میلانی فرد، امید؛ کاهانی، محسن (۱۳۹۷). ساخت هستان‌نگار از پایگاه داده‌های بزرگ: چالش‌ها و روشی برای غلبه بر آن‌ها. *پژوهشنامه کتابداری واطلاع‌رسانی*، ۸ (۱)، ۲۸۷- ۳۱۲.

ساخت هستان‌نگار از پایگاه داده‌های بزرگ: چالش‌ها و روشی برای غلبه بر آن‌ها

امید میلانی فرد[۱]، محسن کاهانی[۲]

تاریخ دریافت: ۱۳۹۶/۰۴/۳۱ تاریخ پذیرش: ۱۳۹۶/۱۲/۲۱ DOI: 10.22067/riis.v0i0.66255

چکیده

مقدمه: استخراج هستان‌نگار از پایگاه داده یکی از روش‌های متداول ساخت هستان‌نگار است. استفاده از روش‌های ارائه شده فعلی برای این کار بر روی پایگاه داده‌های بزرگ مانند پایگاه داده سامانه برنامه‌ریزی منابع سازمان (ERP) چالش‌هایی به همراه دارد. این پژوهش به‌دنبال ارائه راهکاری عملی برای غلبه بر این چالش‌ها است.

روش‌شناسی: این پژوهش با استفاده از روش پژوهش علم طراحی انجام شده و برای ارزیابی صحت و کارایی این روش، الگوریتم‌ها و فرآیند آن پیاده‌سازی و بر روی پایگاه داده یک نمونه ERP فعال در حوزه آموزش عالی آزمون شده است.

یافته‌ها: با استفاده از این روش، یک هستان‌نگار برای آموزش عالی از پایگاه داده یک ERP آموزش عالی ساخته شد. مقایسه این هستان‌نگار با سایر هستان‌نگارهای موجود آموزش عالی، نشان‌دهنده دقت و کارایی بالای این روش است.

نتیجه‌گیری: در نتیجه این پژوهش، روشی مبتنی بر مهندسی معکوس با جزئیات دقیق و کامل ارائه شده است. ابزارهای نرم‌افزاری این روش به‌صورت کد منبع باز، پیاده‌سازی شده و قابل ارتقاء و به‌کارگیری توسط سایر پژوهشگران می‌باشد. در روش پیشنهادی به هر دو مرحله آماده‌سازی و غنی‌سازی در فرآیند ساخت هستان‌نگار توجه شده و الگوریتم‌های تبدیل پایگاه داده به هستان‌نگار بر اساس نیازمندی‌های پایگاه داده‌های بزرگ بهینه‌سازی شده‌اند. همچنین معماری کامل و دارای جزئیات کافی از نقاط قوت این روش نسبت به سایر روش‌های موجود است.

کلیدواژه‌ها: هستان‌نگار، توسعه هستان‌نگار، مهندسی معکوس پایگاه داده.

۱. دانشجوی دکترای مدیریت فناوری اطلاعات، پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)، milanifard@students.irandoc.ac.ir

۲. استاد گروه کامپیوتر دانشگاه فردوسی مشهد، استاد مهمان پژوهشگاه علوم و فناوری ایران (ایرانداک) (نویسنده مسئول)، kahani@irandoc.ac.ir

مقدمه

برای رسیدن به درک مناسبی از یک حوزه[1] دانشی یا کسب و کار، باید مفاهیم موجود در آن و روابط معنایی میان این مفاهیم را شناخت. مدل‌سازی انتزاعی از یک حوزه به همین منظور انجام می‌شود. یکی از ابزارهای مدل‌سازی که در حیطه مهندسی دانش و فناوری اطلاعات کاربرد زیادی دارد، هستان‌نگار[2] است. گروبر (Gruber, 1995) هستان‌نگار را مفهومی می‌داند که به ایجاد فهم مشترک از یک حوزه اشاره داشته و شامل مجموعه‌ای از مفاهیم، روابط میان این مفاهیم، قواعد و نمونه‌هاست[3]. هستان‌نگارها بایستی به قالبی پیاده‌سازی شوند که توسط ماشین قابل خواندن و تفسیر باشند؛ به‌همین دلیل برای پیاده‌سازی هستان‌نگارها زبان‌های استانداردی مانند چارچوب توصیف منبع[4] (RDF) استفاده می‌شود. در حال حاضر زبان هستان‌نگار وب[5] (OWL) محبوبیت و استفاده گسترده‌تری از RDF دار.

کاربرد هستان‌نگارها بسیار متنوع است. از نظر آشچولد و گرونینگر (Uschold & Gruninger, 1996) سه کاربرد اساسی هستان‌نگار در بهبود ارتباطات[6]، افزایش قابلیت همکاری[7] و مهندسی سیستم‌ها است. کاربرد هستان‌نگار در حوزه‌های مختلف علم، فناوری و تجارت همچنان در حال توسعه است. به‌عنوان مثال می‌توان از هستان‌نگار برای مدل‌سازی مفهومی داده (Gailly and Poels, 2010) و یا در موتورهای جستجو برای غنی‌سازی متن پرس‌وجوهای کاربران (با اضافه کردن مترادف‌ها و واژگان مرتبط) و در نتیجه بهبود نتایج جستجو استفاده کرد (Paolucci, Kawamura, Payne and Sycara, 2002).

به‌رغم تمام مزایا، ساخت هستان‌نگار کاری پرهزینه و دشوار است. محققان در راستای کاهش هزینه‌های ساخت و همچنین افزایش دقت و کیفیت محصول نهایی، روش‌شناسی‌های گوناگونی معرفی کرده‌اند. در این روش‌شناسی‌ها سعی شده است الگویی نظام‌مند به‌صورت دستی یا نیمه خودکار برای ساخت ارائه شود. برای ساخت هستان‌نگار می‌توان از منابع مختلف دانشی استفاده کرد که می‌توان آن‌ها را به سه دسته منابع غیر ساخت‌یافته مانند متون زبان طبیعی، نیمه ساخت‌یافته مانند اسناد وب در قالب XML و

1. Domain
2. Ontology
3. Instance
4. Resource Description Framework
5. Web Ontology Language
6. Communication
7. Inter-Operability

HTML و ساخت‌یافته مانند پایگاه‌های داده‌های رابطه‌ای (RDB)[1]، تقسیم کرد (Al-Arfaj & Al-Salman, 2015).

امروزه RDB ها اصلی‌ترین منبع ذخیره و بازیابی اطلاعات در سازمان‌ها هستند و در نتیجه ساختار و نمونه‌های اکثر مفاهیمی که در حوزه کسب و کار سازمان وجود دارد به نحوی در پایگاه داده سیستم‌های عملیاتی و اطلاعاتی سازمان موجود است. سامانه برنامه‌ریزی منابع سازمان[2] (ERP) جامع‌ترین سامانه اطلاعاتی در سازمان است که تمامی بخش‌های عملیاتی و نظارتی سازمان را پوشش می‌دهد؛ به‌همین دلیل پایگاه داده آن، غنی‌ترین منبع دانش ساخت‌یافته برای استخراج هستان‌نگار آن حوزه کسب و کار است.

مسأله اصلی این پژوهش، مشخص نبودن چالش‌های ساخت هستان‌نگار از پایگاه داده‌های بزرگ و راهکارهای غلبه بر این چالش‌ها است. هدف از این پژوهش، ارائه راه‌کاری جامع برای استخراج هستان‌نگار یک حوزه از پایگاه داده ERP آن حوزه است. از آنجا که یک ERP باید داده‌های مربوط به اکثر موضوعات حوزه کسب و کار را ثبت و ضبط کند؛ پایگاه داده آن معمولاً شامل هزاران جدول اطلاعاتی است. استخراج هستان‌نگار از چنین پایگاه داده‌ای چالش‌های مختلفی دارد که در نتیجه این پژوهش باید برای غلبه بر آن‌ها راه‌حل مناسبی یافت شود.

پیشینه پژوهش

پژوهش‌های زیادی برای استخراج هستان‌نگار از RDB صورت گرفته است. مهمترین نقطه اشتراک این روش‌ها، استفاده از قوانینی است که برنرز-لی[3] در سال ۱۹۹۸ ارائه کرده است و به‌عنوان قوانین پایه‌ای نگاشت شناخته می‌شوند. این قوانین که برای تبدیل عناصر RDB به هستان‌نگار تعریف شده‌اند، عبارت‌اند از:

۱- تبدیل هر جدول R به یک کلاس C

۱. یک پایگاه داده رابطه‌ای (Relational Database) مجموعه‌ای از جداول داده‌ای است که می‌توانند با یکدیگر در ارتباط باشند. ساختار اصلی آن شامل جداول، ستون‌ها (فیلدها) و ردیف‌های (رکوردهای) جداول است. ارتباط بین جداول توسط کلیدهای خارجی (FK) و کلیدهای اصلی (PK) مشخص می‌شود.

2. Enterprise Resource Planning
3. Berners-Lee

۲- تبدیل هر رکورد از جدول R به یک گره RDF[1] از نوع C

۳- تبدیل تمام خصوصیات جدول به خصوصیات RDF

۴- برای هر رکورد جدول، تبدیل مقدار هر خصوصیت به خصوصیت معادل در گره RDF

در پژوهش‌های انجام گرفته، دو رویکرد کلی برای استخراج هستان‌نگار قابل شناسایی است:

۱- ساخت هستان‌نگار با استفاده از طرح‌واره پایگاه داده[2] (RDBS): پژوهش‌های دارای این رویکرد به‌دنبال روشی برای ساخت هستان‌نگار تنها با استفاده از RDBS هستند و به محتوای جداول و سایر فراداده‌ها از جمله فرهنگ اصطلاحات از پیش موجود توجهی ندارند.

از روش‌های مهم ارائه شده با این رویکرد می‌توان به DataMaster (Nyulas, Connor and Tu, 2007)، RDB2ONT (Trinh, Barker and Alhajj, 2006) و Automapper (Fisher, Dean and Joiner, 2008) اشاره کرد.

