شده‌اند، وجود دارد. در این روش تفاوتی بین وراثت و بخش‌بندی وجود ندارد. شکل ۲ و ۳، ترجمه شمای رابطه‌ای به گراف آنتولوژی را نمایش می‌دهد. در این مثال، دو رابطه که کلیدهای اصلی یکسان دارند، می‌توانند به یک مفهوم واحد یا به یک سلسله‌مراتب ISA نگاشت شوند. بنابراین، سیستم در این مورد وابسته به تصمیم کاربر است تا ابهام را حل کند. لازم به ذکر است در این بخش از بیضی برای بیان کلاس، از مستطیل برای بیان ویژگی و از پیکانه برای بیان ارتباط استفاده شده است.



شکل ۲: ترجمه شمای رابطه‌ای با استفاده از قانون انتقال وراثت



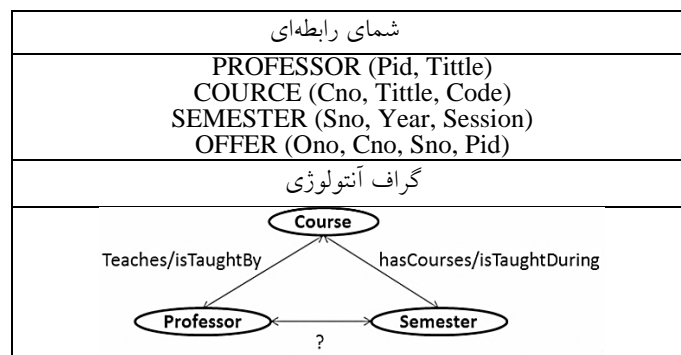
شکل ۳: ترجمه شمای رابطه‌ای با استفاده از قانون انتقال بخش‌بندی

مشکل دیگر این روش استفاده از فرآیند میانی است، یعنی نیاز به تولید مدل میانی از مدل مبدأ و سپس تبدیل این مدل به مدل مقصد است. مدل مبدأ در این روش، شمای رابطه‌ای و مدل میانی به فرمت Frame Logic و مدل مقصد آنتولوژی RDF Schema است. همچنین استفاده از زبان آنتولوژی RDF Schema به‌عنوان نمایش مقصد است، یکی دیگر از مشکلات این روش است، زیرا این زبان دارای کاستی‌های بسیاری نسبت به زبان آنتولوژی OWL است.

لی و همکاران (Li, Du & Wang, 2005) روشی برای تبدیل شمای رابطه‌ای به آنتولوژی OWL ارائه کردند. نویسندگان، روش پیشنهادی‌شان را در دو مقاله ارائه کرده‌اند و یکی از این کارها را خودکار و دیگری را نیمه‌خودکار نام‌گذاری کرده‌اند، در صورتی که قوانین هر دو کار یکسان بوده و دلیلی برای این نام‌گذاری متفاوت ذکر نکرده‌اند. آن‌ها در هر دو کار، ۱۲ قانون برای سیستم انتقال به وسیله توصیفات به زبان انگلیسی، مثال و برخی توصیفات رسمی بیان کرده‌اند. قوانین ارائه شده منجر به تولید کلاس‌های آنتولوژی، ویژگی‌ها، سلسله‌مراتب وراثت، کاردینالیتی و

نمونه‌ها می‌شوند. لی و همکاران (Li et al., 2005)، قانونی را برای یادگیری یک کلاس آنتولوژی OWL بیان کردند، به طوری که اگر چند رابطه برای بیان یک موجودیت استفاده شود و همگی این روابط دارای کلیدهای اصلی مشترک و یکسان باشند، آن‌گاه همه روابط در یک کلاس آنتولوژی OWL یکپارچه می‌شوند. این قانون هنگامی در نظر گرفته می‌شود که بخش‌بندی در شمای رابطه‌ای به کار گرفته شده باشد. همچنین از همین قانون برای یادگیری سلسله‌مراتب ISA نیز استفاده می‌شود. در نتیجه به کارگیری این قانون نیاز به تعامل کاربر دارد.

