
بسم الله الرحمن الرحيم

دانشگاه فردوسی مشهد
دانشکده مهندسی - گروه مهندسی کامپیوتر

پایان‌نامه کارشناسی‌ارشد

پرکردن خودکار فرم‌های وب با استفاده از وب داده

نگارنده

محبوبه دادخواه

استاد محترم راهنما

جناب آقای دکتر محسن کاهانی

شهریور ۱۳۹۰

تقدیر و تشکر

از زحمات بی‌دریغ، راهنمایی‌های ارزنده و همراهی‌های ارزشمند استاد راهنما جناب آقای دکتر محسن کاهانی کمال تشکر را دارم.

همچنین از تمام اعضای آزمایشگاه WTLab[1]، به علت همکاری‌هایشان سپاس‌گزاری می‌کنم.

[1] Web Technology Laboratory

چکیده

فرم‌های وب اصلی‌ترین روش برای دسترسی به حجم قابل توجهی از اطلاعات در وب عمیق هستند. کاربران فرم‌های وب را برای جستجوی این اطلاعات و یا ثبت‌نام در وب‌سایت‌هایی همانند سایت‌های اجتماعی استفاده می‌کنند. پر کردن فرم یک فرآیند تکراری است و بعضی از داده‌های استفاده شده در این فرآیند، ایستا هستند. فرآیند پر کردن را می‌توان با استفاده از تکنولوژی معنایی برای ذخیره‌ی داده‌هایی که کاربر قبلا در فرم‌ها پر کرده و یا برای پیشنهاد مقادیری در پر کردن فرم‌های جدید توسط وب داده بهینه نمود. در این رساله، یک چارچوب برای پر کردن خودکار فرم با استفاده از داده‌های منتشر شده به صورت داده‌های پیوندی برروی وب، ارائه شده‌است. هدف اصلی در این رساله، استفاده از تکنولوژی‌های معنایی برای پر کردن خودکار فرم‌های وب جدید بر اساس وب داده و فرم‌هایی که کاربر قبلا پر کرده‌است، می‌باشد. چارچوب پیشنهادی از یک روش مبتنی بر آنتولوژی به عنوان روش نگاشت استفاده می‌کند. بدین جهت، مفاهیم استفاده شده در دامنه‌های مختلف فرم استخراج شده‌است. ابتکار کلیدی در این چارچوب، استفاده از داده‌های پیوندی به عنوان یک منبع مفید برای فراهم کردن داده در پر کردن فرم‌ها می‌باشد. اگرچه فرآیند پیشنهادی نیاز به میزان کمی از تعامل کاربر دارد، بازخوردهای کاربر در مورد هر فیلد بلافاصله استفاده می‌شود تا مقادیر درستی را برای پر کردن فیلدهای دیگر این فرم و نیز فرم‌های جدید فراهم گردد. نتایج تجربی برروی مخزن فرم TEL8 نشان می‌دهد که در صورت وجود داده‌های پیوندی، استفاده از آن در حوزه‌های مختلف فرم می‌تواند فاز پیشنهاد داده در فرآیند پر کردن را بهبود بخشد. استفاده از وب داده در نه حوزه‌ی مختلف، یک تلاش چالش برانگیز و خلاقانه است که در این چارچوب مورد توجه قرار گرفته‌است. یافته‌ها نشان می‌دهند که داده‌های پیوندی باز کنونی یک منبع مفید در ساختن برنامه‌های کاربردی حوزه‌های مختلف می‌باشد. نتایج ارزیابی نشان می‌دهند که روش پیشنهادی امکانپذیر و موثر است و نتایج راضی کننده می‌باشند.