۲- ساخت هستان نگار خاص دامنه[3]: در این رویکرد به‌جای اینکه اول هستان‌نگار ساخته شده و سپس بر اساس دانش حوزه، پالایش شود؛ از همان ابتدا به مرتبط بودن عناصر آن به حوزه توجه می‌شود. برای این کار می‌توان از دو منبع دانش حوزه یعنی منبع خبرگان حوزه و محتوای پایگاه داده استفاده کرد. بر همین اساس این رویکرد به دو دسته زیر تقسیم می‌شود:

۲-۱- بدون استفاده از مهندسی معکوس: در این رویکردها ابتدا یک گراف RDF از محتوای پایگاه داده تهیه و به کارشناسان خبره حوزه عرضه می‌شود؛ تا نگاشت معانی به‌صورت دستی انجام گیرد. از ابزارهای معروف این رویکرد می‌توان به Tether (Byrne, 2008)، (Ling and Zhou, 2013)، RDB2RDF (Thuy, Thuan, Han, Park and Lee, 2014)، RDB2Onto (Octaviani, Pranolo and Othman, 2015) اشاره کرد.

۲-۲- با استفاده از مهندسی معکوس: ابزارهای معرفی شده بخش قبل روی کاربر انسانی به‌عنوان منبع دانش حوزه تکیه کرده و با فراهم کردن امکانات اولیه تبدیل (غالباً بر اساس قوانین پایه‌ای نگاشت برنرز-

---

۱. چارچوب توصیف منبع (Resource Description Framework) مدلی استاندارد برای تبادل داده بر روی وب است که از آن می‌توان برای مدل‌سازی فراداده استفاده کرد. این استاندارد را می‌توان پدر زبان‌های جدید توصیف هستان‌نگار مانند OWL به‌حساب آورد.

۲. طرح‌واره پایگاه داده (Relational DataBase Schema)، یک ساختار کالبدی است که دیدگاه منطقی به کل پایگاه داده را بازنمون می‌کند. این ساختار شامل تعاریف جداول، فیلدها و قیود مختلف روی فیلدها و جداول است.

3. Domain-Specific

لی) به کاربر در تعریف نگاشت کمک می‌کنند. در رویکردهای مهندسی معکوس به خود پایگاه داده به‌عنوان منبع اصلی دانش توجه شده و این منبع با دانش‌های اختیاری دیگر از جمله منابع خارجی (مانند هستان‌نگارهای موجود و پرس و جو[1]های پایگاه داده) و افراد خبره تکمیل می‌شود. از جمله روش‌های پیشنهادی با این رویکرد می‌توان به (Shen, Huang, Zhu and Zhao, 2006)، DB2OWL (Ghawi and Cullot, 2007)، (Tirmizi, Sequeda and Miranker, 2008)، RDBToOnto (Cerbah, 2008)، SQL2OWL (Alalwan, Zedan and Siewe, 2009)، DM-2-OWL (Albarrak and Sibley, 2009)، (Lubyte and Tessaris, 2009)، (Astrova, 2009)، (Liu, Wang, Bao and Wang, 2010)، OWLminer (Santoso, Haw and Abdul-Mehdi, 2011)، R2O (Khan and Sonia, 2011)، (BLOBELcd, 2014)، (Kaulins and Borisov, 2014) و (Zarembo, 2015) اشاره کرد.

مسأله اصلی این پژوهش ساخت هستان‌نگار خاص دامنه است؛ بنابراین رویکرد دوم مورد توجه است. در این رویکرد هم تمرکز اصلی بر روش‌های مهندسی معکوس است؛ زیرا استفاده از روش‌های غیر مهندسی معکوس بر روی پایگاه داده‌های بزرگ مانند ERP، مستلزم ساخت گراف‌های بسیار بزرگ است که بررسی آن‌ها را توسط خبرگان دشوار و زمان‌بر می‌سازد (Spanos, Stavrou, & Mitrou, 2012).

پر استفاده‌ترین قوانین تبدیل در رویکرد مهندسی معکوس عبارت‌اند از (Spanos, Stavrou, & Mitrou, 2012):

۱- قوانین اولیه: این قوانین، برگرفته از قانون‌های پایه برنرز-لی هستند که برای ایجاد یک خروجی متناسب با OWL[2] تغییر یافته‌اند. به‌صورت خلاصه جدول به کلاس[3]، فیلد غیر کلید خارجی به خصوصیت داده[4]، فیلد کلید خارجی به خصوصیت شیء[5] و رکوردهای جداول به نمونه‌های کلاس نگاشت می‌شوند.

۲- قوانین روابط دوتایی: این قوانین جداولی که بتوان آنها را به‌شکل خصوصیت شیء نگاشت کرد، شناسایی می‌کنند. کلید اصلی چنین جداولی مرکب از کلید خارجی به دو جدول دیگر بوده و فیلد دیگری

1. Query

۲. زبان هستان‌نگار وب (Web Ontology Language) زبان استانداردی برای بازنمایی هستان‌نگارها است که به‌صورت قابل توجهی از زبان‌های قدیمی RDF و RDFS کاربردی‌تر می‌باشد.

3 .Class

۴. خصوصیت داده یکی از عناصر هستان‌نگار است که ویژگی‌های یک یا چند مفهوم (کلاس) را مشخص می‌کند.

۵. خصوصیت شیء، یکی از عناصر هستان‌نگار است که روابط معنایی میان مفاهیم را مشخص می‌کند.

به غیر از این کلیدهای خارجی ندارد. یکی از جداول مورد ارجاع، دامنه و دیگری برد[1] این خصوصیت شیء خواهند بود.

۳- قوانین سلسله مراتب: زمانی‌که دو جدول با کلیدهای اصلی از طریق کلید خارجی با هم متصل شده باشند، یک رابطه پدر-فرزندی شکل می‌گیرد که به‌شکل فراکلاس[2] و زیر کلاس[3] نگاشت می‌شود.

۴- قوانین موجودیت‌های وابسته: این قوانین موجودیت‌هایی که جزئی از یک موجودیت دیگر هستند، شناسایی می‌کنند. یک موجودیت وابسته در RDB، یک جدول دارای کلید اصلی مرکب است که کلید خارجی به جدولی دیگر (جدول مالک) دارد. این موجودیت‌ها به کلاس نگاشت می‌شوند با این حال معنای یافتن رابطه بین کلاس وابسته و کلاس مالک دشوار است. برخی روش‌ها پیشنهاد کرده‌اند که چنین روابطی به‌صورت خصوصیت شیء «بخشی – از»[4] نامگذاری شوند.

۵- قوانین روابط چندتایی: روابط چندتایی در مواردی اتفاق می‌افتد که کلید اصلی یک رابطه، ترکیبی از کلیدهای خارجی به بیش از یک جدول باشد. از آنجا که در OWL روابط چندتایی وجود ندارد، این روابط به روابط دوتایی شکسته می‌شوند.

۶- قوانین تجمیع: با این قوانین موجودیت‌هایی شناسایی می‌شوند که به‌صورت عمودی قطعه قطعه شده‌اند؛ به‌عبارت دیگر جداولی که کلید اصلی یکسان داشته و هر یک، بخشی از اطلاعات یک موجودیت واحد را دارند. این جداول باید با هم تجمیع شده و به یک کلاس نگاشت شوند.

۷- قوانین محدودیت: این قوانین بر اساس قیود قابل تعریف در دستورات ساخت SQL بوده و آنها را به قیود OWL ترجمه می‌کنند. (محدودیت‌هایی مانند یکتایی یا غیرقابل تهی بودن برای یک فیلد)

۸- قوانین نوع داده: برای تبدیل نوع داده‌های SQL به نوع داده‌های OWL به‌کار می‌روند.

استفاده از قوانین تبدیل به‌تنهایی نمی‌تواند منجر به تولید هستان‌نگار مطلوب شود؛ زیرا این قوانین بر پایه فرضیاتی چون در قالب نرمال سوم[5] بودن پایگاه داده، وجود نام‌های معنادار برای جداول و خصوصیات و مشخص بودن تعاریف تمام کلیدهای اصلی و خارجی در RDBS عمل می‌کنند که در عمل این فرضیات به‌خصوص در سامانه‌های اطلاعاتی بزرگ معتبر نیستند. همچنین هستان‌نگار ساخته شده بر اساس مهندسی معکوس بهتر است با کمک گرفتن از سایر منابع دانشی دیگر مانند هستان‌نگارهای موجود

1.Range
2. Super Class
3. Sub Class
4. Is Part of
5. 3NF

حوزه، تکمیل شود. به دلایل ذکر شده، پژوهشگران پیشنهاد کرده‌اند عملیات استخراج هستان‌نگار به‌صورت یک فرآیند سه مرحله‌ای شامل آماده‌سازی، استخراج و غنی‌سازی انجام شود (Idrissi, Baïna and Baïna, 2014).

فهرست روش‌های مهندسی معکوس و پشتیبانی آن‌ها از قوانین و مراحل فرآیند ذکر شده، در جدول شماره ۱ نشان داده شده است. همان‌گونه که در این جدول مشاهده می‌شود، تعداد کمی از روش‌ها به مراحل آماده‌سازی و غنی‌سازی توجه کرده‌اند.