در این روش، اگر کلید اصلی رابطه‌ای، کلید خارجی اش نیز باشد، در این صورت از کلید خارجی برای تعریف دو ویژگی معکوس با نام‌های Has-Part و Is-Part-Of با دامنه و برد مناسب استفاده می‌شود. به علاوه می‌توان از این ارتباط برای نمایش سلسله‌مراتب ISA نیز استفاده کرد. در نتیجه این قانون دارای دو معنا است. اگر رابطه‌ای ارتباطات چندتایی را بیان کند، به تعداد کلیدهای خارجی موجود در آن جدول، ویژگی شیء ایجاد می‌شود و سپس با استفاده از این ویژگی‌های شیء، روابط شرکت کننده در ارتباط، با یکدیگر مرتبط می‌شوند. به کارگیری این قانون باعث ایجاد ارتباطات غیر ضروری می‌شود. به عنوان مثال، اگر شمای رابطه‌ای شکل ۴ که یک ارتباط سه‌تایی را بیان می‌کند، در نظر بگیریم، با استفاده از این قانون، ارتباطات Professor-Course-Semester، Course-Semester، Professor-Semester ایجاد می‌شوند. آیا ارتباط Professor-Semester ضروری است؟ شاید ارتباطات دیگر که بیشتر ایجاد شده‌اند، دوباره تولید شوند! به علاوه این روش، حالتی را که ارتباطات چندتایی دارای صفت یا صفات اضافی باشند، در نظر نمی‌گیرد.



شکل ۴: ترجمه شمای رابطه‌ای با استفاده از قانون انتقال ارتباطات چندتایی لی و همکاران

از آنجایی که لی و همکاران (Li et al., 2005)، هدف خود را ساخت سیستمی برای استخراج آنتولوژی از دامنه‌های گوناگون ذکر کرده‌اند، در نتیجه باید قوانین سیستم انتقال به‌طور کامل به روش توصیف رسمی فرمول‌نویسی شوند.

آستروا و همکاران (Astrova & Kalja, 2007) قوانینی با استفاده از توصیفاتی به زبان انگلیسی و مثال برای بیان یک سیستم انتقال خودکار که شمای رابطه‌ای را به آنتولوژی OWL ترجمه می‌کند، بیان کرده‌اند. همچنین خروجی سیستم‌شان مطابق با محدودیت‌های آنتولوژی‌های نوع OWL-DL نیست. تفاوت اصلی بین کار آستروا و همکاران (Astrova et al., 2007) با کار لی و همکاران (Li et al., 2005)، در نگاشت ارتباطات چندتایی است. در این روش صفت با قید Unique به ویژگی تابعی معکوس نگاشت می‌شود. اگر آنتولوژی هدف از نوع OWL-DL باشد، این قانون صحیح نخواهد بود، زیرا در این نوع آنتولوژی OWL، ویژگی تابعی معکوس فقط برای ویژگی‌های شیء در نظر گرفته می‌شود و نه ویژگی‌های نوع داده. قید Unique، یک قید کلید را تعریف می‌کند و در آنتولوژی‌های واقعی نیز این قید منتهی به محاسبات کنترل‌نشده^۱ می‌شود که توسط هیچ‌کدام از استنتاج‌گرهای موجود OWL، پشتیبانی نمی‌شود. اگر آنتولوژی از نوع OWL-Full باشد، این قانون قابل قبول است.

در این روش، اگر کلید اصلی بخشی از کلید خارجی باشد، آن‌گاه به یک ویژگی شیء با قید کاردینالیتی ۱ نگاشت می‌شود. لی و همکاران این مورد را به ویژگی شیء Has-Part ترجمه کرده‌اند. همچنین این قانون می‌تواند برای شناسایی سلسله‌مراتب وراثت نیز استفاده شود. اگر کلید اصلی در یک جدول به عنوان کلید خارجی همان جدول درج شده باشد، آن‌گاه این کلید خارجی به یک ویژگی شیء با مشخصه‌ی تقارنی نگاشت می‌شود. ولی این قانون وابسته به دامنه است و نمی‌توان آن را در سیستم‌های مستقل از دامنه به کار برد. همچنین اگر کلید اصلی جدول به عنوان کلید خارجی در همان جدول به همراه قید On Delete Cascade درج شده باشد، آن‌گاه این مدل یک ارتباط جزء-کل^۲ را نشان می‌دهد، یعنی جز نمی‌تواند بدون کل موجود باشد و در نتیجه کلید خارجی به ویژگی شیء با مشخصه‌ی تراگذری نگاشت می‌شود که کاملاً هم صحیح است (شکل ۵).