کلید واژه: پر کردن خودکار فرم، نگاشت مبتنی بر آنتولوژی، داده‌های پیوندی، تکنیک‌های معنایی، پیشنهاد داده، تاریخچه‌ی کاربر

فهرست مطلب‌ها

فهرست جدول‌ها

فهرست شکل‌ها

فهرست اختصارات به‌کار رفته در متن

DBLP: Digital Bibliography and Library Project
FDO: Form Data Ontology
FOAF: Friend Of a Friend
HTML: Hyper Text Markup Language
linkedMDB: Linked Movie DataBase
LOD: Linked Open Data
OWL: Web Ontology Language
RDB: Relational Databases
RDF: Resource Description Framework
RDFS: Resource Description Framework Schema
SIOC: Semantically Interlinked Online Communities
SKOS: Simple Knowledge Organization System
SPARQL: Simple Protocol and RDF Query Language
XML: Extensible Markup Language

فصل ۱– مقدمه

۱–۱– مقدمه

امروزه کاربردهای تحت وب نیاز به میزان زیادی از ارتباط متقابل با کاربر دارند. حجم بسیار زیادی از داده‌هایی که کاربران به عنوان ورودی به برنامه‌های کاربردی وب می‌دهند، توسط فرم‌های وب تهیه می‌شود. فرآیند پر کردن فرم‌ها یک فعالیت متناوب و تکراری می‌باشد که نوع داده‌های مورد نیاز برای آن را می‌توان شناسایی نمود. با مشاهده و بررسی فرم‌هایی از برنامه‌های کاربردی وب که در یک حوزه‌ی یکسان، داده‌های مشابهی را از کاربر دریافت می‌کنند، می‌توان نوع و ساختار داده‌های فرم را مشخص نمود [ARA2010C]. به عنوان مثال بسیاری از فرم‌هایی که برای ثبت‌نام و ورود به سایتها استفاده می‌شوند، اطلاعات عمومی و فردی همانند نام و آدرس پست الکترونیک کاربر را تقاضا می‌کنند.

در سال‌های اخیر از تکنیک‌های پر کردن خودکار[۱] و کامل کردن خودکار[۲] برای کمک به کاربر در وارد کردن داده‌ها در فرم‌های وب و پر کردن آن‌ها استفاده شده‌است. کامل کردن خودکار فرم یک ویژگی است که توسط بسیاری از کاربردهای وب فراهم شده‌است. در این تکنیک، بدون اینکه کاربر کلمه و یا عبارت مورد نظر خود را به طور کامل در فرم وارد نماید، سیستم آن را به کاربر پیشنهاد می‌دهد. در اکثر مرورگرها نیز این ویژگی وجود دارد. نحوه‌ی کار آن‌ها به این صورت است که مقادیری که کاربر قبلا درون فرم‌ها وارد کرده‌است را ذخیره و نگهداری می‌کنند. سپس براساس این تاریخچه، مقادیری را برای فیلدی که قبلا مشاهده شده‌است پیشنهاد می‌دهند.

در این پروژه، هدف پر کردن خودکار فرم می‌باشد. پر کردن خودکار فرم یک مکانیزم برای وارد کردن داده‌های مورد نظر کاربر در فرم‌های وب به صورت خودکار می‌باشد. در حال حاضر ابزارهایی در

---

[1] Auto-filling

[2] Auto-completion

مرورگرهای وب بدین منظور وجود دارد. روند کار این ابزارها معمولا بدین صورت است که کاربر داده‌های مورد نیاز برای پر کردن فرم‌ها را در ابتدا و پیش از استفاده از ابزار، توسط فرم از قبل آماده‌ای به ابزار وارد می‌کند. پس از شروع به کار، در صورت مشاهده‌ی عنصری در فرم که مشابه با داده‌های وارد شده توسط کاربر باشد، ابزار از داده‌های ذخیره شده برای پر کردن آن عنصر فرم استفاده می‌کند. روش استفاده شده برای شناسایی عناصر مشابه معمولا از روش‌های انطباق رشته می‌باشد. در این پروژه از روش‌ها و تکنیک‌های معنایی برای پر کردن خودکار فرم‌های وب استفاده می‌شود.

۱-۲- انگیزه

کاربران وب هرروزه با تعداد زیادی فرم مواجه هستند که باید برای جستجو و یافتن اطلاعات و استفاده از سرویس‌ها و امکانات وب، داده‌های خود را در آن‌ها وارد نمایند. فرآیند پر کردن فرم‌ها یک فرآیند تکراری و زمان‌بر است. کمک به کاربر برای پر کردن فرم‌ها به صورت خودکار باعث صرفه‌جویی در زمان و انرژی کاربر می‌گردد.

