نشریه علمی- پژوهشی علوم و فنون نقشه برداری، دوره سوم، شماره ۴، اردیبهشت ماه ۱۳۹۳

وب معنایی- مکانی، چارچوبی برای ذخیره و بازیابی اطلاعات affordance نقاط مورد توجه

رضا عرب شیبانی*[۱]، فرشاد حکیم‌پور[۲]

[۱] دانشجوی کارشناسی ارشد سیستم‌های اطلاعات مکانی - گروه مهندسی نقشه‌برداری- پردیس دانشکده‌های فنی- دانشگاه تهران
rasheibani@ut.ac.ir

[۲] استادیار گروه مهندسی نقشه‌برداری- پردیس دانشکده‌های فنی- دانشگاه تهران
fhakimpour@ut.ac.ir

(تاریخ دریافت اسفند ۱۳۹۲، تاریخ تصویب اردیبهشت ۱۳۹۳)

چکیده

به میزانی که حجم داده‌های تحت وبِ دارای محتوای مکانی افزایش می‌یابد نیاز به رویکردهایی که دارای توانایی مدیریت این حجم داده‌ها در شبکه وب باشند افزایش می‌یابد. گفتمان وب معنایی با تعریفی جدید و متفاوت از پایگاه‌های داده به منظور ذخیره و بازیابی داده‌ها، راه حلی برای مواجهه با چالش‌های ناهمگنی و گستردگی توزیع منابع اطلاعاتی است. از طرفی می‌توان گفت که افراد در فعالیت‌های روزمره خود با مفهوم و معنایی که برای هر مکان خاص قائلند سر و کار دارند. معیارهای گوناگون شناختی برای تعریف یک مکان باعث می‌شود که علاوه بر طبقه‌بندی، اطلاعات توصیفی بیشتری برای مدلسازی ویژگی‌های متمایزکننده(affordance) نقاط مورد نیاز باشد، درحالیکه که تعریف این اطلاعات توصیفی در یک اسکیمای خاص از مدل‌های رابطه‌ای امکان‌پذیر نیست. مدلسازی داده در پایگاه‌های داده گرافی و در قالب داده‌های مکانی- معنایی، تعریف و ذخیره داده‌های مرتبط با مدلسازی مکان بر اساس affordance را امکان‌پذیر می‌کند. در این تحقیق با مروری بر مفاهیم و فناوری‌های مرتبط با کارکردهای وب معنایی، نشان داده خواهد شد که چگونه می‌توان داده‌های مکانی را در قالب‌های معنایی تعریف و تفسیر کرد. همچنین سامانه اطلاعاتی طراحی و پیاده‌سازی شده در این تحقیق، از قالب داده‌های وب معنایی به‌منظور ذخیره و بازیابی داده‌های مکانی- معنایی affordance نقاط مورد توجه، استفاده خواهد کرد. جهت ارزیابی نتایج به‌دست آمده، اطلاعات بازگردانده شده توسط سامانه مکانی- معنایی با اطلاعات همسان در مدل‌های رابطه‌ای مقایسه خواهد شد و نتیجه مقایسه‌ها مورد بررسی و ارزیابی قرار خواهد گرفت.

واژگان کلیدی : وب معنایی، نقاط مورد توجه، بازیابی و ذخیره، GeoSPARQL، مدلسازی، affordance.

* نویسنده رابط

۱- مقدمه

وضعیت کنونی و آینده وب به نحوی است که دربرگیرنده حجم بسیار زیادی از داده‌ها اعم از مکانی و غیرمکانی خواهد بود، از این رو به راهکارهایی بهینه برای مدیریت و استفاده از چنین داده‌هایی نیاز خواهیم داشت. ماهیت داده‌های تحت وب به گونه‌ای است که این داده‌ها ناهمگن هستند و بر روی منبع‌های مختلفی قرار گرفته-اند، که این شرایط کاربر را با سوالاتی از قبیل زیر روبرو می‌کند : «چگونه می‌توان اطلاعات مورد نیاز را یافت و از آن استفاده کرد؟، چگونه می‌توان به منابع دیگری که در زمینه مدنظر دارای اهمیت و ارتباط باشند، متصل شد؟ و یا چگونه می‌توان نتایجی را که دیگر منابع بدست آورده-اند، با نتایج کار خود تلفیق کرد؟»[۴]. دسترسی به منابع مختلف با وضعیت‌های مختلف، شرایط خاص هر منبع داده را اقتضاء می‌کند، به طوری که ممکن است در یک بررسی اولیه، تنوع این شرایط برای منابع مختلف بسیار زیاد و پرهزینه باشد. به صورت خلاصه می‌توان گفت که کاربرِ وب می‌تواند حجم زیاد داده‌ها را در اختیار داشته باشد، امّا این داده‌ها از منظر دسترسی، دارای توزیعِ پراکنده بر روی منبع‌های مختلف هستند و همچنین ساختارهای داده‌های در دسترس، در مقایسه با یکدیگر، ناهمگن است.

از طرف دیگر تکامل روزافزون صفحات وب در سال‌های اخیر به نحوی است که طراحان این صفحات تلاش می‌کنند علاوه بر استفاده از قالب‌های مرسوم مانند متن‌ها و عکس‌ها، امکانات پیشرفته‌تری به طراحی اضافه کنند تا صفحات وب دارای قابلیتِ هرچه بیشتر ارتباط و استفاده از داده‌های متنوعِ تحت وب باشند. رویکرد وب معنایی که برای اولین بار توسط Tim Berners-Lee در سال ۲۰۰۱ مطرح شد را می‌توان به‌صورت زیر خلاصه کرد : «وقتی رایانه‌ها علاوه بر بازیابی، دارای قابلیت درک داده‌های موجود تحت وب باشند، ما گونه‌های جدید از وب و انواع مختلفی از برنامه‌های هوشمند تحت وب را خواهیم داشت»[۴]. به عبارت دیگر می‌توان گفت که ابزار وب ایجاد شده بر مبنای وب معنایی، افزون‌بر قابلیت صفحات وب معمولی، قادر به انتقال اطلاعات معنایی به شکل مرسوم مانند XML و درک این نوع اطلاعات معنایی است. مدلسازی بر اساس affordance روشی است که توانایی مدیریت داده‌هایی را داراست که مدلسازی آنها در سامانه‌های رابطه‌ای دشوار است. هنگامی که اطلاعات مربوط به مکان‌های مختلف با درنظرگرفتن معیارهای غیریکسان صورت بپذیرد، طراحی اسکیمای پایگاه داده رابطه‌ای غیر ممکن و در صورت امکان دارای ناسازگاری و هزینه‌ی زیاد است. توصیفات مکانی داده‌های داوطلبانه، توسط افراد مختلف صورت می‌پذیرد و طبیعتاً مطابق‌با معیارهای مختلف در ذهن افراد است، از این رو انتظار می‌رود مدلسازی بر مبنای affordance دراین موارد دارای کارائی موجه باشد.