1 - Intractable

2 - Whole-Part

شمای رابطه‌ای
Create Table COLLECTION{ COLLECTIONNAME varchar primary key, SUBCOLLECTION varchar references COLLECTION(COLLECTIONNAME) ON DELETE CASCADE};
آنتولوژی OWL
<owl:TransitiveProperty rdf:ID="SubCollection"> <rdfs:domain rdf:resource="#COLLECTION"/> <rdfs:range rdf:resource="#COLLECTION"/> </owl:TransitiveProperty>

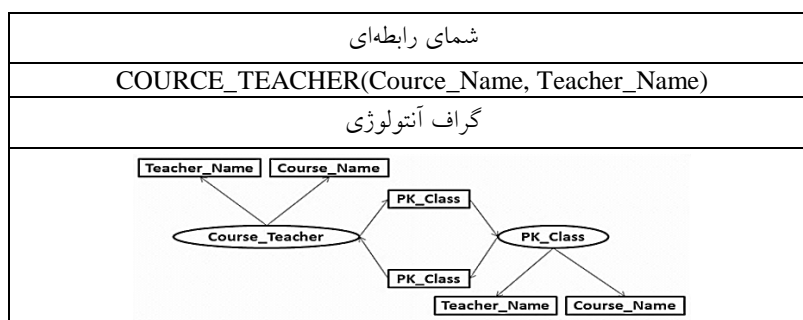
شکل ۵: ترجمه شمای رابطه‌ای با استفاده از قانون انتقال جزء- کل آستروا و همکاران

به‌طور کلی، آستروا و همکاران در این کار از شمای پایگاه داده رابطه‌ای نسبت به لی و همکاران (Li et al., 2005)، معانی بیشتری استخراج کرده‌اند.

اطمینانی و همکاران (Etminani, Kahani & Yanesari, 2009) برای انتقال پایگاه داده‌های رابطه‌ای به آنتولوژی، چهار گام را شامل نگاشت جداول، نگاشت ستون‌ها، نگاشت ارتباطات، نگاشت سطرها در نظر گرفته‌اند. هر ستون در مدل رابطه‌ای کلید اصلی، کلید خارجی یا یک ستون ساده است. روش آن‌ها نیز به‌جز هنگامی که کلید اصلی بیش از یک صفت باشد، مشابه روش‌های پیشین است. در واقع وجه تمایز این کار با کارهای پیشین در تعریف قانون برای حالتی است که کلید اصلی مرکب باشد، زیرا در کارهای پیشین، فقط تأکید بر کلید اصلی ساده بود (البته غیر از جداولی که ارتباطات چندتایی و چندبه‌چند را بیان می‌کنند). اگر کلید اصلی مرکب باشد، آن‌گاه تمامی صفات تشکیل دهنده‌ی آن باید Not Null باشند ولی لزومی ندارد که هر یک از این صفات به تنهایی یکتا باشند، بلکه همه آن‌ها روی هم باید خاصیت یکتایی را داشته باشند.

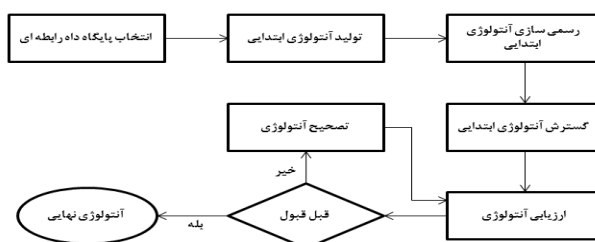
در این کار، کلاسی به نام کلاس کلید اصلی (مثلاً PK_Class) ایجاد می‌شود. هر ستون کلید اصلی در کلاس کلید اصلی تعریف می‌شود. همچنین برای برقراری ارتباط بین کلاس کلید اصلی و کلاس اولیه یک ویژگی شیء که دو کلاس را به یکدیگر مرتبط می‌کند، تعریف می‌شود. کلاس اولیه به‌عنوان دامنه و کلاس کلید اصلی به‌عنوان برد این ویژگی شیء است. همچنین برای اعمال قید

یکتایی روی کلید اصلی، ویژگی شیء تعریف شده باید تابعی معکوس^۱ باشد. به‌عنوان مثال شکل ۶ انتقال رابطه‌ی درس-استاد را که شامل کلید اصلی مرکب است، نشان می‌دهد.