جدول شماره ۱. قوانین مورد استفاده و مراحل استخراج در برخی روش‌های مهندسی معکوس

روش	مراحل فرآیند استخراج			شماره قانون تبدیل							
	آماده‌سازی	استخراج	غنی‌سازی	۱	۲	۳	۴	۵	۶	۷	۸
(Shen et al., 2006)		X		X	X	X			X	X	X
DB2OWL (Ghawi and Cullot, 2007)		X		X	X	X				X	X
(Tirmizi, Sequeda and Miranker, 2008)		X		X	X	X		X		X	X
RDBToOnto (Cerbah, 2008)	X	X	X	X	X	X					
SQL2OWL (Alalwan, Zedan and Siewe, 2009)		X		X	X	X		X	X	X	X
(Lubyte and Tessaris, 2009)		X		X	X			X		X	
DM-2-OWL (Albarrak and Sibley, 2009)		X		X	X	X				X	X
(Astrova, 2009)		X	X	X	X	X	X	X		X	X
(Liu et al., 2010)		X		X	X			X		X	
OWLminer (Santoso, Haw and Abdul-Mehdi, 2011)	X	X		X	X	X			X	X	X
R2O (Khan and Sonia, 2011)		X		X	X	X	X	X		X	
(Blobel et al., 2014)		X		X	X				X		
(Kaulins and Borisov, 2014)		X	X	X	X	X	X	X	X	X	X
(Zarembo, 2015)		X		X	X	X				X	X

در مرحله آماده‌سازی لازم است فراداده مناسبی برای پایگاه داده ایجاد شود. برای ساخت چنین فراداده‌ای باید جداول و فیلدها به‌صورت مناسبی برچسب‌گذاری و روابط بین جداول که در کدهای برنامه نهفته هستند، شناسایی شوند (Idrissi, Baïna and Baïna, 2014). این جنبه از آماده‌سازی را که به تکمیل دانش در خصوص عناصر پایگاه داده (جداول و فیلدها و روابط آن‌ها) اشاره دارد با عنوان «کامل بودن» در نظر می‌گیریم. از آنجا که در این پژوهش ساخت هستان‌نگار خاص حوزه مدنظر است، باید در مرحله آماده‌سازی، مرتبط بودن موجودیت‌های پایگاه داده (جداول و فیلدها) با حوزه مورد نظر مشخص شود. این جنبه از مرحله آماده‌سازی را که مشخص کردن عناصر مرتبط و غیر مرتبط با حوزه توجه دارد، «مرتبط

بودن» می‌خوانیم. همان‌طور که در جدول ۱ دیده می‌شود، دو روش RDBToOnto و OWLMiner به مرحله آماده‌سازی پایگاه داده توجه داشته‌اند. مقایسه این دو روش از نظر دو جنبه «کامل بودن» و «مرتبط بودن» مرحله آماده‌سازی، در جدول شماره ۲ نشان داده شده است. روشRDBToOnto تعریف محدودیت به‌صورت دستی را برای مشخص کردن ارتباط موجودیت‌های پایگاه داده با حوزه مورد نظر پیشنهاد داده است. انجام چنین کاری بدون ابزارهای کمکی برای تشخیص میزان ارتباط یک موجودیت با حوزه مورد نظر برای پایگاه داده ERP کاری زمان‌بر و دشوار خواهد بود که در این روش به چنین ابزاری اشاره نشده است. همچنین RDBToOnto برای ساخت فراداده و جنبه کامل بودن مرحله آماده‌سازی، طرحی ارائه نداده است. همان‌طور که اشاره شد نداشتن فراداده مناسب به‌خصوص برای پایگاه داده‌های بزرگ، کار استخراج هستان‌نگار را با چالش جدی مواجه می‌کند. از سوی دیگر روش OWLminer بدون توجه به مرتبط بودن موجودیت‌های پایگاه داده با حوزه هستان‌نگار، تنها به شناسایی روابط سلسله مراتبی بین جداول با توجه به محتوای آن‌ها توجه داشته است و این شکل از رابطه جداول را به‌عنوان فراداده ثبت و در مرحله استخراج مورد استفاده قرار می‌دهد.

جدول ۲. مقایسه روش‌های دارای مرحله آماده‌سازی

روش	مرتبط بودن	کامل بودن
RDBToOnto	تعریف محدودیت به‌صورت دستی توسط کاربر	
OWLminer		شناسایی روابط سلسله مراتبی با دانش پیش‌زمینه

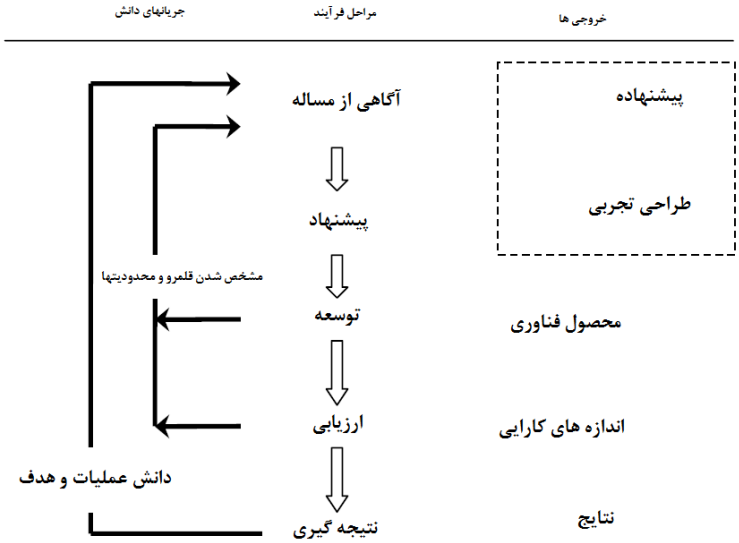
با توجه به این موارد ذکر شده، هیچکدام از روش‌های پیشنهادی برای غلبه بر چالش‌های استخراج هستان‌نگار از پایگاه داده‌های بزرگی چون ERP، راه حل کامل و مناسبی ارائه نداده‌اند. در ادامه این نوشتار روش‌شناسی این پژوهش بیان می‌شود و در بخش سوم، روشی عملی برای ساخت هستان‌نگار از RDB معرفی و در خصوص آن بحث می‌شود. در بخش چهارم پیاده‌سازی و ارزیابی عملی این روش مورد بررسی قرار گرفته و نتایج حاصل از به‌کارگیری روش پیشنهادی ارائه می‌گردد.

روش‌شناسی

از آنجا که خروجی مورد نظر این پژوهش روشی برای تولید یک محصول[1] فناوری اطلاعات (هستان‌نگار یک حوزه) از روی یک محصول دیگر فناوری اطلاعات (پایگاه داده یک ERP) است،

1. Artifact

رویکرد انجام این پژوهش، پژوهش علم طراحی[1] (DSR) انتخاب شده است. روش‌شناسی عمومی DSR در شکل ۱ نشان داده شده است (Vaishnavi and Kuechler, 2015).



شکل ۱. روش‌شناسی عمومی DSR

بر اساس این مدل عمومی، ابتدا با بررسی پیشینه و روش‌های موجود مشخص شد که راه حل مناسبی برای مسأله اصلی پژوهش یعنی استخراج هستان‌نگار از پایگاه داده ERP وجود ندارد. بر اساس الگوی تقسیم و حل[2] در DSR (Vaishnavi and Kuechler, 2015) چالش‌هایی که در استفاده از روش‌های فعلی بر روی پایگاه داده ERP وجود داشتند شناسایی و در چهار دسته تولید فراداده اولیه، قوانین تبدیل، غنی‌سازی و ارزیابی قرار گرفتند. در چرخه «پیشنهاد – توسعه – ارزیابی» برای چالش‌های هر دسته، با مشورت متخصصان راه‌حل‌هایی پیشنهاد شدند که در نهایت با ادغام این راه‌حل‌ها روش مناسب مهندسی معکوس پایگاه داده‌های بزرگ برای ساخت هستان‌نگار ایجاد گردید. این روش بخشی از یک روش‌شناسی کامل ساخت هستان‌نگار با عنوان انتیرانداک[3] است که در آن از رویکرد سه سویه سازی در منابع دانشی و روش‌های ساخت، برای بالا بردن قابلیت اطمینان و کیفیت محصول نهایی (هستان‌نگار هدف) استفاده شده است (میلانی‌فرد و کاهانی، ۱۳۹۶). روش‌شناسی انتیرانداک از سه منبع دانشی شامل

1. Design Science Research
2. Divide and Conquer pattern
3. Ontirandoc

هستان‌نگارهای موجود، مستندات غیر ساخت‌یافته مرتبط با حوزه و پایگاه داده عملیاتی مرتبط با حوزه استفاده می‌کند. به ازای هر یک از این منابع دانشی، روش و ابزار متناسب آن طراحی و در قالب یک سامانه یکپارچه برای انجام کامل فرآیند ساخت هستان‌نگار، پیاده‌سازی و آزمون شده است.

برای ارزیابی روش پیشنهادی، از الگوی اثبات عملی[1] در DSR استفاده شد (Vaishnavi and Kuechler, 2015). در الگوی اثبات عملی باید راه حل در عمل پیاده‌سازی و اجرا شود. به‌همین منظور پیمانه‌های[2] نرم‌افزاری روش پیشنهادی پیاده‌سازی شدند. سامانه دانشگاه فردوسی مشهد (سدف)[3] به‌عنوان نمونه مورد مطالعه انتخاب گردید و با استفاده از روش پیشنهادی پیاده‌سازی شده، هستان‌نگار حوزه آموزش و پژوهش از آن استخراج شد. در بخش بعد روش پیشنهادی به همراه شرح چالش‌ها و راه حل آنها بیان شده است.

یافته‌ها

در روش پیشنهادی استخراج هستان نگار از پایگاه داده، فرآیند ۴ مرحله‌ای آماده‌سازی، استخراج، غنی‌سازی و صحت سنجی در نظر گرفته شده است. معماری این روش در شکل ۲ مشاهده می‌شود[4]. در این پژوهش بر روی مراحل آماده‌سازی و استخراج به روش مهندسی معکوس تمرکز دارد و در ادامه به تفضیل در این خصوص بحث خواهد شد. مراحل غنی‌سازی و صحت سنجی در روش‌شناسی کامل انتیرانداک (میلانی فرد و کاهانی، ۱۳۹۶) و پس از استخراج هستان‌نگار از سه منبع دانشی (پایگاه داده،