فعالیت‌های تحقیقاتی متنوعی که برای ذخیره[۱] و پرس و جو[۲] از اطلاعات معنایی- مکانی موجود در شبکه وب صورت گرفته است. S. Scheider در سال ۲۰۰۸ با استفاده از معیارهایی تعریف شده و داده‌های معنایی خیابان‌ها مثل ویژگی‌های جاده‌ها، محدودیت‌های دور زدن و نحوه اتصال تقاطع خیابان‌ها، تقاطع‌ها را به‌عنوان نقاط مطلوب در پایگاه داده ای مدلسازی کرد [۲]. هم چنین در سال ۲۰۱۰، O. Fonts با استفاده از استانداردهای ساده‌ی موجود مثل فیلترهای WFS، جستجو و بازیابی اطلاعات معنایی مکانی تحت وب را مدلسازی نمود [۹]. در ادامه روند تحقیقات و مدلسازی در این زمینه Yu-Wei Lin در سال ۲۰۱۱ سعی کرد مدلی را برای استخراج و ذخیره اطلاعات معنایی داده‌های مکانی داوطلبانه[۳] طراحی کند تا اصطلاحاً تبادل دانش مکانی امکانپذیر شود[۱]. در همین سال۲۰۱۱ Mao Ye نیز چارچوبی را ارائه داد به‌گونه‌ای که در سرویس‌های مکانی ابتدا ویژگی‌های مکانی نقاط موردتوجه با استفاده از داده‌های معنایی تعریف شوند و سپس عملیات پرس‌وجو و تعیین رسته‌ها وکلاس‌ها فراهم آید[۸]. در سال ۲۰۱۲ Alkyoni Baglatzi با استفاده از یک مدلسازی نسبتاً پیچیده پیشنهاد داد که از اطلاعات معنایی بعنوان یک لایه‌ی اطلاعاتی میانی در نقشه‌های تحت وب استفاده شودکه دراین صورت پرس‌وجوی معنایی- مکانی علاوه بر پرس‌وجوهای دیگر امکانپذیر می‌گردد [۱۱]. M. Ruta در سال ۲۰۱۲ با استفاده از مدلسازی ذخیره و پرس‌وجوی داده‌های معنایی مربوط به

۱ storage
۲ retrieval
۳ Voluntary Geographic Information (VGI)

نقاط مورد توجه سامانه‌ای را طراحی کرده است که در آن امکانِ یافتنِ مسیرهای بهینه‌ای که هم در معیارهای معنایی صدق می‌کنند و هم معیارهای مکانی را برآورده می‌کنند را فراهم می‌کند. [۱۰]. هم‌چنین در راستای استانداردسازی قالب گردآوری اطلاعات معنایی- مکانی R. Battle در سال ۲۰۱۲، طی تحقیقی جهت معرفی GeoSPARQL ، از این ایده دفاع کرد که این سندِ استاندارد قابلیت تبدیل شدن به استاندارد مشترکِ بین مراکز تحقیقاتی و تجاری را داراست [۵].

در پژوهش پیش رو کارکرد وب معنایی در ذخیره و بازیابی استطاعت[۱] (affordance) موجود در یک رسته خاص از نقاط مورد توجه (Point Of Interests) مربوط به کشور انگلیس پیاده سازی شده و نتایج آن در مقایسه با نتایج فرآیندهای مدلسازی و پرس‌وجوی مشابه در پایگاه‌های داده رابطه‌ای تحلیل شده است. در این راستا استانداردهای تعریف شده وب معنایی برای ذخیره و بازیابی داده‌های مکانی بررسی شده و ابزارهای موجود فعلی در دامنه فناوری اطلاعات که از استانداردهای مکانی- معنایی تبعیت می‌کنند، مرور گردیده‌اند. در بخش پیش رو فعالیت‌ها و نتایج تحقیقاتی گوناگونی که در این زمینه برای پیشبردِ طراحی سیستم‌های مکانی- معنایی صورت گرفته است بررسی خواهد شد.

۲- کاربرد وب معنایی در مدیریت اطلاعات

فناوری وب معنایی به دنبال ایجاد بستری جهت استفاده از داده‌های معنایی و هم‌چنین کاهش ناهمگنی میان داده‌ها است. این فناوری ترکیبی است از یک روش نام‌گذاری با خاصیت گستردگی زیاد به نام URI[۲]، همراه با یک قالب رسمی به منظور بیان اطلاعات مانند RDF[۳] و OWL[۴] و همچنین یک سازوکار مشترک جهت پرس‌وجو به نام SPARQL[۵]. ماهیت چندوجهی URIها وضعیت دسترسی به داده‌هایی که روی کارسازهایی[۶] که از نظر فیزیکی یا الکترونیکی از همدیگر جدا هستند، را بهبود می‌دهد. استفاده از اصطلاحات مشترک و استاندارد جهت بیان اطلاعات، تلاش کاربران برای درک و استفاده از داده‌ها را کاهش می‌دهد. SPARQL نیز یک زبان پرس-وجوی مشترک برای دسترسی به اطلاعاتی که در قالب اصطلاحات اشاره شده ذخیره شده اند را فراهم می‌کند.

یک مثال عینی از کاربرد وب معنایی در شبکه‌ی وبی که امروزه مورد استفاده قرار می‌گیرد نشان‌هایی[۷] است که در شبکه‌های اجتماعی مختلف مانند Flicker مورد استفاده قرار می‌گیرد. به‌گونه‌ای که کاربران اصطلاحات مشترکی را برای توصیف اطلاعات با محتوای یکسان استفاده می‌کنند. این قابلیت موجب تسهیل جستجو و درک محتوای اطلاعات شده است (مانند پرس‌وجوی تمام عکس‌های یک سایت که در یک ساحل مشخص اقیانوس گرفته شده اند). اما نشان‌های درون یک سایت یا منبعِ مشخص، با سایت‌ها یا منابع دیگر سازگاری ندارند، به طوری که اگر حتی یک نشان مشابه (مانند ساحل) را جستجو کنیم، سیستم‌های موجود دارای قابلیت جستجو در سرتاسر وب نیستند.