شکل ۶: ترجمه شمای رابطه‌ای با استفاده از قانون کلید اصلی مرکب اطمینانی و همکاران

برای پیاده‌سازی این روش از چارچوب جِنَا^۲ استفاده شده است. همچنین از پایگاه داده رابطه‌ای برت^۳ به‌عنوان مطالعه‌ی موردی^۴ و برای اعتبارسنجی آنتولوژی استخراج شده نیز از استدلال‌گرهای^۵ Fact++ و Pellet از نرم‌افزار پروتگ^۶ استفاده شده است. ژو و همکاران (Zhou, 2011) روشی نیمه‌خودکار برای ساخت آنتولوژی با استفاده از شمای رابطه‌ای و گسترش آن با استفاده از مخزن دانش واژگانی وردنت ارائه کرده‌اند. ورودی این سیستم شمای رابطه‌ای و خروجی آنتولوژی OWL است. فرآیند ساخت آنتولوژی در شکل ۷ نشان داده شده است.



شکل ۷: چارچوب کلی فرآیند ساخت آنتولوژی ژو و همکاران

- 1 - Inverse Functional
- 2 - Jena
- 3 - Birt < [Http://www.eclipse.org/birt/phenix/db](http://www.eclipse.org/birt/phenix/db)>
- 4 - Case Study
- 5 - Reasoners
- 6 - Protégé

نویسندگان برای نگاشت شمای رابطه‌ای به آنتولوژی OWL، هفت قانون تعریف کرده‌اند که بسیار شبیه به قوانین تعریف شده در کارهای پیشین است. فرآیند گسترش آنتولوژی شامل دو گام است. نخست، افزودن مترادف‌هاست. طی این فرآیند نام هر موجودیت درون آنتولوژی با واژه‌های وردنت مقایسه می‌شود و از این طریق مترادف‌ها استخراج می‌شوند. اگر آنتولوژی موجود، فاقد مترادف‌های استخراج شده باشد، آن واژگان به آنتولوژی افزوده می‌شوند. دوم، افزودن سلسله‌مراتب‌هاست. اغلب مفاهیم حاصل از شمای رابطه‌ای با یکدیگر ارتباط معنایی دارند، به‌ویژه هنگامی که آنتولوژی از چند شمای رابطه‌ای استخراج می‌شود. به‌عنوان مثال آنتولوژی استخراجی می‌تواند همزمان دارای مفاهیم دانشجو و دانش‌آموز باشد. بنابراین در این گام با استفاده از وردنت، موجودیت‌های آنتولوژی با یکدیگر مقایسه می‌شوند و اگر دو موجودیت از سلسله‌مراتب واژگان در وردنت پیروی کنند و این سلسله‌مراتب در آنتولوژی موجود نباشد، آن‌گاه به آنتولوژی افزوده می‌شود.

از آنجایی که این روش نیمه‌خودکار است، مهندس آنتولوژی باید آنتولوژی حاصل را ارزیابی و تکمیل کند. برخی ارتباطات را نمی‌توان به‌طور خودکار استخراج کرد و بنابراین لازم است که توسط مهندس آنتولوژی به‌صورت دستی به آنتولوژی افزوده شوند. به‌عنوان مثال دو مفهوم دانشجو و اعضای هیأت علمی دارای ارتباطات درس یاد دادن و درس یاد گرفتن هستند (اعضای هیأت علمی به دانشجویان درس یاد می‌دهند و دانشجویان از اعضای هیأت علمی درس یاد می‌گیرند). بنابراین این ارتباطات باید به‌صورت دستی به آنتولوژی نهایی اضافه شوند.

روش‌های مبتنی بر تجزیه و تحلیل صفحات HTML

اغلب پایگاه داده‌های رابطه‌ای قابل دسترسی از طریق اینترنت، دارای صفحات HTML هستند که اگر تجزیه و تحلیل شوند، می‌توان از آن‌ها برای استخراج شمای پایگاه داده استفاده کرد. این موضوع، ایده‌ی اصلی روش‌هایی است که در این گروه جای می‌گیرند.