1. Demonstration pattern
2. Module

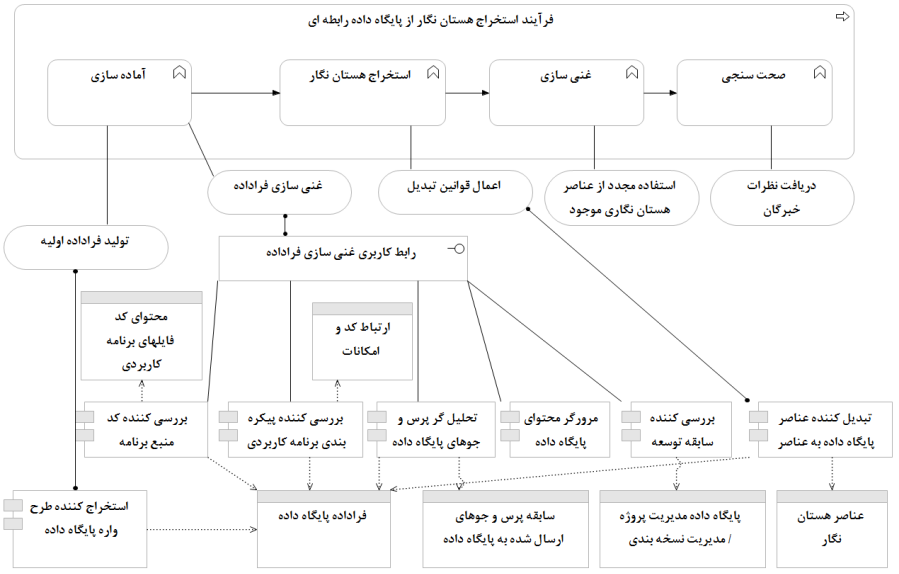
۳. این سامانه یکپارچه فرآیندهای گوناگون حوزه آموزش عالی را در بخش‌های مختلف آموزشی، پژوهشی، دانشجویی و اداری و مالی، پشتیبانی می‌کند. سدف در بیش از ۳۰ دانشگاه و مرکز آموزش عالی کشور نصب و در حال استفاده است. تراکنش‌های مربوط به فرآیندهای تحت پوشش سدف در یک پایگاه داده بزرگ شامل بیش از ۳۰۰۰ جدول اطلاعاتی ثبت و ضبط می‌شود. به‌دلیل گستردگی پوشش فرآیندهای حوزه آموزش عالی و استفاده از آن در مؤسسات آموزش عالی مختلف، سدف نمونه مناسبی از یک ERP برای آزمون روش پیشنهادی است. اطلاعات بیشتر در خصوص این سامانه از http://www.dpm.ir/Sadaf/About قابل دریافت است.

۴. این شکل با زبان توصیف معماری آرچیمیت از زبان‌های توصیف معماری ذکر شده در استاندارد ISO/IEC/IEEE 42010 ترسیم شده است.

مستندات متنی و هستان‌نگارهای موجود) اجرا می‌شوند و به‌همین دلیل در این نوشتار به اختصار مورد اشاره قرار می‌گیرند.

ایده اصلی مرحله آماده‌سازی، ساخت یک فراداده کامل و مرتبط با حوزه برای پایگاه داده است. برای انجام این کار دو سرویس کاربردی طراحی شده است:

۱- سرویس تولید فراداده پایه: این سرویس کاربردی با مراجعه به RDBS، ساختار اولیه مورد نیاز شامل نام و مشخصات جداول و فیلدها را استخراج و در پایگاه فراداده ثبت می‌کند.



شکل ۲. معماری روش مهندسی معکوس پایگاه داده ERP به هستان نگار

۲- سرویس غنی‌سازی فراداده: این سرویس کاربردی با فراهم ساختن یک رابط کاربری و سه پیمانه نرم‌افزاری پشتیبان به کاربران امکان می‌دهد با استفاده از دانش پیش زمینه، فراداده اولیه را تکمیل کنند. اطلاعات تکمیلی با کمک این سرویس شامل برچسب‌گذاری معنایی هر جدول و فیلد، ثبت ارتباطات بین جداول، تعیین حوزه‌های کسب و کار مرتبط با جدول و همچنین وضعیت استفاده از جداول و فیلدها می‌باشد. پیمانه‌های نرم‌افزاری پشتیبان بر اساس انواع دانش پیش زمینه موجود به شرح زیر طراحی شده‌اند:

- بررسی‌کننده کد منبع برنامه کاربردی: یکی از منابع مهم دانش در مورد پایگاه داده، کد برنامه کاربردی است که می‌تواند نقص مربوط به نامگذاری ناقص و نامفهوم جداول و فیلدها را تا حدی

جبران کند (Idrissi, Baïna and Baïna, 2014). بسیاری از ارتباطات بین جداول و معنای کدهای مختلف ذخیره شده در فیلدها در لایه منطق برنامه‌های کاربردی قرار دارند. با بررسی متون مرتبط مشخص می‌شود در این زمینه پژوهش چندانی صورت نگرفته است؛ مرتبط‌تری پژوهش در این حوزه توسط ژائو، چنگ و دیلون (Zhao, Chang & Dillon, 2008) انجام شده که پیشنهاد می‌دهد در مهندسی معکوس RDBS برای رسیدن به هستان‌نگار از دانش مستخرج از کد منبع برنامه تحت وب استفاده شود. این پیمانه یک راه حل عملی برای چنین کاری ارائه می‌دهد. با استفاده از این پیمانه کاربران می‌توانند با انتخاب یک جدول، محتوای قطعاتی از کد که در آن به جدول مورد نظر ارجاعاتی وجود دارد، مشاهده کنند و با خواندن آن بخش از کد به معنای واقعی جدول و فیلدهای آن و همچنین ارتباطات آن با سایر جداول پی ببرند و این معنا را در قالب برچسب زدن به جدول یا فیلدها، تعریف کلیدهای خارجی جدید و یا مشخص کردن عنوان مقادیر مجاز فیلدها در فراداده ثبت کند.

- بررسی‌کننده پیکره‌بندی برنامه کاربردی: سامانه‌های عملیاتی مجموعه امکانات خود را از طریق منوهای[1] مختلف در دسترس کاربران قرار می‌دهند. مقصود از پیکره‌بندی در اینجا، ارتباط بین هر بخش از کد با منوها یا امکانات سامانه است. این پیکره‌بندی اغلب در پایگاه داده مربوط به سامانه ذخیره شده است. این پیمانه در تعامل با پیمانه بررسی‌کننده کد منبع، این امکان را فراهم می‌کند تا کاربران بتوانند برای یک جدول خاص، فهرست منوها یا امکاناتی که به نحوی به آن جدول ارجاع دارند، مشاهده کنند. علاوه بر اینکه عنوان یک منو به‌تنهایی می‌تواند برای معنابخشی به‌کاربرد یا ماهیت وجودی یک جدول مفید باشد، کاربر می‌تواند با مراجعه به منوهای مربوطه در برنامه کاربردی و مشاهده خروجی برنامه (فرم‌های ورود اطلاعات یا گزارشات مختلف)، به دانش بیشتری در خصوص جدول و فیلدهای آن دست یابد.
- تحلیل‌گر پرس و جوهای پایگاه داده: تقریباً تمامی سامانه‌های اطلاعاتی بزرگ مانند ERP، مکانیزم‌هایی برای ثبت سابقه[2] پرس و جوهای ارسال شده به پایگاه داده دارند. این پرس و جوها منبع بسیار خوبی برای شناسایی ارتباط بین جداول هستند. همان‌گونه که در قبل گفته شد، برخی از ارتباطات بین جداول تنها در لایه منطق برنامه کاربردی تعریف می‌شوند و در RDBS نشانی از آن‌ها وجود ندارد. این پیمانه با تحلیل پرس و جوهای ارسال شده، ارتباطات بین جداول و شرایط آن‌را

1. Menu
2. log

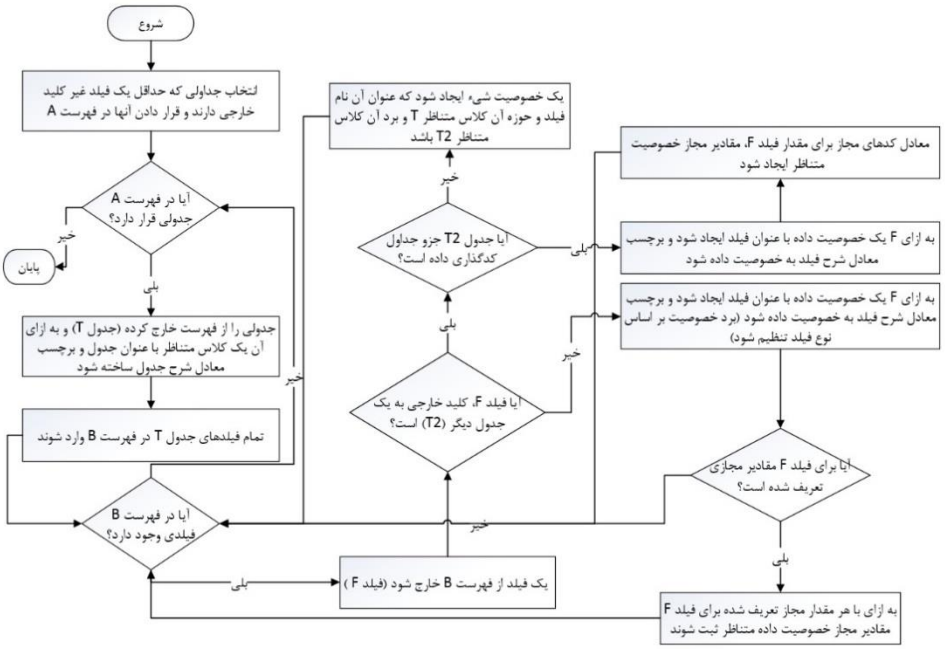
شناسایی و به‌عنوان کلیدهای خارجی بالقوه به کاربر پیشنهاد می‌دهد تا در صورت تأیید به‌عنوان یک اتصال کلید خارجی در فراداده ثبت شود.