تلفیق وب معنایی با داده‌های مکانمند با دو چالش اساسی روبرو است. نخستین چالش تعریف مجموعه واژگان و استانداردهای مناسبی است که داده‌های مکانی را به گونه‌ای مدل کنند که علاوه بر تبعیت از قالب استاندارد RDF و زبان پرس‌وجوی SPARQL، با قالب-های استاندارد معروف مکانی تعریف شده توسط OGC[۸] مثل GML[۹] سازگاری داشته باشند. چالش دوم مربوط به استفاده از داده‌های مکانی در ساختارهای وب معنایی، تصمیم‌گیری برای استفاده از یک فن‌آوری واحد است. در این پژوهش راهکارهای موجود برای چالش‌های اشاره شده پوشش داده می‌شود.

۲-۱- استانداردهای ذخیره و بازیابی داده‌های مکانی- معنایی

به‌عنوان نخستین الگو استاندارد GeoPos84 توسط مؤسسه W3C ارائه شد. این استاندارد به منظور بیان نقاط

---

۱ Affordance در اینجا به معنی «استطاعت فراهم یا ارائه کردن امکانات» بوده و در بخش ۳ تشریح می‌گردد.

۲ Uniform Resource Identifier

۳ Resource Description Framework

۴ Ontology Web Language

۵ یک سرنام بازگشتی از واژه‌های SPARQL Protocol And RDF Query Language

۶ Server

۷ Tag

۸ Open Geospatial Consortium

۹ Geographic Markup Language

مورد توجه در سیستم مختصات جهانی WGS84 با استفاده از طول و عرض جغرافیایی است. این استاندارد تنها بیانِ موقعیت یک نقطه را مدنظر دارد و از بقیه استانداردهای OGC پشتیبانی نمی‌کند. در ادامه GeoRSS معرفی شد که استاندارد بیان اطلاعات مکانی درقالب خورد وب[1] است. خوردِ وب سامانه‌ای است که برای کاربر امکان پیگیری بخشهایی از وب که به طور متوالی به روز می‌شوند را فراهم می‌کند. GeoRSS از محتوای مکانی شامل نقطه، مجموعه نقطه‌ها، خط‌ها، چندضلعی‌ها و توصیفات آن‌ها پشتیبانی می‌کند. در سمتِ کاربر برنامه‌هایی وجود دارند که GeoRSS را دریافت کرده و نمایش و یا تحلیل این داده‌ها را میسر می‌کنند. اما GeoSPARQL قالب استاندارد اصلی مورد استفاده در ذخیره وبازیابی داده‌های معنایی- مکانی است. «با در نظر گرفتن استاندارد SPARQL و مشخصات تعریف شده در آن، زبان GeoSPARQL توسط مؤسسه OGC به گونه‌ای طراحی شده است که از نحوه ثبت استاندارد داده‌های مکانی تحت وب در زمینه وب معنایی پشتیبانی کند» [۷]. سند GeoSPARQL دربرگیرنده مجموعه‌ای تعریف شده از واژگان به‌منظور بیان مختصات اشیای مکانی در قالب RDF است. به‌علاوه این استاندارد یک بخش به‌SPARQL می‌افزاید تا پرس‌وجو از داده‌های ذخیره شده معنایی- مکانی امکان‌پذیر شود. استاندارد GeoSPARQL شامل کلاس‌هایی است که در قالب معنایی، اشیای مکانی را تعریف می‌کنند. بخش‌های دیگر این استاندارد شامل مدلسازی توپولوژی بین اشیای هندسی، نحوه نگارش پرس‌وجوهای مکانی و استفاده از تابع‌های مکانی مختلف در پرس‌وجوها است. قسمت RDFS[2]، نحوه بررسی مطابقت RDFها با RDFS تعریف-شده را مشخص می‌کند. جهت کاستن از پیچیدگی نگارش پرس‌وجوهای GeoSPARQL نیز، قسمت بازنویسی پرس‌وجوها پیش‌بینی شده است.

۲-۲- فناوری‌های موجود

برای انتخاب یک فناوری از میان فناوری‌های موجود گزینه‌های زیادی در دسترس نیست چرا که بر خلاف پیشرفت‌های گسترده‌ای که در زمینه فناوری‌های مرتبط با ذخیره RDF بوجود آمده است، تنها تعداد محدودی فناوری موجود است که از داده‌های معنایی- مکانی پشتیبانی می‌کنند. علاوه بر آن از میان این تعداد محدود فناوری که توانایی مدیریت داده‌های معنایی- مکانی را دارا هستند، تعداد محدودی از آنها از استاندارد GeoSPARQL پشتیبانی می‌کنند. برخی از فناوری‌های موجود که ادعا می‌شود از این استاندارد پشتیبانی می‌کنند تنها برخی از بخش‌های GeoSPARQL را پشتیبانی می‌کنند و درصدی از بخش‌های این سند استاندارد را پوشش نمی‌دهند. در جدول ۱ وضعیت فناوری‌های موجود را بر اساس میزان رایج بودن استفاده در بین کاربران مشاهده می‌کنیم. [۶]

۳-مدلسازی نقاط بر اساس Affordance

مفهوم استطاعت در سال ۱۹۸۶ توسط آقای Gibson [۱۲] معرفی شد. ایشان «فرآیند درک محیط اطراف را مبتنی بر استخراج ویژگیهای ثابت محیطی دانست که باعث تحریک قوه‌های شناختی بشری می‌شوند» [۳] که این ویژگی‌ها را استطاعت نامید. می‌توان گفت که استطاعت یعنی پتانسیلِ قابل درکِ هر چیز برای انجام فعالیت‌هایی که آن فعالیت‌ها بعنوان معیارهایی برای تعریف رسته[3]های مختلف محسوب می‌گردند. برای مثال قلم شی‌ای است که پتانسیل نوشتن دارد پس توانایی نوشتن یکی از استطاعت‌های قلم محسوب می‌گردد.

تعریف مدل‌های بر اساس استطاعت[4] یکی از انواع مدلسازی است که برای بیان ویژگی‌های مکانی نقاط استفاده می‌گردد. از آنجایی که استطاعت، امکاناتی است که هر محل در اختیار افراد قرار می‌دهد، «مسأله‌ای که بر اساس استطاعت حل می‌شود هم به نیروهای شناختی افراد مرتبط است و هم جنبه‌های مهندسی را در بر می‌گیرد» [۳]. بخاطر همین مسأله‌ی روبرو بودن با افراد و اشیا می‌توان از این مدل انتظارِ بیانِ اطلاعات مکانی داشت.