بن‌سلمان و همکاران (Benslimane, Malki, Amghar & Hassane, 2006)، از روش مهندسی معکوس برای تولید یک سیستم نیمه‌خودکار استفاده کرده‌اند که می‌تواند آنتولوژی OWL را از پایگاه داده رابطه‌ای به‌دست آورد. این روش معانی موجود در صفحات HTML را استخراج کرده و سپس از اطلاعات استخراج شده برای بهبود شمای رابطه‌ای استفاده می‌کند. در ادامه شمای

غنی شده به‌عنوان توصیف‌کننده منبع برای سیستم انتقال به‌منظور تولید آنتولوژی استفاده می‌شود. استفاده از صفحات HTML هنگامی مفید است که دسترسی به پایگاه داده‌های رابطه‌ای وجود نداشته باشد. از آنجایی که صفحات HTML به‌طور منظم نیاز به بازسازی دارند، پس هرگونه تغییر در ساختار صفحات HTML می‌تواند آنتولوژی ساخت شده را دچار مشکل کند.

در این کار، قوانین انتقال مدل رابطه‌ای به مدل آنتولوژی با قوانین انتقال مورد استفاده توسط لی و همکاران (Li et al., 2005)، یکسان و قوانین ساخت نمونه‌های آنتولوژی همان قوانین استفاده شده توسط استایونوویچ و همکاران (Stojanovic et al., 2002) است. این روش الگوریتم‌های زیادی دارد که نیاز به ارزیابی دارند. در این روش موتورهای استخراج و غنی‌سازی، اطلاعات اضافی زیادی تولید می‌کنند که می‌توان آن‌ها را به‌تنهایی از شمای رابطه‌ای به‌دست آورد. همچنین باید اشاره کرد که این روش از سربار زیادی رنج می‌برد، زیرا برای تولید آنتولوژی نیاز به استفاده از ۵ موتور است.

آستروا و همکاران (Astrova & Stantic, 2006) از تجزیه و تحلیل صفحات HTML برای تولید آنتولوژی به زبان Frame Logic استفاده کرده‌اند. در این روش صفحات HTML برای تولید مدل مفهومی ER تجزیه و تحلیل شده و سپس مدل استخراجی به آنتولوژی ترجمه می‌شود. این کار هیچ روشی برای استخراج سلسله‌مراتب وراثت که به‌عنوان یکی از مهم‌ترین جنبه‌ها در فرآیند ایجاد و گسترش آنتولوژی‌ها در نظر گرفته می‌شود، پیشنهاد نمی‌دهد. نویسندگان بیان کرده‌اند که برای استخراج قیود می‌توان از روش‌های یادگیری ماشین استفاده کرد، بدین شکل که روش یادگیری استنتاجی^۱ روی نمونه‌های به‌دست آمده از صفحات HTML اعمال می‌شود تا ارتباطات بین نمونه‌ها کشف شود. همچنین آنتولوژی استخراجی فاقد نمونه است، زیرا برای انتقال نمونه‌ها، قانونی ذکر نشده است.

1 - Inductive

روش‌های مبتنی بر تجزیه و تحلیل مدل‌های داده‌ای مفهومی

این ایده که پایگاه داده‌های رابطه‌ای و آنتولوژی یک مدل مفهومی دارند و در نتیجه ترجمه یک مدل مفهومی به دیگری آسان‌تر است و معانی کمتری از دست می‌رود، اساس شکل‌گیری روش‌های این گروه است.