- مرورگر محتوای پایگاه داده: در خصوص استنتاج معنای جداول بر اساس تحلیل ماشینی محتوای RDB کارهای اندکی صورت گرفته که از آن میان می‌توان به OWLminer و RDB2Onto اشاره کرد؛ با این وجود برای پایگاه داده‌های بسیار بزرگ مانند ERP، اجرای الگوریتم‌های پیشنهادی کارایی لازم را ندارد. این پیمانه به کاربران امکان می‌دهد محتوای جداول را مشاهده کنند و امکاناتی برای فیلتر کردن اطلاعات بر اساس شرایط مختلف فراهم می‌آورد. آنچه کاربر از محتوای جداول برداشت می‌کند به تسلط او به حوزه بستگی دارد. به‌دلیل مستقل بودن از سایر پیمانه‌ها می‌توان در آینده با توسعه الگوریتم‌های کارآمد، این پیمانه را به حالت نیمه–خودکار ارتقاء داد.
- بررسی‌کننده سابقه توسعه کد: یکی از منابع دانش پیش زمینه که در ادبیات موجود به آن اشاره‌ای نشده است، پایگاه داده سامانه‌های کنترل نسخه[1] و مدیریت پروژه است. برای توسعه سامانه‌های اطلاعاتی بزرگ معمولاً سامانه‌های پشتیبانی چون مدیریت پروژه و کنترل نسخه وجود دارند که حاوی اطلاعات ارزشمندی در خصوص بخش‌های مختلف سامانه از جمله کد و ساختار پایگاه داده هستند. چنانچه چنین پایگاه‌هایی در دسترس باشد می‌توان با استفاده از این پیمانه به سابقه کارهای مرتبط با یک جدول یا فیلد اطلاعاتی (دلایل ایجاد یا تغییر ساختار) دسترسی پیدا کرد و با نمایش این اطلاعات به کاربر امکان برچسب‌گذاری مناسب جدول یا فیلد را فراهم ساخت.

پس از مرحله آماده‌سازی، مرحله استخراج هستان‌نگار قرار دارد که در آن از سرویس اعمال قوانین تبدیل استفاده می‌شود. این سرویس توسط پیمانه نرم‌افزاری مربوطه، عناصر RDB را به عناصر هستان‌نگار تبدیل می‌کند. قوانین هشت‌گانه بیان شده در بخش پیشینه نظری در این تبدیل استفاده می‌شوند. نکته مهم در روش پیشنهادی استفاده از فراداده است. به این ترتیب که تنها جداول مرتبط با حوزه مورد نظر هستان‌نگار برای تبدیل انتخاب می‌شوند و در زمان اعمال قوانین ۲ تا ۶، علاوه بر کلید خارجی حاصل از RDBS به کلیدهای خارجی که با استفاده از سایر منابع دانشی کشف و در فراداده ثبت شده‌اند نیز توجه می‌شود. همچنین در قانون تبدیل محدودیت، علاوه بر مورد اخیر، فهرست‌های شمارشی (محدودیت مقادیر مجاز برای فیلد) نیز بر اساس فراداده شناسایی شده‌اند و به‌عنوان محدودیت خصوصیت به هستان‌نگار منتقل می‌شوند. نامگذاری عناصر هستان‌نگار بر اساس عنوان عناصر RDB است. اما

1. Version Control

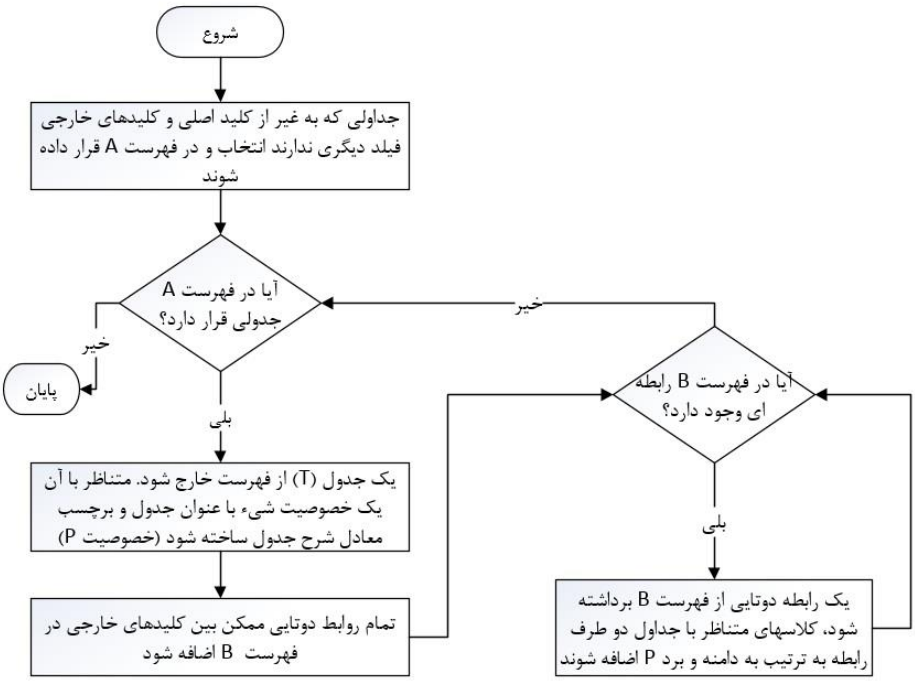
برچسب‌های معنایی ثبت شده در فراداده به هستان‌نگار منتقل، و به‌عنوان برچسب کلاس‌ها و خصوصیات ثبت می‌شوند.



شکل ۳. نمودار گردش اعمال قوانین پایه، محدودیت و نوع داده

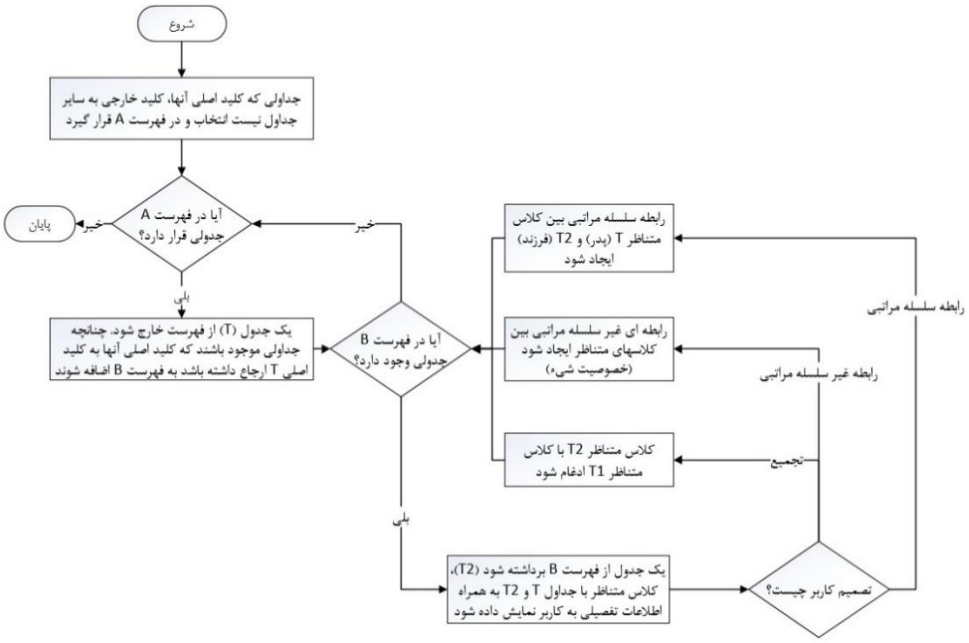
شکل ۳ نمودار گردش[1] طراحی شده برای اعمال قوانین پایه، محدودیت و نوع داده را نشان می‌دهد. از جمله نکات منحصر بفرد در این طراحی، توجه به جداول کدگذاری داده‌ها در سامانه‌های بزرگی مانند ERP است. این نوع خاص از جداول معمولاً برای نگهداری معنای کدهای مختلف از مقادیر فیلدها (مقادیر محدود و شمارشی) مورد استفاده قرار می‌گیرند و نباید به‌عنوان یک خصوصیت شیء در نظر گرفته شوند.

1. Flow chart



شکل ٤. نمودار گردش اعمال قوانین روابط دوتایی و چندتایی

در شکل ۴ نمودار گردش طراحی شده برای اعمال قوانین روابط دوتایی و روابط چندتایی نشان داده شده است. برای اعمال قوانین سلسله مراتب و تجمیع، یک رویکرد جدید طراحی شده است.



شکل ۵. نمودار گردش اعمال قوانین سلسله مراتب و تجمیع

همان‌گونه که در نمودار شکل ۵ مشاهده می‌شود، جداولی که ممکن است با هم یک مفهوم را تشکیل دهند و یا رابطه پدر – فرزندی داشته باشند، شناسایی، و به کاربر نمایش داده می‌شود تا با مداخله کاربر تجمیع یا ساخت سلسله مراتب انجام شود.

پس از انجام مراحل آماده‌سازی و استخراج، دو مرحله غنی‌سازی و ارزیابی قرار دارد. به‌صورت خلاصه مرحله غنی‌سازی از ترکیب دانش استخراج شده با عناصر هستان‌نگار مستخرج از دو منبع دانش دیگر استفاده می‌کند و در مرحله ارزیابی، با استفاده از تبدیل عناصر هستان‌نگار به نوشتار متنی قابل تفسیر توسط خبرگان و تجمیع و اعمال نظرات آن‌ها، اعتبار محصول نهایی تأیید می‌شود. جزئیات تفصیلی این دو مرحله در روش‌شناسی انتیرانداک (میلانی فرد و کاهانی، ۱۳۹۶) بیان شده است.

بحث و نتیجه‌گیری

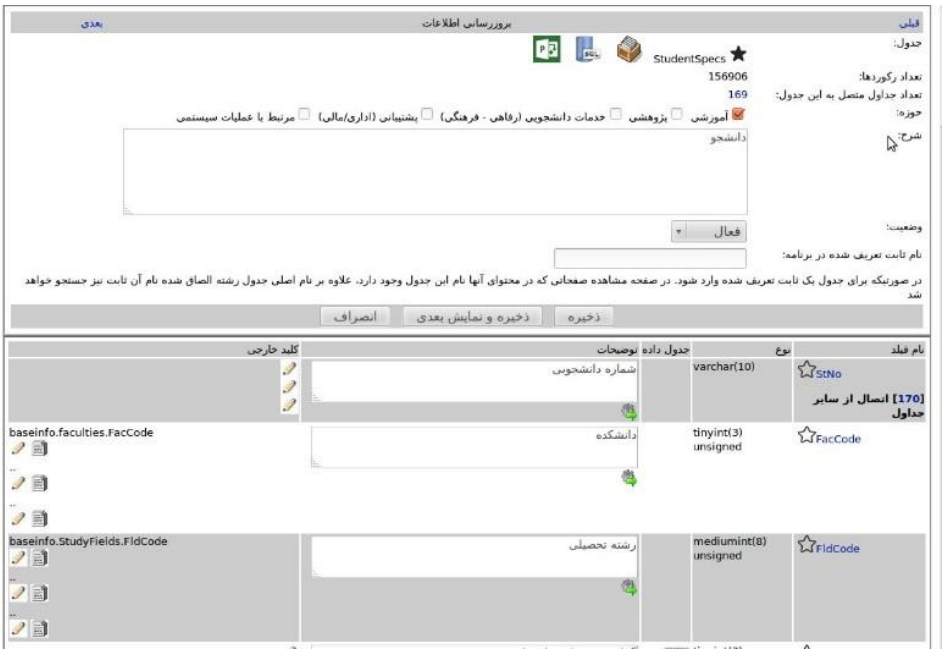
در این بخش ابتدا ابزار توسعه یافته برای روش پیشنهادی به‌صورت مختصر شرح داده شده و سپس در مورد هستان‌نگار حاصل از اجرای روش بر روی سدف (نمونه ERP آموزش عالی) و ارزیابی آن بحث می‌شود.