به عنوان مثال، فرض کنید کاربری به دنبال یافتن یک قطعه موسیقی خاص باشد. برای جستجوی این قطعه کاربر باید در سایت‌های مختلف موسیقی، اطلاعات قطعه را وارد نماید. قسمت قابل توجهی از این اطلاعات در بسیاری از سایتها مشترک می‌باشد. به عنوان مثال نام قطعه، نام نویسنده، نام آهنگساز، نام خواننده و گروه اجرا کننده‌ی موسیقی و ... باید در فیلدهای مختلف فرم در سایتهای موسیقی وارد شود. در این حالت کاربر باید در فرم هر سایت موسیقی، اطلاعات تکراری و یکسانی را به صورت دستی پر نماید. با وجود یک ایزار برای پر کردن خودکار فرم می‌توان در انجام این فرآیند به کاربر کمک نمود. بدین ترتیب که پس از بدست آوردن اطلاعات از منابع مختلف، داده‌ها را بجای کاربر در عناصر مختلف فرم قرار داد.

۱-۳- روش پیشنهادی[1]

در این پروژه از روش‌ها و تکنیک‌های معنایی برای پر کردن خودکار فرم‌های وب استفاده شده و چارچوبی برای انجام این کار ارائه گردیده است. روش استفاده شده برای قسمت‌های مختلف پروژه به شرح زیر می‌باشد:

روش نگاشت[2]: از روش مبتنی بر آنتولوژی استفاده شده‌است. پس از بررسی اطلاعات عناصر داده‌ای فرم‌های وب، آنتولوژی داده‌های موجود در فرم‌های وب ایجاد شده و در هنگام نگاشت، نام عناصر با مفاهیم تعریف شده در آنتولوژی مقایسه می‌گردد.

شناسه‌ی عناصر[3]: از میان اطلاعات استخراج شده برای هریک از عناصر وب، ویژگی نام به عنوان شناسه‌ی یک عنصر انتخاب گردید. ویژگی برچسب عناصر نیز در بسیاری از مواقع بیانگر مفهوم و معنی عنصر می‌باشد اما همیشه در دسترس نمی‌باشد. بنابراین در هنگام نگاشت، از ویژگی نام عناصر استفاده گردیده‌است.

بازخورد کاربر[4]: در این سیستم بازخورد کاربر دریافت شده و در تکمیل و تصحیح اطلاعات فرم استفاده می‌شود. در صورتیکه کاربر داده‌های جدیدی را در یک فرم وارد نماید، پس از آن، از این داده‌ها برای پر کردن فرم‌های بعدی استفاده می‌شود. همچنین اگر کاربر داده‌های یکی از فیلدهای اصلی در جستجوی داده‌های هر حوزه‌ی فرم که در مجموعه قوانین جستجو مشخص شده‌است را تغییر دهد، سیستم از این داده‌ها برای ادامه‌ی کار استفاده می‌کند.

---

[1] Proposed Method

[2] mapping

[3] Field Identifier

[4] User Feedback

منبع داده[۱]: در این سیستم دو منبع داده برای پر کردن فرم‌های جدید وجود دارد. اولین منبع، وب داده می‌باشد. یکی از مهمترین پروژه‌ها در وب داده، ابر داده‌های پیوندی باز[۲] می‌باشد. داده‌های زیادی در این ابر به صورت داده‌های پیوندی منتشر شده‌است. پس از بررسی مجموعه داده‌های موجود در این ابر، مجموعه داده‌ی DBpedia به عنوان مجموعه داده‌ی هدف برای جستجوی داده‌ها انتخاب گردید. منبع دوم، تاریخچه‌ی کاربر می‌باشد. داده‌های پر شده در فرم‌های قبلی به صورت معنایی در یک مخزن داده‌های معنایی ذخیره شده و سپس برای پر کردن فرم‌های جدید به‌کار می‌روند. در هنگام جستجوی داده، در ابتدا برروی مخزن محلی داده جستجو انجام شده و در صورت عدم وجود داده، از وب داده استفاده می‌گردد.

مجموعه قوانین جستجو[۳]: پس از بررسی داده‌های فرم‌ها و نیز مدل داده‌ای[۴] اطلاعات منتشر شده به صورت داده‌های پیوندی در مجموعه داده‌های هدف، فیلدهای اصلی در هر حوزه از فرم شناسایی شده و مجموعه قوانین جستجو برای یافتن دیگر داده‌های آن حوزه از فرم برروی ابر داده‌های پیوندی مشخص شده‌اند. همچنین مجموعه‌ای از مسندها برای نوشتن عبارات جستجو مورد ارزیابی قرار گرفته و در جستجو استفاده می‌شوند.