شو و همکاران (Xu, Cao, Dong & Su, 2004)، روشی برای تبدیل شمای ER به آنتولوژی OWL بیان کرده‌اند که در آن قوانین سیستم انتقال به شکل رسمی توصیف شده است. در این کار ابزار خودکاری ارائه شده است که می‌تواند شمای ER در قالب XML را خوانده و تجزیه و تحلیل کند. در ادامه، این روش از قوانین انتقال به منظور ترجمه شمای ER به آنتولوژی OWL-DL استفاده می‌کند. آنتولوژی OWL-DL حاصل در دو فرمت انتزاعی^۱ و تبادل^۲ در دسترس خواهد بود. در این روش فقط کلیدها و صفات ساده در نظر گرفته می‌شود و کلیدها و صفات مرکب موجود در شمای ER در نظر گرفته نمی‌شود. در این روش به منظور دستیابی به آنتولوژی مورد نظر، از قوانینی برای انتقال شمای ER به نمایش انتزاعی OWL و سپس از این نمایش برای تولید نمایش تبدلی (RDF/XML) استفاده می‌شود. شمای ER ورودی برای سیستم انتقال توسط ابزار Power Design ایجاد شده است. این ابزار نمی‌تواند از همگی ویژگی‌ها و قابلیت‌های موجود در شمای ER استفاده کند. به عنوان مثال این ابزار از ارتباطات چندتایی که در آن‌ها n بزرگتر از ۲ باشد ($n > 2$) و صفت در ارتباط پشتیبانی نمی‌کند.

اوپادیا و همکاران (Upadhyaya & Kumar, 2005) هدفشان را استخراج آنتولوژی به زبان OWL از شمای EER ذکر کرده‌اند. در فرآیند ترجمه شمای EER به شمای رابطه‌ای، اطلاعات معنایی مشخصی از دست می‌رود. همچنین در فرآیند ترجمه مدل EER به مدل آنتولوژی اطلاعات معنایی بیشتری نسبت به ترجمه مدل رابطه‌ای به مدل آنتولوژی قابل استخراج است. مشکل اصلی این روش ناشی از عدم وجود، به روز نبودن و یا عدم دسترسی به مدل‌های داده‌ای مفهومی مانند EER در فرآیند ساخت پایگاه داده‌های رابطه‌ای است. این روش به طرز چشمگیری به کارشناس دامنه، برای تجزیه و تحلیل شمای EER به منظور استخراج اطلاعات معنایی بیشتر در راستای بهبود آنتولوژی حاصل وابسته است. همچنین در این روش صفات مرکب موجود در شمای EER در نظر

1 - Abstract

2 - Exchange

گرفته نمی‌شوند. در این روش، اگر صفتی کلید اصلی باشد، به ویژگی نوع داده با مشخصات تابعی و تابعی معکوس و محدودیت کاردینالیته حداقل ۱ نگاشت می‌شود. نگاشت کلید اصلی به ویژگی نوع داده همراه با مشخصه تابعی قابل قبول است ولی برخی ابزارها از مشخصه تابعی معکوس برای ویژگی نوع داده پشتیبانی نمی‌کنند. همچنین در این روش برای شناسایی وراثت یک‌گانه، چندگانه و تجمیع^۱ قوانین متمایز و درستی تعریف شده است.

نتیجه‌گیری

پس از بررسی روش‌های موجود برای انتقال پایگاه داده‌های رابطه‌ای به آنتولوژی، این روش‌ها براساس منبع ورودی به سه دسته تقسیم شدند. منبع ورودی برای پایگاه داده رابطه‌ای می‌تواند شمای رابطه‌ای، مدل‌های داده‌ای مفهومی مانند ER/EER و صفحات HTML باشد. در هر دسته به‌طور خلاصه به مهم‌ترین روش‌های موجود و کاستی‌هایشان پرداخته شد. در پایان ملاحظه می‌شود که شمای رابطه‌ای بهترین منبع برای تولید آنتولوژی از پایگاه داده رابطه‌ای، هنگامی که ساختار پایگاه داده در دسترس و مالک پایگاه داده همکاری لازم را داشته باشد، است. ساده‌ترین منبع قابل دسترسی که نیاز به همکاری مالک پایگاه داده ندارد، صفحات HTML مرتبط با پایگاه داده است. صفحات HTML ارتباط بین فرم‌های وب‌سایت و پایگاه داده را برقرار می‌کنند ولی نمی‌توان آن‌ها را به‌عنوان مدل واقعی پایگاه داده رابطه‌ای در نظر گرفت، در نتیجه نمی‌توان معانی نهفته در پایگاه داده‌های رابطه‌ای را از این منبع استخراج کرد. همچنین از نظر دسترسی به داده‌ها باید خاطر نشان کرد که بازیابی همه‌ی رکوردهای پایگاه داده رابطه‌ای از طریق وب‌سایت امری غیرممکن است. همچنین به‌طور کلی سه احتمال وجود دارد که به موجب آن‌ها ممکن است که مدل‌های داده‌ای مفهومی بیان‌گر پایگاه داده رابطه‌ای نباشد:

- ممکن است طی فرآیند مهندسی پایگاه داده، مدل مفهومی ایجاد نشود.
- ساخت مدل مفهومی همیشه نخستین گام در فرآیند ایجاد پایگاه داده است، بنابراین ممکن است این مدل بیان‌گر آخرین تغییرات و اصلاحات پایگاه داده نباشد. معمولاً تغییرات در سطح پیاده‌سازی، در مدل مفهومی منعکس نمی‌شود.

1 - Aggregation

- ممکن است تمرکز اصلی طراحان در فرآیند طراحی پایگاه داده، معطوف به ایجاد شمای فیزیکی باشد، بنابراین مدل مفهومی ممکن است بیانگر پیاده‌سازی واقعی پایگاه داده نباشد.
- ممکن است مدل مفهومی پس از طراحی و ایجاد پایگاه داده دیگر در دسترس نباشد.
- ممکن است ابزارهای مدل‌سازی مفهومی پایگاه داده از همه خصوصیات مدل‌های مفهومی پشتیبانی نمی‌کنند.

به‌علاوه به دلایل زیر نمی‌توان با استفاده از مدل مفهومی، یک آنتولوژی کامل تولید کرد:

- انواع داده‌ای به‌صورت صریح مشخص نشده است، بنابراین طراح باید آن‌ها را به‌صورت دستی اضافه کند. همچنین در این نوع شما، قیود **Not Null** و **Unique** وجود ندارد.
- برخی خصوصیات صفت مانند تقارنی، تراگذری و شمارشی در دسترس نیستند.
- هیچ محدودیت مقداری وجود ندارد. (به‌عنوان مثال سن شخص باید عدد مثبتی باشد).
- هیچ نمونه‌ای وجود ندارد، در نتیجه نمی‌توان یک پایگاه دانش ایجاد کرد.

بنابراین با توجه به معایب اشاره شده، می‌توان نتیجه گرفت که استفاده از مدل مفهومی پایگاه داده برای استخراج آنتولوژی مناسب نیست. در نتیجه نمی‌توان انتظار داشت که آنتولوژی استخراجی از مدل مفهومی غنی‌تر و کامل‌تر از آنتولوژی استخراجی از شمای رابطه‌ای باشد. با در نظر گرفتن مزایا و معایب مدل مفهومی و رابطه‌ای، در صورتی که هر دو شما در دسترس باشند، می‌توان از مزایا و معایب هر دو مدل برای تولید آنتولوژی استفاده کرد. بنابراین می‌توان ابتدا از شمای رابطه‌ای آنتولوژی استخراج کرد و سپس از آنتولوژی استخراج شده از مدل **ER/EER** مرتبط با شمای رابطه‌ای مورد نظر برای بهبود و یا ارزیابی آنتولوژی تولیدی به‌منظور به‌دست آوردن معانی اضافی یا از دست‌رفته استفاده نمود. خودکار یا نیمه‌خودکار بودن روش‌های بررسی شده یا توسط ارائه دهندگان روش ذکر شده یا از طریق به کار بردن این دو واژه در متن اصلی استنتاج شده است. لازم به ذکر است که هیچ‌گونه فاکتور(های) رسمی که به‌طور دقیق مرز بین خودکار یا نیمه‌خودکار بودن روش را مشخص کند، وجود ندارد و نویسندگان بسته به این که آنتولوژی استخراج شده به چه مقدار کار دستی برای تکمیل شدن نیاز دارد، روش را خودکار یا نیمه‌خودکار عنوان کرده‌اند. در جدول ۱، روش‌های بررسی شده به ترتیب زمانی، از قدیم به جدید مرتب شده است.