---

۱ Web-Feed
۲ RDF Schema
۳ Category
٤ Affordance-based Models

جدول ۱- مقایسه فناوری‌های موجود برای ذخیره و بازیابی داده‌های وب معنایی

نام فناوری	هدف طراحان	نرم افزارهای پایه	شاخص‌گذاری مکانی	وضعیت پشتیبانی از بخش‌های GeoSPARQL	وضعیت پشتیبانی از توابع مکانی
Parliament	فناوری ذخیره سه‌گانه‌ها	Berkley DBMS	R-Tree	تمام بخش‌ها	تمام توابع مکانی تعریف شده GeoSPARQL
USeekM	فناوری ذخیره سه‌گانه‌ها	Sesame[1]	R-Tree	پشتیبانی محدود	توابع مکانی موجود در پایگاه داده PostGIS
AllegroGraph	فناوری ذخیره گراف پنج‌گانه‌ها	Sesame	بدون شاخص	عدم پشتیبانی	چند تابع محدود هندسی
Oracle Spatial and Graph	مدیریت داده‌های پیوندی مکانی	Oracle 11g DBMS	B-Tree	بخش‌هایی شبیه به GeoSPARQL	غالب توابع هندسی
Staborn	ذخیره داده‌های پیوندی زمانی مکانی	Sesame	B^+-Tree	عدم پشتیبانی	توابع هندسی موجود در PostGIS

اگر مدل یک مکان را بتوان به گونه‌ای مشخص صورت‌بندی کرد، در محیط‌های محاسباتی مانند یک سیستم اطلاعات مکانی، باعث تسهیل بررسی‌ها و تحلیل مشخصات و موقعیت‌ها و رفتارهای محیطی می‌شود. در یک مدل بسیار مفصل، که از مفاهیم استطاعت برای بیان یک مکان استفاده شده است، ویژگی‌های فیزیکی، فعالیت‌های انسانی، عامل‌های اجتماعی و اقتصادی و فرهنگی ،ذهنیت‌های افراد مختلف در مورد مکان مورد نظر، اسامی و نمادهای بیانگر محل و نوع افراد استفاده کننده از محل مورد نظر به عنوان معیارهای اصلی این نوع مدلسازی تعریف شده است.[۱۱] مثلاً برای توصیف مفهوم رستوران به عنوان یک مکان، معیارهای مختلفی مثل ارائه غذای صبحانه، ارائه غذاهای خانگی، نحوه ارتباط با خدمتکاران، وجود روزنامه، نحوه روابط اجتماعی در محل، میزان فضای اختصاصی به افراد در محل و بسیاری دیگر از معیارها به عنوان استطاعت‌های رستوران در نظر گرفته می‌شود.

ممکن است که علاوه بر وجود استطاعت‌های ثابت برای یک رسته، هرنمونه از رسته یک مجموعه استطاعت مختصِّ خود را دارا باشد. برای مثال در نظر بگیرید که از میان مجموعه رستوران‌های در حال بررسی، تنها یک رستوران دارای محل استراحت باشد و یا تنها چند رستورانِ محدود نوشیدنی‌های رژیمی مختلف را ارائه کنند و بسیاری از ویژگی‌های دیگر که ممکن است تنها مختص یک رستوران باشد. در این صورت اگر بخواهیم در پایگاه‌های داده رابطه‌ای این ویژگی‌ها را مدل کنیم، تعداد زیادی ستون به جدول اضافه می‌شود که اکثر محتوای این ستون‌ها دارای مقادیر صفر یا null است و تنها یک یا چند ذخیره محدود از ستون دارای مقداری بامعنا هستند. در این صورت ستون‌های اضافه شده به مدل رابطه‌ای ممکن است باعث ناسازگاری در مدل شوند. علاوه بر آن این کار باعث پیچیده شدن و پرهزینه شدن فرآیند مدلسازی خواهد شد. از طرف دیگر در روش وب معنایی برای ذخیره و بازیابی، می‌توان هر استطاعت را به عنوان یک سه‌گانه[2] ذخیره نمود که عملاً پیچیدگی اضافی به مجموعه داده مدلسازی شده وارد نمی‌کند و امکان جستجو و پرس‌وجو داده‌های ذخیره شده نیز فراهم می‌آید. از این جهت برتری روش وب معنایی در مقایسه با روش پایگاه‌های داده رابطه‌ای انتظار می‌رود.

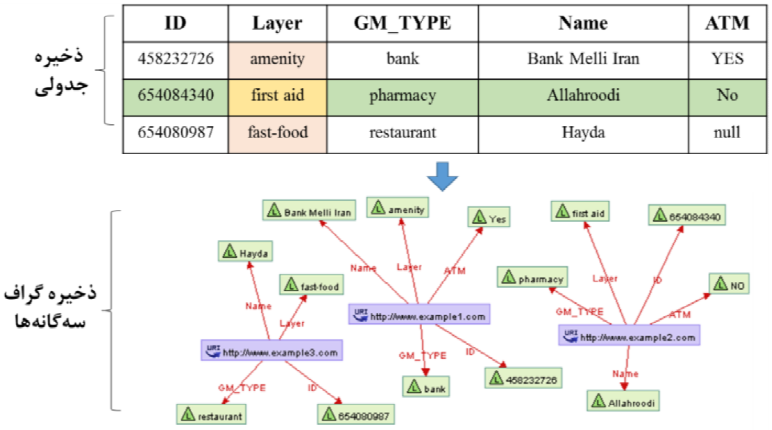
۴-مدل مفهومی

داده‌هایی که در پایگاه‌های داده معنایی ذخیره می‌شوند ابتدا باید به قالب داده سه‌گانه تبدیل شوند تا امکان تزریق آنها به پایگاه‌های داده گرافی (که برای

[1] پروژه Sesame یک چارچوب برای پرس‌وجو و ذخیره و تحلیل داده‌های معنایی (RDF) است که از زبان SPARQL استفاده می‌کند.