زمان اعتبار داده‌ها[۵]: برای هر یک از حوزه‌های فرم یک دوره‌ی زمانی اعتبار تعیین می‌شود. این دوره‌ی زمانی بسته به نوع داده‌های آن حوزه متفاوت است. در صورتیکه از زمان آخرین تغییر داده‌ها به اندازه‌ی بیش از دوره‌ی اعتبار آن‌ها گذشته باشد، آن داده‌ها قابل استفاده نیستند. انجام این کار

---

[1] Data Source

[2] LOD cloud

[3] Query Rule Set

[4] Data model

[5] Data Validation Time

باعث می‌شود تعداد اصلاحاتی که کاربر به دلیل پر شدن فرم توسط سیستم با داده‌های درستی که مدنظر کاربر نیستند انجام می‌دهد، کاهش یابد.

قابل به ذکر است که در روش پیشنهادی ارائه شده در این پروژه، بهبود روش از لحاظ زمانی و امنیت داده‌های کاربر مورد توجه نبوده‌است.

۱-۴- ابتکارات[1] پایان نامه

تا کنون کارهای تحقیقاتی فراوانی برای شناخت فرم‌های وب و داده‌های آن‌ها انجام شده‌است. یکی از اهداف انجام این تحقیقات پر کردن خودکار فرم‌های وب می‌باشد. در تحقیقات انجام شده در زمینه‌ی پر کردن خودکار فرم‌های وب هیچگاه به منبع داده‌های مورد نیاز توجه نشده‌است. امروزه با وجود وب داده‌ها و حجم زیاد داده‌هایی که به صورت پیوندی منتشر شده‌اند، ابر داده‌های پیوندی به عنوان یک منبع باز و در دسترس در برنامه‌های کاربردی تحت وب مورد مطالعه و بررسی قرار گرفته‌است. در این پروژه علاوه بر استفاده از داده‌های تاریخچه‌ی کاربر، استفاده از داده‌های منتشر شده برروی ابر داده‌های پیوندی به عنوان یک منبع خارجی در پر کردن فرم‌های وب مورد مطالعه قرار گرفته‌است. در اکثر تحقیقاتی که در استفاده از وب داده در برنامه‌های کاربردی انجام شده‌است، تنها یک حوزه از اطلاعات مورد توجه بوده‌است. به عنوان مثال تنها از داده‌های حوزه‌ی موسیقی و یا انتشارات استفاده شده‌است. دلیل این امر وجود داده‌های بیشتر در این حوزه‌ها می‌باشد. در این پروژه کارایی داده‌های موجود برروی وب داده در هشت حوزه‌ی مختلف بررسی شده‌است.

در این پروژه همچنین در فرآیند پر کردن فرم، سیستم ارتباط متقابلی را با کاربر حفظ می‌کند. بازخورد کاربر در تصحیح و تکمیل داده‌های فرم‌ها به‌کار گرفته می‌شود. داده‌های یافت شده به کاربر نمایش داده می‌شوند و بلافاصله پس از تغییر توسط کاربر، داده‌های جدید به صورت خودکار برای پر کردن بقیه‌ی عناصر همان فرم و نیز فرم‌های بعدی مورد استفاده قرار می‌گیرند. داده‌های ذخیره شده

---

[1] contributions

به عنوان تاریخچه‌ی کاربر به صورت معنایی و با استفاده از آنتولوژی داده‌های فرم در یک مخزن RDF محلی نگهداری می‌شوند.

همچنین برای داده‌های هریک از حوزه‌های فرم یک دوره ی زمانی اعتبار درنظر گرفته می‌شود که قابل تعیین است. در صورتیکه در تاریخچه‌ی کاربر، داده برای پر کردن فرم جدید موجود باشد اما زمان اعتبار آن گذشته باشد، احتمال اینکه همان داده‌های قبلی مورد نظر کاربر باشند بسیار کم است. بنابراین اگر فرم را با همان داده‌ها پر کنیم کاربر باید داده‌های وارد شده توسط سیستم را اصلاح نماید. دوره‌ی اعتبار پیشفرض تعیین شده برای هر حوزه‌ی فرم به صورت تجربی انتخاب شده‌است.