پیمانه استخراج‌کننده فراداده اولیه دارای یک رابط کاربری برای انتخاب پایگاه داده‌های مورد نظر است. از آنجا که ERP مورد مطالعه به‌صورت مداوم در حال توسعه و تغییر می‌باشد، این پیمانه به‌صورت زمان‌بندی شده و خودکار هر روز اجرا شده و تغییرات ساختاری جداول اطلاعاتی پایگاه داده را به فراداده منتقل می‌کند.

فراداده استخراج شده توسط رابط کاربری غنی‌سازی فراداده، قابل پالایش است. در این رابط کاربری می‌توان بر اساس شرایط انتخابی، جداولی را انتخاب و فراداده آن‌ها را مشاهده و ویرایش کرد. کاربر با کلیک روی عنوان هر جدول می‌تواند به فراداده آن وارد شده، آن‌را تکمیل و اصلاح کند. در فرم ویرایش فراداده جدول (شکل 6)، پیمانه‌های نرم‌افزاری مختلف به کاربر در شناسایی بهتر مفهوم و معنای جدول و فیلدهای آن کمک می‌کنند.

کاربر می‌تواند با کلیک روی آیکون‌های مربوطه در جلوی عنوان جدول موارد زیر را مشاهده کند:

۱- مشاهده محتوای قابل فیلتر جدول: نمایش داده‌ها که به‌صورت عمودی و افقی قابل فیلتر باشند، این امکان را برای کاربر فراهم می‌کند تا با بررسی محتوای جدول به دانش بهتری در خصوص معنای جدول و فیلدهای آن دست پیدا کند.



شکل ٦. ویرایش فراداده جدول و فیلدهای آن

۲- مشاهده پرس و جوهای ارسال شده از سوی زیرسامانه‌های مختلف ERP به پایگاه داده که محتوی نام جدول انتخابی هستند. با استفاده از این امکان، کاربر می‌تواند به دانشی در خصوص کاربرد اطلاعات آن جدول در سامانه‌های اطلاعاتی دست پیدا کند.

۳- بررسی سابقه و محتوای کدهای سامانه: با استفاده از این امکان کاربر می‌تواند فهرست فایل‌هایی که در محتوای آن‌ها به جدول انتخابی اشاره شده است، مشاهده کند. با توجه به مشخص بودن ارتباط بین این فایل‌ها و منوهای سدف، کاربر می‌تواند با مراجعه به منوی مربوطه از کاربرد عملی جدول آگاه شود.

در صفحه ویرایش فراداده، جدول، کاربر می‌تواند فراداده فیلدهای جدول را تنظیم کند. از مهمترین اطلاعات مورد نیاز برای استخراج هستان‌نگار، ارتباط بین جداول است. ابزار پیاده‌سازی شده به کاربر کمک می‌کند ارتباطات ثبت نشده در فراداده اولیه را با سرعت بیشتری پیدا و ثبت کند. برای این کار کافی است کاربر روی آیکون مربوط به ویرایش کلید خارجی جدول کلیک کند تا پیشنهادهایی از سوی سیستم به او نشان داده شود. این پیشنهادها بر اساس مشابهت عنوان برچسب و عنوان فیلد می‌باشد. همچنین فیلدهایی که بیشترین ارجاع را داشته‌اند نیز برای انتخاب سریع‌تر نشان داده می‌شوند.

پس از تکمیل فراداده، کاربر می‌تواند با استفاده از یک رابط کاربری، قوانین تبدیل مهندسی معکوس را بر روی جداول حوزه‌های مورد نظر خود اجرا و هستان‌نگار اولیه را تشکیل دهد (شکل ۷). در زمان اجرای قوانین تبدیل، پیشنهادهای تجمیع کلاس و ادغام خصوصیات و روابط نیز تولید و ذخیره می‌شوند. کاربر می‌تواند با کلیک روی دکمه‌های مربوطه، این پیشنهادها را مشاهده و پس از بررسی، اعمال یا رد کند.

شکل ۷. اجرای قوانین تبدیل و استخراج هستان‌نگار

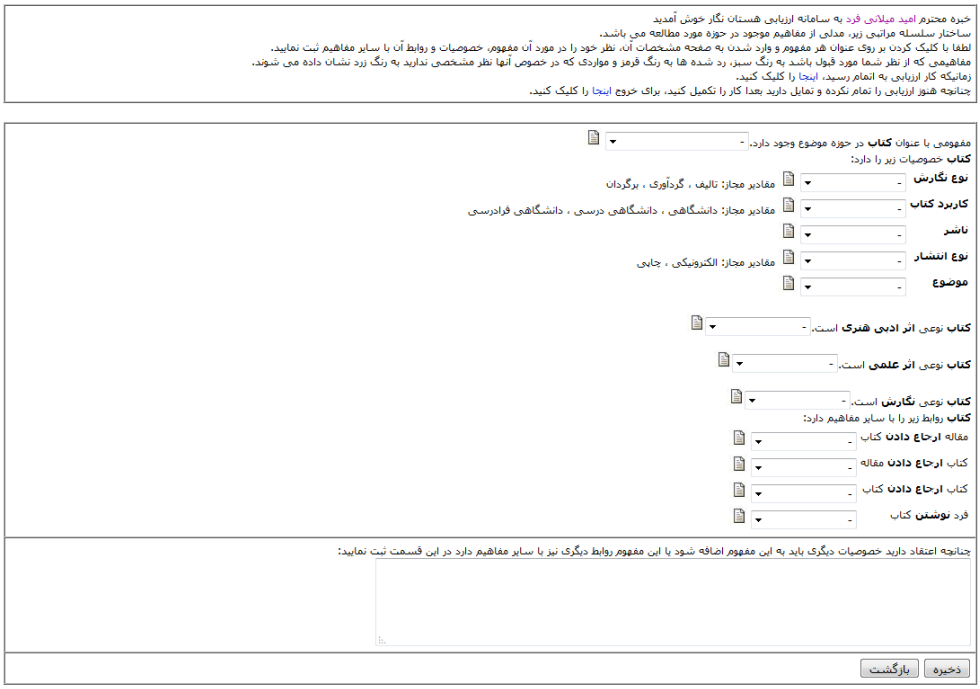
با اجرای قوانین تبدیل بر روی پایگاه داده سدف (در حوزه آموزش و پژوهش) و پس از اعمال پیشنهادهای تجمیع کلاس و ادغام خصوصیت، هستان نگار نهایی شامل ۱۵۶ کلاس و ۶۵۵ خصوصیت ایجاد شد.

برای ارزیابی این هستان‌نگار می‌توان از روش‌های گوناگون ذکر شده در پیشینه پژوهشی استفاده کرد. این روش‌ها در دسته‌های زیر تقسیم‌بندی می‌شوند (Brank, Grobelnik, & Mladenić, 2005):

۱. مقایسه با یک استاندارد طلایی[1]
۲. استفاده در یک کاربرد و ارزیابی نتایج حاصل
۳. مقایسه با یک منبع داده (به‌عنوان مثال یک مجموعه از مستندات) در مورد حوزه هستان‌نگار.
۴. بررسی انطباق با استانداردها و نیازمندی‌ها به‌وسیله خبرگان.

از آنجا که به‌صورت کلی ممکن است در حوزه مورد نظر هستان‌نگار، استاندارد طلایی یا منبع داده‌ای مناسب برای مقایسه در دسترس نباشد، در روش‌شناسی انتیرانداک که روش مهندسی معکوس پیشنهادی در این نوشتار، بخشی از آن‌را تشکیل می‌دهد، ارزیابی توسط خبرگان (رویکرد چهارم) مبنای اصلی ارزیابی قرار گرفته است. برای این کار عناصر هستان‌نگار تبدیل به گزاره‌هایی نزدیک به زبان طبیعی شده و در یک فرم الکترونیکی، از خبرگان در مورد صحت آنها سؤال می‌شود (شکل ۸) همچنین خبرگان می‌توانند یادداشت‌های اصلاحی برای هر عنصر ثبت کنند. عناصری که روی آنها اتفاق نظر وجود دارد در هستان‌نگار تثبیت شده و پس از اصلاحات لازم بر روی سایر عناصر، چرخه ارزیابی تکرار می‌شود تا در نهایت بر روی تمامی عناصر هستان‌نگار اتفاق نظر صورت گیرد.

1. Golden Standard



شکل ۸. صحت سنجی الکترونیکی هستان‌نگار توسط خبرگان

جهت ارزیابی هستان‌نگار حاصل از مهندسی معکوس سدف، به‌صورت خاص، از رویکرد اول و سوم نیز استفاده شد:

۱- مقایسه هستان‌نگار با استاندارد طلایی: به‌منظور شناسایی استاندارد طلایی، با روش گمانه‌زنی هستان‌نگار[1]، هستان‌نگارهای مرتبط با حوزه شناسایی شدند. از میان موارد یافت شده، هستان‌نگار LUBM[2] که به‌عنوان یک هستان‌نگار معیار حوزه دانشگاهی شناخته شده است (Gil, Borges, Contreras, & Martin-Bautista, 2009)، به‌عنوان استاندارد طلایی در نظر گرفته هستان‌نگار LUBM با روش نگاشت نیمه خودکار هستان‌نگار با سایر هستان‌نگارهای یافت شده و همچنین هستان‌نگار حاصل از این پژوهش مقایسه شد. در روش نگاشت نیمه خودکار، عناصر هستان‌نگار اول، توسط کاربر به عناصر هستان‌نگار دوم نگاشت می‌شوند و در فرآیند نگاشت، یک پیمانه نرم‌افزاری روش‌شناسی انتیرانداک به کاربر کمک می‌کند تا عناصر مشابه‌شکلی و معنایی را در هستان‌نگار دوم بیابد و در نتیجه دقت و سرعت فرآیند نگاشت را بالا می‌برد.