[2] Triple

ذخیره داده‌های معنایی- مکانی استفاده می‌شوند) فراهم شود. بخشی از داده‌های استفاده شده در این تحقیق از منابع دسترسی آزاد اخذ شده اند که در قالب جدولی در پایگاه‌های داده رابطه‌ای ذخیره شده بودند. برای تطبیق مدل این داده‌ها با مدلسازی مدنظر این تحقیق، این داده‌ها از قالب رابطه‌ای به قالب گرافی و به شکل سه گانه‌ها تبدیل شده‌اند. همانطور که در شکل ۱ مشاهده می‌شود به ازای وجود هر داده در پایگاه‌داده رابطه‌ای یک یال به گراف داده اضافه خواهد شد. در صورت عدم وجود یک داده، در پایگاه‌داده رابطه‌ای حجمی از حافظه به آن اختصاص داده خواهد شد که محتوای آن null است، اما در گراف داده‌های به صورت سه‌گانه، یال متناظر بدون اشغال حجمی از حافظه از مجموعه گراف داده حذف خواهد شد.

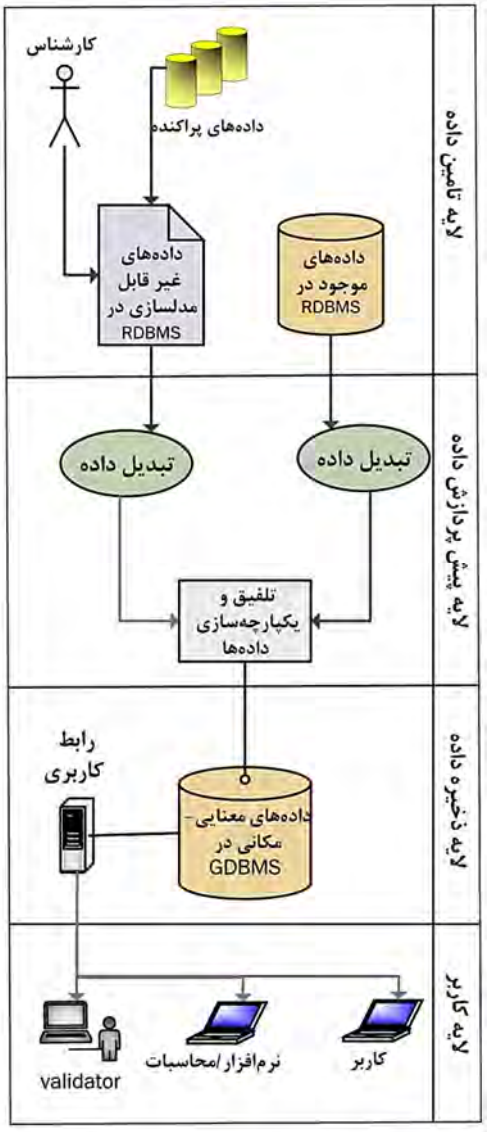


شکل ۱- مقایسه نوع ذخیره داده در مدل رابطه‌ای (ذخیره جدولی) و مدل معنایی (ذخیره سه‌گانه)

شکل ۲ مراحل پایه‌ای جهت طراحی و اجرای سامانه اطلاعاتی معنایی- مکانی مدنظر پژوهش ما را بیان می‌کند. ابتدا می‌توان گفت که به طورعام دوگونه اطلاعات دردست خواهد بود، دسته اول اطلاعاتی که مدلسازی آنها در پایگاه‌های داده رابطه‌ای امکانپذیر است و دسته دوم اطلاعات مرتبطی که یا به صورت پراکنده برروی منابع گوناگون موجودند و یا ذهنیت افراد کارشناس مرتبط دربرگیرنده آن داده‌ها است. باز هم به صورت کلی می‌توان گفت که داده‌های دسته دوم(افراد و منابع غیر رابطه‌ای)، قابل مدلسازی به صورت رابطه‌ای نیست اما امکان مدلسازی این نوع داده‌ها به صورت معنایی مکانی امکان‌پذیر است.

به منظور یکپارچه‌سازی (تبدیل به قالب داده معنایی- مکانی و تلفیق دو دسته داده) ابتدا باید هر دو نوع داده به قالب سه‌گانه تبدیل شوند. نوع تبدیل قالب از رابطه ای به گرافی شامل اختصاص URIهای استاندارد و بامعنا، استفاده از توابع مکانی استاندارد و تشخیص. حذف بخشهای بدون داده است. در این تحقیق ما از داده‌های OSM که به صورت رابطه‌ای ذخیره‌شده اند استفاده کردیم و تمرکز بیشتر بر روی استخراج و تبدیل مختصات نقاط موردتوجه از این منبع بود. با استفاده از توابع استاندارد تعریف شده در GeoSPARQL مختصات به قالب معنایی- مکانی تبدیل شدند و در صورت وجود، اطلاعات توصیفی OSM نیز بصورت سه‌گانه تبدیل شدند و به نقطه موردتوجه مرتبط پیوست داده شدند.

در مرحله بعدی روند مدلسازی اطلاعات affordance جمع آوری شده از منابع مرتبط به قالب معنایی- مکانی تبدیل شده و پس از تلفیق در یک پایگاه داده گرافی که از استانداردهای معنایی- مکانی مدنظر (GeoSPARQL) پشتیبانی می‌کند به صورت یکپارچه ذخیره خواهند شد که آنرا یک گراف داده می‌نامند. به منظور پرس‌وجو توسط زبان استاندارد GeoSPARQL به یک رابط میان کاربران و پایگاه داده برای تبدیلات احتمالی نیاز خواهد بود.