۱-۵- ساختار پایان‌نامه

در این پایان‌نامه با توجه به تکنیک‌های معنایی، یک روش برای پر کردن خودکار فرم‌های وب با استفاده از وب داده ارائه شده‌است. ساختار پایان‌نامه بدین شکل است که در فصل دوم مروری بر کارهای انجام شده در زمینه‌ی پر کردن فرم می‌پردازیم. چارچوب پیشنهادی برای پر کردن خودکار فرم و اجزاء آن در فصل سوم شرح داده می‌شود. معماری سیستم و فرآیند انجام کار در همین فصل ارائه می‌شوند. در فصل چهارم جزئیات پیاده‌سازی و نتایج حاصل بیان می‌گردند. فصل پنجم به نتیجه‌گیری و پیشنهادهایی برای کارهای آتی اختصاص یافته‌است.

فصل ۲- مرورادبیات

پر کردن خودکار فرم یک مکانیزم برای وارد کردن داده‌های مورد نظر کاربر در فرم‌های وب به صورت خودکار می‌باشد. در حال حاضر ابزارهایی در مرورگرهای وب بدین منظور وجود دارد و در هنگامیکه کاربر یک صفحه وب حاوی یک فرم را مشاهده می‌کند، با یک کلیک ماوس می‌تواند امکان پر کردن خودکار را فعال نماید. ابزار Google Toolbar Auto-fill [GTA2011] یکی از ابزارهای موجود است که ساده‌ترین شکل پر کردن خودکار را انجام می‌دهد و تنها برای فرم‌های ثبت نام که اطلاعات شخصی کاربر را نیاز دارند کار می‌کند. افزونه‌ی Firefox Auto-fill Forms [MAF2011] برای مرورگر موزیلا فایرفاکس نیز یکی دیگر از این ابزارها است. این ابزار نیز تنها به داده‌های شخصی کاربر محدود است اما به کاربر اجازه‌ی افزودن داده‌هایی اضافه بر داده‌های پیشفرض را می‌دهد. در هر دو این ابزارها کاربر باید یک فرم بخصوص و از قبل آماده که حاوی تعدادی از فیلدهای پایه‌ای همانند نام و آدرس می‌باشد را قبل از استفاده از ابزار پر نماید. در مرورگر سافاری نیز یک ویژگی برای استفاده‌ی مجدد از داده‌هایی که قبلا کاربر در فرم‌ها وارد نموده‌است وجود دارد. در اکثر این ابزارها از روش انطباق رشته برای تطبیق نام فیلد و نام عنصر موجود در فرم از قبل آماده استفاده می‌شود.

در این فصل کارهای تحقیقاتی انجام شده در زمینه‌ی پر کردن خودکار فرم‌های وب را مرور و بررسی می‌نماییم. از آنجاییکه یکی از اهداف این پروژه استفاده از وب معنایی در این زمینه می‌باشد، کارهای انجام شده در این زمینه با توجه به استفاده از وب معنایی مورد نظر بوده‌اند. در هنگام پر نمودن فرم‌های وب، یکی از وظایف اصلی، شناسایی عناصر فرم و داده‌هایی است که باید در هریک از آن عناصر قرار گیرد. در چند ابزار معرفی شده قبل، گفته شد که ابتدایی‌ترین روش برای انجام این کار، استفاده از روش‌های انطباق رشته برای نام فیلدها و عناصر یک فرم است. در این پروژه از آنتولوژی برای انجام این کار استفاده شده است. به همین دلیل کارهای انجام شده در این زمینه نیز بررسی

شده‌اند. همچنین از آنجاییکه نیاز داریم آنتولوژی داده‌های فرم را ایجاد و سپس استفاده نماییم، مرور مختصری بر روش‌های ایجاد آنتولوژی نیز انجام گرفته است.

مهمترین قسمت این پروژه استفاده از وب داده و داده‌های منتشر شده برروی ابر داده‌های پیوندی برای پر کردن خودکار فرم‌ها می‌باشد. در هیچیک از کارهای انجام شده‌ی قبلی از یک منبع داده‌ی خارجی برای انجام این کار استفاده نشده‌است؛ بلکه داده‌های قبلی کاربر و یا یک مجموعه داده محدود بکار رفته‌است. برای استفاده از وب داده مروری نیز بر تعدادی از کارهای انجام شده در زمینه‌ی استفاده از وب داده انجام داده‌ایم.