1. Ontology Dowsing (https://www.w3.org/wiki/Ontology_Dowsing)
2. Lehigh University Benchmark (http://swat.cse.lehigh.edu/projects/lubm)

جدول شماره ۳. مقایسه هستان نگارهای حوزه آموزش و پژوهش با LUBM

نام هستان‌نگار	درصد پوشش کلاس‌ها	درصد پوشش خصوصیات
هستان‌نگار حاصل از مهندسی معکوس	۹۷,۵	۹۶,۳
هستان‌نگار دانشگاه[1] (دانشگاه تورنتو)	۹۷,۵	۸۸,۸۹
هستان‌نگار مرجع آموزش عالی[2]	۷۲,۵	۷۰,۳۸
هستان‌نگار انجمن پژوهشی وب معنایی[3]	۶۷,۵	۵۹,۲۶
قالب اطلاعات پژوهشی مشترک اروپا[4]	۶۰	۵۱,۸
هستان نگار حوزه آموزشی و پژوهشی[5]	۵۵	۷۷,۷۸
هستان نگار دانشگاه[6]	۵۲,۵	۴۸,۱۵
سمات[7] (رووف نژاد، کاهانی و مهارتی، ۱۳۹۴)	۵۰	۵۱,۸۵

پس از نگاشت LUBM با سایر هستان‌نگارها، میزان پوشش مفاهیم این هستان‌نگار در سایر هستان‌نگارها مشخص شد که در جدول شماره ۳ نشان داده شده است. این میزان پوشش نشان‌دهنده دقت مفاهیم (دقت کلمات کلیدی انتخاب شده برای مفاهیم) در هستان‌نگار است (احمدی، عصاره، حسینی و حیدری، ۱۳۹۶). همان‌گونه که در جدول دیده می‌شود هستان‌نگار حاصل از مهندسی معکوس کامل‌ترین پوشش را دارد. تنها کلاس نگاشت نشده، کلاسی با عنوان کار[8] است که برای دسته‌بندی فعالیت‌های حوزه آموزش (تدریس و کار پژوهشی) استفاده شده است.

۲- مقایسه با منبع داده: کتاب تعاریف و مفاهیم آماری علوم، تحقیقات و فناوری[9] برای این کار در نظر گرفته شد. در این کتاب که حاصل صدها ساعت کار کارشناسی است، مفاهیم آماری مورد استفاده در حوزه آموزش عالی، شناسایی و توصیف شده‌اند. در ابتدا واژگان مهم مرتبط با حوزه آموزش و پژوهش از متن کتاب استخراج و سپس در قالب یک هستان‌نگار به‌صورت دستی مدل‌سازی شد. هستان‌نگار ساخته شده از روی منبع داده با هستان‌نگارهای دیگر یافت شده در این حوزه و همچنین هستان‌نگار حاصل از

1. http://www.cs.toronto.edu/semanticweb/maponto/MapontoExamples/univ-cs.owl
2. Higher Education Reference Ontology (HERO), http://sourceforge.net/projects/heronto/
3. Semantic Web Research Community (SWRC), http://swrc.ontoware.org/ontology
4. Common European Research Information Format (CERIF), http://www.dfki.de/~brigitte/CERIF/CERIF2004_1.1FDM/cerif.owl
5. http://vivoweb.org/files/vivo-isf-public-1.6.owl
6. http://www.webkursi.lv/luweb05fall/resources/university.owl

۷ سامانه ملی اطلاعات تحقیقاتی ایران (سمات)

8. Work

۹ کتاب تعاریف و مفاهیم آماری علوم، تحقیقات و فناوری، تدوین شده توسط کمیته آماربخشی موسسه پژوهش و برنامه‌ریزی آموزش عالی، ۱۳۸۷

مهندسی معکوس با روش نگاشت نیمه خودکار عناصر مقایسه شد که نتایج آن در جدول شماره ۴ نشان داده شده است.

جدول ٤. مقایسه هستان نگارهای حوزه آموزش و پژوهش با هستان‌نگار مستخرج از کتاب مفاهیم آماری علوم، تحقیقات و فناوری

نام هستان‌نگار	درصد پوشش کلاس‌ها	درصد پوشش خصوصیات
هستان‌نگار حاصل از مهندسی معکوس	۷۷٫۸۶	۶۵٫۲۷
هستان‌نگار دانشگاه (دانشگاه تورنتو)	۱۷٫۸۶	۴٫۱۹
هستان‌نگار مرجع آموزش عالی	۲۲٫۸۶	۲۰٫۳۶
هستان‌نگار انجمن پژوهشی	۱۶٫۴۳	۱۱٫۹۸
چارچوب مشترک اطلاعات پژوهشی اروپا	۱۷٫۸۶	۱۶٫۱۷
هستان نگار حوزه آموزشی و پژوهشی	۳۲٫۱۴	۱۸٫۵۶
هستان نگار دانشگاه	۱۰٫۷۱	۱۰٫۷۸
سمات	۲۴٫۲۹	۱۶٫۱۷

همان‌گونه که در جدول ۴ مشاهده می‌شود، هستان‌نگار حاصل از مهندسی معکوس، به میزان قابل توجهی پوشش بالاتری از مفاهیم را دارد. اغلب مفاهیمی که توسط هستان‌نگار حاصل از مهندسی معکوس پوشش داده نشده‌اند را می‌توان در سه دسته طبقه‌بندی کرد:

- مفاهیمی که برای گروه‌بندی مفاهیم دیگر استفاده می‌شوند: مفاهیمی مانند نهاد پژوهشی، نهاد علمی، کمیته، فناوری، سند و ... . این‌گونه مفاهیم به‌صورت مستقیم از روش مهندسی معکوس، قابل استخراج نیستند زیرا در پایگاه داده وجود ندارد. برای شناسایی چنین مواردی یک بخش بهبوددهنده به انتهای روش مهندسی معکوس اضافه شد که با بررسی خصوصیات و روابط کلاس‌های شناسایی شده، دسته‌بندی‌هایی به کاربر پیشنهاد می‌دهد تا در صورت تمایل به هستان‌نگار اضافه کند. از جمله دسته‌های شناسایی شده با کمک این افزونه می‌توان به کلاس نگارش (کلاس پدر برای کلاس‌های کتاب، مقاله، گزارش فنی و طرح پژوهشی) اشاره کرد.
- مفاهیمی که بالاتر از سطح یک دانشگاه و در سطح کلان آموزش عالی مطرح هستند و به‌همین دلیل قابل استخراج از پایگاه داده یک سامانه دانشگاهی نیستند. از جمله این مفاهیم می‌توان به هیأت امنا و آزمون کاردانی به کارشناسی اشاره کرد.

- مفاهیمی که در پایگاه داده سدف پیاده‌سازی نشده‌اند. با وجود گستردگی پشتیبانی سدف از فرآیندهای دانشگاهی، برخی موارد همچون پارک علم و فناوری و ارتقای رتبه علمی هنوز در این سامانه پیاده‌سازی نشده‌اند و در نتیجه قابل استخراج از پایگاه داده آن نیستند.

همان‌گونه که توصیف شد، استفاده از روش مهندسی معکوس پیشنهادی این نوشتار، منجر به تولید محصولی با دقت بالا شده است. علاوه بر این می‌توان به مزایای دیگر استفاده از روش پیشنهادی برای استخراج هستان‌نگار به شرح زیر اشاره کرد:

۱- این روش به‌صورت عملیاتی پیاده‌سازی شده و کد منبع آن باز[1] است. این ویژگی به پژوهش‌گران اجازه می‌دهد با اعمال تغییرات مورد نظر خود در کد برنامه آن‌را برای کاربردهای دیگر سفارشی‌سازی[2] کنند.

۲- معماری این روش به‌صورت پیمانه بندی[3] است. هر پیمانه به‌صورت مجزا و بدون تأثیر بر سایر اجزاء معماری قابل تکمیل و بهینه‌سازی است. به‌عنوان مثال «بررسی‌کننده کد منبع برنامه» در حال حاضر به‌شکل نیمه خودکار عمل می‌کند و تنها محتوای کد صفحات مرتبط با جدول را به کاربر نشان می‌دهد. این پیمانه با انجام پژوهش‌های آینده در خصوص تجزیه و تحلیل کد قابل ارتقاء می‌باشد.

۳- از آنجا که ارتباط بین عناصر پایگاه داده و هستان‌نگار حاصل توسط پیمانه اجرای الگوریتم‌های تبدیل ذخیره‌سازی می‌شود، می‌توان ادعا کرد رویکرد دوم در ارزیابی هستان‌نگار (استفاده در یک کاربرد و ارزیابی نتایج حاصل) به‌صورت ضمنی حین اجرای این روش انجام می‌شود. زیرا کاربرد هر یک از اجزاء هستان‌نگار در یک نمونه عملی (پایگاه داده عملیاتی) مشخص و قابل ردگیری می‌باشد.