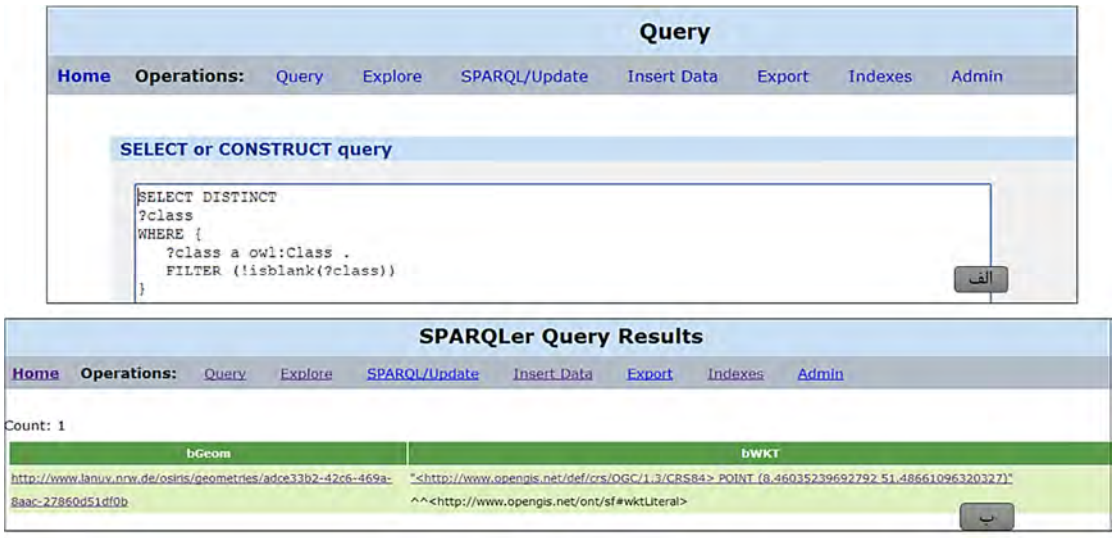


شکل ۲- روند اجرایی سامانه اطلاعاتی معنایی- مکانی

ممکن است داده‌های بازیابی شده، مستقیماً توسط کاربر مورد نیاز و استفاده باشد همچنین ممکن است یک نرم‌افزار کاربر سامانه باشد و اطلاعات بازیابی شده را به عنوان ورودی یا بخشی از عناصر موردنیاز برای اجرای برنامه‌های بالادستی استفاده کند. یک ناظر موقت جهت نظارت و ارزیابی اطلاعات بازیابی شده از لحاظ رعایت قالب استاندارد و محتوای منطبق بر پرس‌وجو باید به API سامانه دسترسی داشته باشد و در صورت لزوم ویرایش سیستم با سطح دسترسی بیشتر، به پایگاه داده دسترسی داشته باشد و قابلیت ویرایش داده‌ها برای او فراهم باشد.

۵- پیاده‌سازی و ارزیابی نتایج

با مدنظر قرار دادن این نکته که داده‌های مورد نظر این تحقیق، اطلاعات مربوط به نقاط مورد توجه هستند، از یک مجموعه داده شامل نقاط با نوع‌های مختلف استفاده شده است. این مجموعه داده شامل رسته‌های متنوعی از نقاط مثل مکان‌های تاریخی، نظامی، فروشگاه‌ها، فرودگاه‌ها و نقاط گردشگری است که این داده‌ها از مجموعه نقشه‌های با دسترسی آزاد OSM کشور انگلیس(B-box=[-6.236 50.303 1.670 55.986]) استخراج شده است و به قالب‌های قابل استفاده در GeoSPARQL تبدیل شده‌اند. هم چنین از میان فناوری‌های موجود جهت پیاده‌سازی فرآیند ذخیره و بازیابی که در بخش ۳-۲ اشاره شد، از Parliament استفاده شد که در شکل ۳ تصویری از محیط نرم‌افزاری این ابزار برای ذخیره و پرس‌وجو و بازیابی داده‌ها مشاهده می‌شود. علت انتخاب Parliament بعنوان فناوری اصلی در این فرآیند پیاده‌سازی، تبعیت کامل آن از استانداردهای GeoSPARQL و پشتیبانی از تمامی توابع مکانی تعریف شده در این استاندارد، توسط Parliament است.



شکل ۳- محیط کاری Parliament، (الف) پرس‌وجو از داده‌ها، (ب) بازیابی اطلاعات و نتیجه پرس‌وجو

در فرآیند پیاده‌سازی و با استفاده از مجموعه داده اشاره شده چندین هزار محتوای داده ذخیره شده است و پرس‌وجوهای مختلف از آن به عمل آمده است. نقاط مورد استفاده در این مجموعه، تحت کلاس ex:PointOfInterest (زیرکلاس geo:Feature) ذخیره شده اند. این کلاس داده‌ها می‌تواند در سه‌گانه‌های دیگر به داده‌های توصیفی مربوط به خود متصل شوند که در این تحقیق از اطلاعات مختصات، نام، ID نقاط و رسته نقطه مورد توجه (درصورت وجود) استفاده شده است. برای برخی نقاط با استفاده از منابع در دسترس، اطلاعاتی که در پایگاه‌داده مدلسازی نشده بودند و به صورت متن و سند در وب موجود هستند با رویکرد affordance مدلسازی شده و با تبدیل به قالب RDF به مجموعه اطلاعات ضمیمه شد. اگر تمامی Affordance نقاط در سه‌گانه‌های ذخیره شده مختلف در نظر گرفته‌شوند، می‌توان با گرفتن یک پرس‌وجو "تمامی اطلاعات نقطه مورد توجه" همه اطلاعات ذخیره شده مربوط به آن نقطه را بازیابی کرد. در بخش الف شکل ۴ همه اطلاعات مربوط به یک پارک با نام خاص پرس‌وجو شده است. هم چنین به عنوان مثال برای ذخیره مختصات یک نقطه از یک عبارت بصورتی مطابق بخش ب شکل ۴ استفاده می‌شود.

```
SELECT ?s ?p ?o
WHERE  {
   ?s gn:name "Park Restaurant" .
   ?s ?p ?o
}
```
الف

```
<http://www.opengis.net/ont/geosparql
#asWKT>
"<http://www.opengis.net/def/crs/OGC/
1.3/CRS84>
 POINT (8.4603523 1.4866109)"^
^<http://www.opengis.net/ont/sf#wktLi
teral> .
```
ب

شکل ۴- نمونه بیان و پرس‌وجو در استاندارد GeoSPARQL، (الف) پرس‌وجوی تمام اطلاعات مرتبط با یک نقطه، (ب) بیان مختصات یک نقطه

در نگاه اول ممکن است این نوع بیان اطلاعات مکانی طولانی به نظر برسد، اما این نوع بیان داده‌های مکانی همزمان با ذخیره مختصات نقطه، سیستم مختصات مربوط به نقطه را نیز ذخیره می‌نماید و با اختصاص یک URI به هر محتوای داده امکان دسترسی مستقیم به داده‌ها را از هر جایی تحت وب فراهم می‌آورد. موضوع اختصاص URIها دقیقاً راه حل RDF برای مساله پراکندگی توزیع منابع داده بر روی کارسازهای مختلف داده تحت وب است که امکان دسترسی به منابع مختلف را برای کاربران گسترده فراهم می‌آورد.

```
SELECT ?p
WHERE {
?p a ex:S.AIRP ;
geo:hasGeometry ?pgeo .
?pgeo geo:asWKT ?pwkt .
FILTER(geof:distance(?pgeo, ?pwkt,
units:km) < 50)
}
```
الف

```
SELECT ?link ?name ?lat ?lon
WHERE  {
   ?link gs:within(51.129725
-0.895386 51.833232 0.645447) .
?link gn:name ?name .
   ?link gn:featureCode ex:AIRP .
   ?link geo:lat ?lat .
   ?link geo:long ?lon
}
```
ب