۲-۱- دریافت معانی عناصر فرم‌های وب

در حال حاضر تعداد بسیار زیادی برنامه‌ی تحت وب به صورت برخط برروی اینترنت در حال استفاده می‌باشند. اکثر این برنامه‌ها برای ذخیره و بازیابی داده‌های خود از پایگاه داده‌ها استفاده می‌نمایند. داده‌هایی که در این پایگاه داده‌ها ذخیره می‌شوند می توانند خصوصی و یا داده‌های باز و عمومی باشند اما دسترسی به این داده‌ها کنترل شده و تحت اختیار برنامه‌ی تحت وب می‌باشد. این داده‌ها همانگونه که گفته شد حجم بسیار زیادی دارند و با عنوان وب عمیق شناخته می‌شوند. در وب عمیق، این حجم زیاد از اطلاعات تنها می‌توانند از طریق رابط‌های پرسجوی[۱] یک پایگاه داده وابسته به برنامه‌ی تحت وب، قابل دسترس باشند و موتورهای جستجوی عمومی و متداول نمی‌توانند با این رابط‌ها تعامل داشته باشند. بنابراین رابط‌های پرسجو که در قالب فرم‌های جستجو در صفحات وب وجود دارند، تنها امکان دسترسی و اصلی‌ترین روش دستیابی به اطلاعات موجود در پایگاه داده‌های برنامه‌های تحت وب می‌باشند.

در پر کردن خودکار فرم‌های وب، اولین کار پس از تجزیه‌ی فرم و یافتن عناصر فرم در صفحه‌ی وب، درک معنی آن‌ها می‌باشد. این بخش یکی از قسمتهای اصلی در پر کردن فرم می‌باشد و در مقالات و

---

[1] Query Interface

طرح‌های تحقیقاتی مختلف به چالشهای موجود در مورد آن پرداخته شده‌است. این بحث که با عنوان ترجمه‌ی پرسجو[۱] و یا نگاشت محدودیت میان رابط‌های پرسجوی وب[۲] نیز شناخته می‌شود، مورد توجه بسیاری از محققان در زمینه‌ی وب عمیق قرار گرفته است. در این قسمت به بررسی تعدادی از کارهای انجام شده در این زمینه خواهیم پرداخت.

رابط‌های پرسجو در وب عمیق، برای نمایش داده شدن به کاربر در صفحات HTML تعبیه شده‌اند که از این صفحات برای دریافت درخواست‌های پرسجوی کاربر استفاده می‌شود. در واقع رابط پرسجوی یک پایگاه داده‌ی وابسته به برنامه‌ی کاربردی تحت وب، از گروهی از فیلدهای وابسته به دامنه[۳] تشکیل شده‌است. کاربر می‌تواند تمامی نیازمندیهای جستجوی خود را از طریق این فیلدها در یک رابط پرسجوی مجتمع شده قرار دهد و پس از ارسال درخواست، تمامی منابع اطلاعاتی وابسته به آن رابط، مورد جستجو قرار خواهند گرفت.

روشهای مختلفی برای انطباق عناصر یک رابط پرسجو با یک نوع داده یا مفهوم وجود دارد که ساده‌ترین آن‌ها انطباق رشته‌ای نام عنصر با نام یک مفهوم می‌باشد. در [WAN2009] از مفهوم آنتولوژی برای درک معنی یک رابط جستجو استفاده شده و چارچوبی برای پرکردن خودکار فرم با استفاده از این روش پیشنهاد شده‌است. فرآیند کلی پرکردن فرم در این روش به چهار مدل تقسیم شده‌است: ساختن آنتولوژی[۴]، استخراج شِما[۵]، نگاشت آنتولوژی[۶] و ترجمه‌ی پرسجو. این روش را می‌توان به صورت مجموعه‌ای از قوانین نگاشت محدودیت که توسط این چهار مدل به طور خودکار اعمال می‌شوند دانست که می‌تواند پرسجوها را از رابط مجتمع شده به رابط‌های پایگاه‌داده‌ی متفاوت