۴- هستان‌نگار استخراج شده با روش مهندسی معکوس به‌دلیل استفاده از یک پایگاه داده عملیاتی، دارای جزئیات بسیار بیشتری نسبت به سایر هستان‌نگارها می‌باشد و از این نظر پوشش بالاتری نسبت به فضای واقعی حوزه مورد نظر ارائه می‌دهد. این تفاوت در سطح جزئیات در جدول شماره ۵ برای هستان‌نگار حوزه آموزش و پژوهش نشان داده شده است. به‌عنوان مثال کلاس دانشجو در هستان‌نگار حاصل از مهندسی معکوس دارای ۳۰ خصوصیت داده و ۳۲ رابطه مستقیم با سایر مفاهیم، و ۵۱ خصوصیت داده و ۲۳ رابطه با سایر مفاهیم از طریق ارث بری از کلاس پدر (کلاس فرد) است؛ به‌عبارت دیگر یک نمونه از کلاس دانشجو در این هستان‌نگار ۸۱ خصوصیت داده و ۵۵ رابطه با سایر نمونه‌ها دارد، در

1. Open Source
2. Customize
3. Modular

حالی‌که بالاترین سطح جزئیات کلاس دانشجو در سایر هستان نگارها مربوط به هستان‌نگار مرجع آموزش عالی است که دارای ۱۱ خصوصیت داده و ۱۷ رابطه با سایر مفاهیم است.

جدول ۵. مقایسه هستان نگارهای حوزه آموزش و پژوهش از نظر تعداد عناصر

نام هستان‌نگار	تعداد کلاس	تعداد خصوصیت
هستان‌نگار حاصل از مهندسی معکوس	۱۵۶	۶۵۵
هستان‌نگار دانشگاه (دانشگاه تورنتو)	۵۲	۳۴
هستان‌نگار مرجع آموزش عالی	۵۶	۱۴۲
هستان‌نگار انجمن پژوهشی	۵۴	۶۶
چارچوب مشترک اطلاعات پژوهشی اروپا	۲۰۷	۱۸۳
هستان نگار حوزه آموزشی و پژوهشی	۳۸۴	۱۶۳
هستان نگار دانشگاه	۶۹	۴۴
سمات	۱۵۸	۱۵۹

با وجود کامل‌تر بودن هستان‌نگار حاصل از اجرای روش مهندسی معکوس نسبت به سایر هستان‌نگارهای موجود، برای رسیدن به یک هستان‌نگار مرجع نیاز به استفاده همزمان از منابع اطلاعاتی و روش‌های دیگر وجود دارد. زیرا تمام مفاهیم در پایگاه داده، مدل نشده و مورد استفاده قرار نگرفته‌اند. به‌همین منظور هم روش‌شناسی انتیرانداک سه سویه سازی منبع و روش را پیشنهاد داده است (میلانی فرد و کاهانی، ۱۳۹۵).

فهرست منابع

احمدی حمید، عصاره فریده؛ حسینی بهشتی، ملوک السادت و حیدری، غلامرضا (۱۳۹۶). طراحی سامانه نیمه خودکار ساخت هستی‌شناسی به کمک تحلیل هم‌رخدادی واژگان و روش C-value (مطالعه موردی: حوزه علم‌سنجی ایران). *پژوهشنامه پردازش و مدیریت اطلاعات*. ۱ (۳۳)، ۱۸۷-۲۱۸.

رووف نژاد، مرضیه؛ کاهانی، محسن و مهارتی، یعقوب. (۱۳۹۴). *طراحی الگوی هوش تجاری مبتنی بر آنتولوژی داده‌های پژوهشی (مورد مطالعه دانشگاه فردوسی مشهد)*. پایان‌نامه کارشناسی ارشد، دانشگاه فردوسی مشهد.

کمیته آماربخشی (۱۳۸۷). *کتاب تعاریف و مفاهیم آماری علوم، تحقیقات و فناوری*، تهران: موسسه پژوهش و برنامه‌ریزی آموزش عالی

میلانی فرد، امید و کاهانی، محسن (۱۳۹۶). انتیرانداک: سامانه یکپارچه توسعه مشارکتی هستان‌نگار فارسی، *پژوهشنامه پردازش و مدیریت اطلاعات*. تاریخ پذیرش: (۱۳۹۶/۱۵/۰۵).

Alalwan, N., Zedan, H., & Siewe, F. (2009, October). Generating OWL ontology for database integration. In *Advances in Semantic Processing, 2009. SEMAPRO'09. Third International Conference*, (pp. 22-31). IEEE.

Al-Arfaj, A., & Al-Salman, A. (2015). Ontology construction from text: challenges and trends. *International Journal of Artificial Intelligence and Expert Systems (IJAE)*, *6*(2), 15-26.

Albarrak, K. M., & Sibley, E. H. (2009, August). Translating relational & object-relational database models into OWL models. In *Information Reuse & Integration, 2009. IRI'09. IEEE International Conference on* (pp. 336-341). IEEE.

Astrova, I. (2009). Rules for mapping SQL relational databases to OWL ontologies. In *Metadata and Semantics* (pp. 415-424). Springer US.

Berners-Lee, T. (1998). Relational databases and the semantic web (in design issues). *World Wide Web Consortium*.

BLOBELcd, B. (2014, May). Conceptual model formalization in a semantic interoperability service framework: Transforming relational database schemas to OWL. In *PHealth 2014: Proceedings of the 11th International Conference on Wearable Micro and Nano Technologies for Personalized Health, 11–13 June 2014, Vienna, Austria* (Vol. 200, p. 35). IOS Press.

Brank, J., Grobelnik, M., & Mladenić, D. (2005). A survey of ontology evaluation techniques.

Byrne, K. (2008, September). Having triplets-Holding cultural data as RDF. In *Proceedings of the ECDL 2008 Workshop on Information Access to Cultural Heritage*.

Cerbah, F. (2008, December). Mining the content of relational databases to learn ontologies with deeper taxonomies. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01* (pp. 553-557). IEEE Computer Society.

El Idrissi, B., Baïna, S., & Baïna, K. (2014). A methodology to prepare real-world and large databases to ontology learning. In *Enterprise Interoperability VI* (pp. 175-185). Springer International Publishing.

Fisher, M., Dean, M., & Joiner, G. (2008, April). Use of OWL and SWRL for Semantic Relational Database Translation. In *OWLED (spring)*.

Gailly, F., & Poels, G. (2010). Conceptual modeling using domain ontologies: Improving the domain-specific quality of conceptual schemas. *10th Workshop on Domain-Specific Modeling* (p. 18). ACM.

Ghawi, R., & Cullot, N. (2007, September). Database-to-ontology mapping generation for semantic interoperability. In *VDBL'07 conference, VLDB Endowment ACM* (pp. 1-8).

Gil, R., Borges, A. M., Contreras, L., & Martin-Bautista, M. J. (2009, March). Improving ontologies through ontology learning: a university case. In *Computer Science and Information Engineering, 2009 WRI World Congress on* (Vol. 4, pp. 558-563). IEEE.

Gruber, T. R. 1995. "Toward principles for the design of ontologies used for knowledge sharing." *International journal of human-computer studies, 43(5)* 907-928.

Kaulins, A., & Borisov, A. (2014). Building Ontology from Relational Database/Ontoloģiju izveide no relāciju datubāzes/Построение онтологии по реляционной базе данных. *Information Technology and Management Science*, *17*(1), 45-49.

Khan, S., & Sonia, K. (2011). R2o: Relation to ontology transformation system. *Journal of Information & Knowledge Management*, *10*(01), 71-89.

Ling, H., & Zhou, S. (2013). Mapping relational databases into owl ontology. *International Journal of Engineering and Technology (IJET)*, *5*(6), 4735-4740.

Liu, P., Wang, X., Bao, A., & Wang, X. (2010, November). Ontology automatic constructing based on relational database. In *Grid and Cooperative Computing (GCC), 2010 9th International Conference on* (pp. 412-415). IEEE.

Lubyte, L., & Tessaris, S. (2009, August). Automatic extraction of ontologies wrapping relational data sources. In *International Conference on Database and Expert Systems Applications* (pp. 128-142). Springer Berlin Heidelberg.

Nyulas, C., O'Connor, M., & Tu, S. (2007, July). DataMaster–a plug-in for importing schemas and data from relational databases into Protege. In *10th international Protégé conference* (pp. 15-18).

Octaviani, D., Pranolo, A., & Othman, S. (2015, October). RDB2Onto: an approach for creating semantic metadata from relational educational data. In *Science in Information Technology (ICSITech), 2015 International Conference on* (pp. 137-140). IEEE.

Paolucci, M., Kawamura, T., Payne, T. R., & Sycara, K. (2002, June). Semantic matching of web services capabilities. In *International Semantic Web Conference* (pp. 333-347). Springer Berlin Heidelberg.

Santoso, H. A., Haw, S. C., & Abdul-Mehdi, Z. T. (2011). Ontology extraction from relational database: Concept hierarchy as background knowledge. *Knowledge-Based Systems*, *24(3)*, 457-464.

Shen, G., Huang, Z., Zhu, X., & Zhao, X. (2006, November). Research on the Rules of Mapping from Relational Model to OWL. In *OWLED*.

Spanos, D. E., Stavrou, P., & Mitrou, N. (2012). Bringing relational databases into the semantic web: A survey. *Semantic Web*, *3(2)*, 169-209.

Thuy, P. T. T., Thuan, N. D., Han, Y., Park, K., & Lee, Y. K. (2014, January). RDB2RDF: completed transformation from relational database into RDF ontology. In *Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication* (p. 88). ACM.

Tirmizi, S. H., Sequeda, J., & Miranker, D. (2008, September). Translating sql applications to the semantic web. In *International Conference on Database and Expert Systems Applications* (pp. 450-464). Springer Berlin Heidelberg.

Trinh, Q., Barker, K., & Alhajj, R. (2006, February). Rdb2ont: A tool for generating owl ontologies from relational database systems. In *Telecommunications, 2006. AICT-ICIW'06. International Conference on Internet and Web Applications and Services/Advanced International Conference on* (pp. 170-170). IEEE.

Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The knowledge engineering review*, *11(2)*, 93-136.

Vaishnavi, V. K., & Kuechler, W. (2015). *Design science research methods and patterns: innovating information and communication technology*. Crc Press.

Zarembo, I. (2015, June). Automatic Transformation of Relational Database Schema into OWL Ontologies. In *Environment. Technology. Resources. Proceedings of the International Scientific and Practical Conference* (Vol. 3, pp. 217-222).

Zhao, S., Chang, E., & Dillon, T. (2008, July). Knowledge extraction from web-based application source code: An approach to database reverse engineering for ontology development. In *Information Reuse and Integration, 2008. IRI 2008. IEEE International Conference on* (pp. 153-159). IEEE.