شکل ۵- نمونه استفاده از توابع مکانی در پرس‌وجوهای مرتبط مکانی در چارچوب GeoSPARQL

به‌عنوان مثالی از پرس‌وجوی مکانی در نظر بگیریم که می‌خواهیم این پرس‌وجو را انجام دهیم: " کدام فرودگاه‌ها درون یک محدوده تعریف شده قرار دارند؟". در این صورت به یک تابع مکانی within نیاز خواهیم داشت. در استاندارد GeoSPARQL این تابع به صورت gs:within موجود است. اطلاعات بازیابی شده توسط پرس‌وجوی فوق که یک گراف از سه‌گانه‌ها است، در شکل ۶ ملاحظه می‌گردد. در پرس و جوی مشابه اگر فرودگاه‌های به فاصله کمتر از پنجاه کیلومتر از مرکز لندن مدنظر باشد می‌توان طبق قسمت الف شکل۵ که از تابع مکانی geof:distance استفاده کرده عمل کرد.

ارزیابی مدلسازی معنایی- مکانی از دو جهت قابل بررسی است. نخست نتایج بازیابی شده از پرس‌وجوهایی که هم در پایگاه‌های داده رابطه‌ای و هم در پایگاه‌های داده گرافی قابل نگارش هستند مقایسه شدند. در واقع این مقایسه، مقایسه بین نتایج بازیابی شده به زبان SQL و GeoSPARQL است. به‌این منظور توابع مکانی برروی داده‌های محدوده مورد نظر در پایگاه داده رابطه‌ای

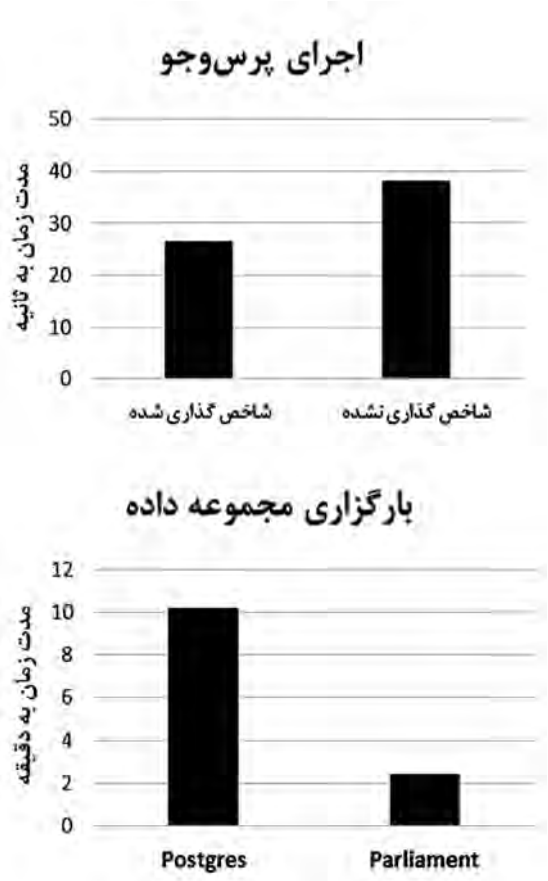
امتحان شده و به زبان SQL پرس‌وجو صورت گرفته شد. پایگاه داده PostGIS 2 for PostgreSQL8.4 به عنوان پایگاه داده رابطه‌ای تحت بررسی انتخاب شد. همچنین داده‌های محدوده مورد نظر که به قالب مکانی- معنایی تبدیل شدند و در پایگاه داده گرافی Parliament ذخیره شدند، توسط زبان استاندارد GeoSPARQL مورد پرس‌وجو قرار گرفتند.

شکل ۶- (الف)گراف داده‌ای که توسط پرس‌وجوی GeoSPARQL بازیابی شده است (ب) نمایش داده‌های بازیابی شده روی نقشه

پرس‌وجوهای بررسی شده از نقطه نظر غیرمکانی دارای پیچیدگی زیادی نیستند اما همانطور که در نمونه-های GeoSPARQL مشاهده شد هر چه نوع پرس‌وجوی ما از نظر مکانی دارای پیچیدگی بیشتری باشد به نگارش عناصر بیشتر و توابع بیشتری برای بازیابی اطلاعات مطلوب نیاز است. ازاین رو از میان تمامی اطلاعات موجود در فایل خروجی OSM تنها از نقاط استفاده شد و از میان کل لایه‌های اطلاعاتی لایه‌های GM_TYPE، ID، Layer، Name_en انتخاب شدند. برای تبدیل قالب ابتدا فایل خروجی OSM توسط نرم‌افزار Global Mapper V14.0.0 به قالب shapefile تبدیل شد و سپس برای تبدیل به قالب سه‌گانه‌های RDF از برنامه متن باز TripleGeo V3 استفاده شد. نکته قابل توجه درمورد داده‌های مدنظر این تحقیق این بود که مجموعاً ۱۲۹۰۱۲ نقطه مورد توجه به پایگاه داده تزریق شد که در پایگاه داده رابطه‌ای با در نظر گرفتن ۹۸۷ ستون برای هر نقطه 10^9 * 1.273 رکورد تشکیل شد که تنها حدود پنج درصد عناصر جدول رابطه‌ای غیر null و بامعنا بودند که به وضوح ناکارآمدی مدلسازی رابطه ای در مورد داده‌های تُنُک[1] را نشان می‌داد.

علاوه بر مقایسه پایگاه‌های داده رابطه‌ای و گرافی، دیدگاه دوم برای ارزیابی مدلسازی این پژوهش استفاده از شاخص‌های مکانی است. در شکل ۷ مشاهده می‌شود که در صورت استفاده از شاخص مکانی (که پایگاه داده Parliament واجد قابلیت شاخص مکانی و مکانی- زمانی برای داده‌های معنایی است) مدت زمان مورد نیاز برای انجام پرس‌وجوها بهبود می‌یابد. هم‌چنین مدت زمان مورد نیاز برای بارگزاری یک مجموعه داده عظیم با محتوای یکسان در دوپایگاه داده مقایسه شده است .