---

[1] Query translation

[2] Constraint mapping across web query interfaces

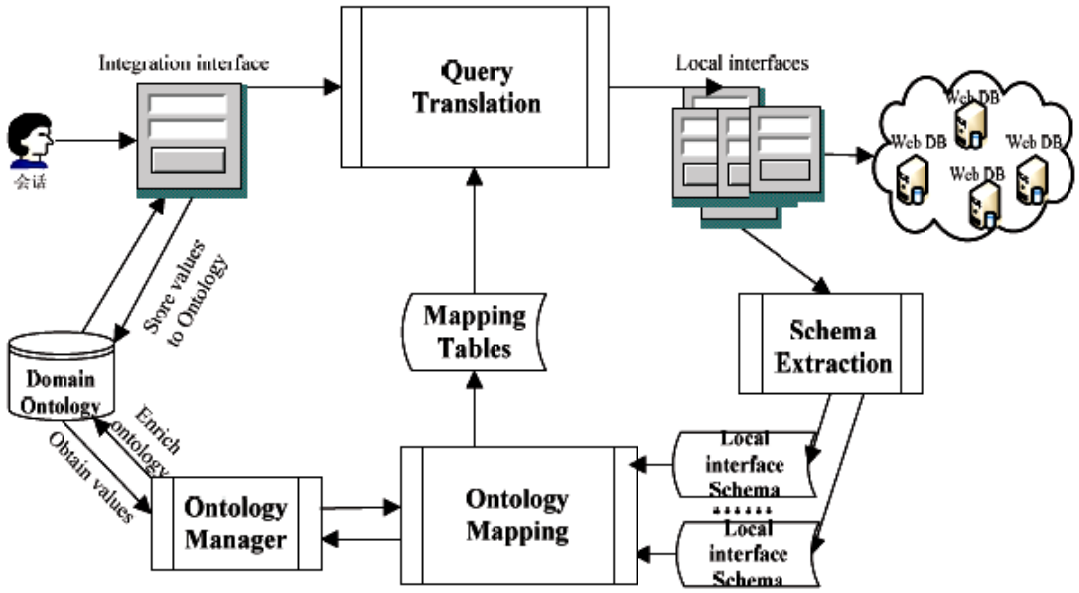
[3] Domain-related attributes

[4] Ontology construction

[5] Schema extraction

[6] Ontology mapping

وب ترجمه کند. در ادامه به بررسی این چهار مدل می‌پردازیم. چارچوب پیشنهادی در شکل ۲-۱ قابل مشاهده می‌باشد.



شکل ۲-۱ چارچوب پرکردن خودکار فرم‌های وب [WAN2009]

فرآیند انجام کار در چارچوب پیشنهادی در شش گام بیان شده‌است. در اولین گام، مدیریت آنتولوژی قرار دارد که مسئول انجام وظایف مرتبط با آنتولوژی است. همچنین قیدهای یک پرسجو که کاربر در رابط پرسجوی مجتمع شده قرار میدهد در قسمت مدیریت آنتولوژی نگهداری می‌شود. در گام دوم، هر نمونه‌ی پرسجو در رابط پرسجوی مجتمع، با قیود متناظر و مقادیر نمونه ترکیب می‌شود. این قیود و مقادیر هر دو از فایل آنتولوژی که به زبان OWL نوشته شده‌است استخراج می‌شوند.گام سوم شامل استخراج شِما می‌باشد که برای دریافت و تحلیل فیلدها و کنترل‌های پرسجو استفاده می‌شود. ورودی بخش استخراج شِما، رابط‌های محلی می‌باشند و مجموعه شِمای رابط‌های جستجو به عنوان خروجی این بخش می‌باشد. در گام چهارم، از مدل نگاشت آنتولوژی برای ثبت روابط انطباق میان رابط‌های جستجوی محلی و رابط مجتمع استفاده می‌شود. خروجی مدل نگاشت آنتولوژی، جداول نگاشت محلی-آنتولوژی-مجتمع می‌باشند. در مورد داده‌های موجود در این جداول در ادامه توضیح داده خواهد شد. در پنجمین گام، مترجم پرسجو با استفاده از جداول نگاشت، فرم‌های جستجوی محلی را