جدول ۱: ویژگی‌های عمومی روش‌های بررسی شده در یک نگاه

روش‌ها	سطح خودکار بودن	مدل		استفاده از مدل‌های میانی	پایه‌سازی شده	هدف اصلی
		منبع	مقصد			
استایانوویچ و همکاران	خودکار	شمای رابطه‌ای	RDFS	بلی	بلی	حاشیه‌نویسی صفحات پویای وب
شو و همکاران	خودکار	مدل ER	OWL	خیر	بلی	یادگیری آنتولوژی
لی و همکاران (۱)	خودکار	شمای رابطه‌ای	OWL	خیر	بلی	یادگیری آنتولوژی
لی و همکاران (۲)	نیمه‌خودکار	شمای رابطه‌ای	OWL	خیر	بلی	یادگیری آنتولوژی
اوپادیای و همکاران	نیمه‌خودکار	مدل EER	OWL	خیر	بلی	یادگیری آنتولوژی
آستروا و همکاران (۱)	نیمه‌خودکار	صفحات HTML	Frame Logic	بلی	خیر	یادگیری آنتولوژی
بن‌سلیمان و همکاران	نیمه‌خودکار	صفحات HTML	OWL	بلی	بلی	حاشیه‌نویسی صفحات پویای وب
آستروا و همکاران (۲)	خودکار	شمای رابطه‌ای	OWL	خیر	خیر	یادگیری آنتولوژی
اطمینانی و همکاران	خودکار	شمای رابطه‌ای	OWL	خیر	بلی	یادگیری آنتولوژی
ژو و همکاران	نیمه‌خودکار	شمای رابطه‌ای	OWL	خیر	خیر	یادگیری آنتولوژی

مراجع

شادگار ب.، عصاره ع. و هراتیان نژادی آ. (۱۳۸۹). وب معنایی: مفاهیم و تکنیک‌ها. تهران:

ارمغان.

Astrova I. and Kalja A. (2007). Rule-Based Transformation of SQL Relational Databases to OWL Ontologies, In The Proceedings of 2nd International Conference on Metadata & Semantic Research (CD-ROM), Ionian Academy, Corfu, Greece.

Astrova I. and Stantic B. (2006). An HTML Forms Driven Approach to Reverse Engineering of Relational Databases to Ontologies, In The Proceeding of the 23rd IASTED International Conference on Databases and Applications (DBA), 246-251, Innsbruck, Austria.

Benslimane S. M., Benslimane D., Malki M., Amghar Y. and Hassane H. S. (2006). Acquiring OWL Ontologies from Data-Intensive Web Sites, ICWE'06: In The Proceedings of the 6th International Conference on Web Engineering, 361-368, New York, USA.

Etminâni K., Kâhâni M. and Yânesârî N. R. (2009). Building Ontology from Relational Database, NDT2009: In The Proceeding of First International

- Conference on Networked Digital Technology, 555-557, Ostrava, Czech Republic.
- He B., Patel M., Zhang Z. and Chang K. (2007). Accessing the Deep Web, *Communications of the ACM*, 50(5), 94-101.
- Li M., Du X. and Wang S. (2005). Learning Ontology from Relational Database, In *The Proceedings of 4th International Conference on Machine Learning and Cybernetics (ICMLC'2005)*, 6, 3410-3415, Hong Kong, China.
- Li M., Du X. and Wang S. (2005). A Semi-automatic Ontology Acquisition Method for the Semantic Web, *Advances in Web-Age Information Management*, 3739, 209-220, Springer, Berlin, Germany.
- Stojanovič L., Stojanovič N. and Volz R. (2002). Migrating Data-Intensive Web Sites into the Semantic Web, In *Proceeding of 2002 ACM Symposium in Applied Computing*, 1100-1107, Madrid, Spain.
- Upadhyaya S. and Kumar P. (2005). ERONTO: A Tool for Extracting Ontologies from Extended E/R Diagrams, In *Proceedings of SAC'05 ACM Symposium on Applied Computing*, Santa Fe, New Mexico, USA.
- Xu Z., Cao X., Dong Y. and Su W. (2004). Formal Approach and Automated Tool for Translating ER Schemata into OWL Ontologies, *Lecture Notes in Computer Science*, 3056, 464-476, Springer, Berlin, Germany.
- Zhou X., Xu G.a and Liu L. (2011). An Approach for Ontology Construction based on Relational Database, *International Journal of Research and Reviews in Artificial Intelligence*, 1(1), 16-19.