۱ Sparse



شکل ۷- مقایسه مدت زمان مصرف شده جهت بارگزاری و اجرای پرس‌وجو

۶-نتیجه‌گیری و افق آینده

در این تحقیق همراه با معرفی و مروری بر مفاهیم وب معنایی، و فناوری‌های موجود در این زمینه ، نحوه کاربرد آن برای مدیریت اطلاعات affordance نقاط مورد توجه بررسی شد. اگر چه پایگاه‌های داده که از مدل وب معنایی استفاده می‌کنند هنوز به گستردگی استفاده از پایگاه داده‌های رابطه‌ای نرسیده‌اند اما توانایی ارائه راه حل برای مقابله با چالش‌های تعامل‌پذیری و پراکندگی منابع اطلاعاتی وهم‌چنین تُنُک بودن اطلاعات در یک ناحیه خاص مدنظر، گسترش هر روزه انواع داده‌های مکانی تحت وب و نیاز به استفاده از آن‌ها نوید اقبال بیشتر به این رویکرد را می‌دهد.

در این تحقیق با استفاده از داده‌های معمول OSM مربوط به کشور انگلیس و تبدیل آن‌ها به داده‌ها در قالب RDF و استفاده از توصیفات نقاط مکانی با رویکرد بر مبنای affordance، توانایی گفتمان وب معنایی در مدیریت داده‌های ناهمگن مکانی سنجیده شده است.

استفاده از داده‌های مربوط به کشور ایران در مدلسازی مکانی- معنایی با دو مشکل روبرو است نخست از نظر کمی می‌توان گفت که تعداد بسیار اندکی از داده‌های مکانی در سایت OSM ثبت شده است به طوری که در یک محدوده مربعی‌شکل به ضلع دو کیلومتر در مرکز تهران تنها سیصد نقطه مورد توجه ثبت شده است در حالی که در محدوده ای با همین مساحت در شهر لندن حدود صدهزار داده ثبت شده است. مانع دوم مدلسازی مکانی- معنایی برای ایران دشواری و یا عدم دسترسی به اطلاعات مربوط به یک نقطه مورد توجه است در حالی که در مجموعه داده بررسی شده در این مقاله با چنین مانعی مواجه نشدیم. در صورت دسترسی آسانتر و بیشتر به اطلاعات مربوط به مکان‌های مورد توجه در ایران مدلسازی به روش معنایی- مکانی می‌تواند نقش کمک کننده‌ای داشته باشد.

آنچنان که مشاهد شد مانند پایگاه داده‌های رابطه ای امکان استفاده از توابع مکانی نیز در پرس‌وجوهای وب معنایی فراهم آمده است و کاربران قادر هستند تا از مقدار معتنابه داده در پایگاه داده‌های RDF استفاده کنند. بسیاری از فناوری‌ها هنوز به طور کامل از تمامی توابع و کارکردهای GeoSPARQL پشتیبانی نمی‌کنند اگر چه ممکن است از استانداردهای موجود مشابه GeoSPARQL تبعیت کنند. از طرف دیگر مدلسازی بر مبنای affordance توانایی تعریف مکان‌ها را به گونه‌ای داراست که هم فهم بشری ساده‌تری را نیاز دارد و هم امر محاسبات جستجوی نقاط مورد توجه را تسهیل می‌کند. تلفیق دو مفهوم وب معنایی و رویکرد affordance مبنا می‌تواند به عنوان یک ابزار قابل اتکا و کارا در ذخیره سازی و بازیابی اطلاعات مکانی در نظر گرفته شود. کاهش سطح پیچیدگی در نگارش پرس‌وجوها که به تلاش بشری کمتری جهت درک آن‌ها نیاز داشته باشد، پیروی مؤسسات و فناوری‌های مختلف از یک استاندارد واحد از جمله GeoSPARQL و تلاش برای ارائه شاخص گذاری برای اطلاعات مکانی RDF به گونه‌ای کاراتر تبدیل داده‌ها از مدل‌های مختلف کلاسیک و از منابع متنوع به مدل بر مبنای affordance از جمله چالش‌های پیش روی گفتمان این پژوهش است.

مراجع

[1] Lin, Y.-W. (2011), A Qualitative Enquiry into OpenStreetMap Making. New Review of Hypermedia and Multimedia, PP. 53-71

[2] Simon Scheider, W.K. (2008), Affordance-based Categorization Of Road Network Data Using A Grounded Theory Of Channel Network. Journal of Geographical Information Science, PP. 1-19

[3] Consortium, O.G. (2012), OGC GeoSPARQL - A Geographic Query Language for RDF Data, http://www.opengis.net/doc/IS/geosparql/1.0OGC

[4] Siegfried Handschuh, S.S. (2003), Annotation for the Semantic Web, Amsterdam the Netherlands: IOSPress PP. 240

[5] Robert Battle, D.K. (2012), Enabling the Geospatial Semantic Web with Parliament and GeoSPARQL. Semantic Web journal,PP. 355-370

[6] Gibson, J.J. (1986), the Ecological Approach To The Visual Perception Of Pictures. Routledge

[7] GeoKnow committee (2012), GeoKnow Making the Web an Exploratory for Geospatial Knowledge, PP. 114

[8] Ye, M., Shou, D., Lee, W.C., Yin, P., Janowicz, K(2011), On The Semantic Annotation Of Places In Location-Based Social Networks, in The 17th ACM SIGKDD international conference on Knowledge Discovery and Data mining New York, NY, USA

[9] Michele Ruta, F.S., Saverio Ieva,Giuseppe Loseto, Eugenio Di Sciascio (2012), Semantic Annotation of OpenStreetMap Points of Interest for Mobile Discovery and Navigation.

[10] Alkyoni Baglatzi, M.K., Marinos Kavouras(2012). Semantifying OpenStreetMap. In Proceedings of the 5th International Terra Cognita Workshop - Foundations, Technologies and Applications of the Geospatial Web. 2012. CEUR Workshop Proceedings

[11] Jordan, T., Raubal, M., Gartrell, B. and Egenhofer, M.J,(1998) An Affordance-Based Model Of Place In Gis, in 8th Int. Symposium on Spatial Data Handling

[12] O. Fonts, J.H., L. Díaz,C. Granell,(2010) Opensearch-Geo: The Simple Standard For Geographic Web Search Engines IV JORNADAS DE SIG LIBRE