نشریه علمی - پژوهشی علوم و فنون نقشه برداری، دوره ششم، شماره ۱، شهریور ماه ۱۳۹۵

ارائه یک مدل مفهومی، برای بهبود کیفیت ذخیره‌سازی داده‌های جغرافیایی داوطلبانه، در زمینه تناسب داده با هدف کاربر

محمد عشقی*[۱]، علی‌اصغر آل‌شیخ[۲]

[۱]دانشجوی کارشناسی ارشد سیستم‌های اطلاعات مکانی - دانشکده مهندسی ژئودزی و ژئوماتیک - دانشگاه صنعتی خواجه نصیرالدین طوسی
m.eshghi@mail.kntu.ac.ir

[۲] استاد دانشکده مهندسی ژئودزی و ژئوماتیک - دانشگاه صنعتی خواجه نصیرالدین طوسی
(قطب علمی مهندسی فناوری اطلاعات مکانی)
alesheikh@kntu.ac.ir

(تاریخ دریافت آبان ۱۳۹۴، تاریخ تصویب اسفند ۱۳۹۴)

چکیده

بررسی کیفیت، از جمله مهمترین چالش‌های موجود بر سر راه داده‌های جغرافیایی داوطلبانه می‌باشد. امروزه، اندازه‌گیری کیفیت این داده‌ها عمدتاً شامل فرآیندهای مستند کردن و اندازه‌گیری خطا با کمترین نگرانی در مورد برآورده کردن نیازهای مختلف و متنوع کاربران است. علی‌رغم طیف وسیع موجود در پذیرش شاخصِ "تناسب برای هدف" به عنوان المان کلیدی ارزیابی کیفیت اطلاعات جغرافیایی، این شاخص عملاً در تولید و به اشتراک‌گذاری داده‌ها مورد توجه قرار نمی‌گیرد. بنابراین، هدف این مقاله، ارائه مدلی مفهومی است که با استفاده از آن بتوان داده‌های ورودی به پایگاه داده داوطلبانه را به صورت درست و بهینه ذخیره‌سازی کرد. عوارض جغرافیایی و در نتیجه داده‌های اخذ شده از آن‌ها دارای کاربردهای متفاوتی هستند. با توجه به این موضوع، ممکن است کاربرد متفاوتی از داده در هنگام استفاده، مد نظر کاربران براساس نیازهایشان باشد. همچنین، ممکن است که کاربردهای مختلفی از یک داده توسط کاربران مختلف در تولید و ذخیره‌سازی آن داده، مد نظر باشد. در نتیجه، در این مقاله با طراحی یک مدل مفهومی، سعی در اعمال تمهیداتی در فرآیند ذخیره‌سازی این نوع از داده‌ها شده است تا فرصتی برای استفاده بهتر از داده‌ها ایجاد شود. این کار با تعریف مجموعه برچسب‌هایی انجام شده است که بیانگر نقش داده در کاربردهای مختلف هستند. در نتیجه، ذخیره‌سازی داده‌ها بهبود یافته که می‌تواند منجر به پردازش و ارائه بهتر داده‌ها براساس پرسش (نیاز و هدف) کاربر شود. برای پیاده‌سازی مفهوم پیشنهادی، از الگوریتم اندازه‌گیری تشابه معنایی Zhao استفاده شده است. داده‌های موجود و همچنین داده‌های ورودی جدید، با توجه به مدل مفهومی ارائه شده پردازش و ذخیره می‌شوند. برای ارزیابی، پاسخ ارائه شده به پرسش‌های مختلف کاربران در دو پایگاه داده با نوع ذخیره‌سازی موجود (فعلی) و جدید (با اعمال مدل پیشنهادی) در دو سناریو مقایسه شده است. همچنین نتایج پاسخ‌های ارائه شده به گروهی از کاربران، بهبود عملکرد پایگاه داده جدید را به میزان بیش از ۵۰ درصد نشان می‌دهد.

واژگان کلیدی: اطلاعات جغرافیایی داوطلبانه، کیفیت داده، تناسب برای هدف

* نویسنده رابط

۱- مقدمه

در سال‌های اخیر و با توجه به رشد سریع فناوری‌هایی نظیر وب ۲[۱]، رایانش ابری[۲]، و معنا[۳]، تغییرات عمیقی در زمینه داده، اطلاعات و خصوصاً داده‌های مکان‌مرجع[۴] صورت پذیرفته است. به وجود آمدن مرحله جدیدی در اینترنت به‌وسیله حرکت از وب۱[۵] به وب۲ امکان استفاده از دانش و داده‌های تولید شده توسط مردم عادی را که در برخی موارد تنها منبع داده می‌باشند، فراهم کرده است [۱،۲]. در نتیجه ساختار تولید داده از حالت بالا-به-پایین[۶] یعنی از سمت دولت به سمت مردم به حالت پایین-به-بالا[۷] که در آن نقش اصلی را مردم ایفا می‌کنند، تغییر یافته است [۳،۴،۵]. این نوع از داده‌ها که توسط عموم مردم تولید می‌شود، مفهومی را تحت عنوان داده‌های داوطلبانه[۸] ایجاد کرده‌اند [۱،۲]. در حال‌حاضر، برنامه‌ها و پروژه‌هایی در سرتاسر دنیا در این زمینه وجود دارند که از جمله آن‌ها می‌توان به Maps-Google Flicker، و Panoramia اشاره کرد [۶]. بنابراین، با توجه به جریان داده عظیمی که بوجود آمده است، ضرورت بهترین استفاده از این داده‌ها اجتناب‌ناپذیر است. برای این منظور، نیاز به زیرساخت‌هایی برای مدیریت و بهره‌برداری از این جریان داده احساس می‌شود که از جمله آن‌ها، می‌توان به ذخیره‌سازی و پردازش بهینه داده‌ها اشاره کرد [۶].

اعتبار[۹]، اطمینان[۱۰]، و کیفیت[۱۱] از جمله مهمترین مسائل در خصوص داده‌های مکانی داوطلبانه[۱۲] (VGI) هستند [۷،۸،۹،۱۰،۱۱،۱۲]. VGI منبع عظیمی از داده‌ها می‌باشد که از چالش‌های مهم آن، نبود اعتبار و اطمینان است، علت این مسئله، تولید داده‌ها توسط مردم عادی و بدون داشتن تخصص لازم است که با روند تولید داده‌ها توسط متخصصین تفاوت بسیاری دارد [۸]. برای مثال، از تولیدکنندگان متخصصِ داده‌های مکانی انتظار می‌رود تا داده‌هایی با سطح دقت و صحت معینی را تولید کنند، در حالیکه، ممکن است تولیدکنندگان داده‌های VGI یا درک درستی از عارضه مورد نظر نداشته باشند و یا تنها توانایی تشخیص و درک جنبه‌های محدودی از یک عارضه مکانی را داشته باشند [۱۳]. Crompvoets و همکاران (۲۰۱۰) بیان می‌کنند که نگرانی اصلی در رابطه با اطلاعات مکانی داوطلبانه مربوط به دقت و صحت در زمینه مکانی بوده است و حتی در صورت رفع این نگرانی‌ها، بُعد دیگری از کیفیت اطلاعات جغرافیایی داوطلبانه، یعنی "تناسب برای هدف[۱۳]"، هنوز به عنوان یک مسئله خودنمایی می‌کند که نتیجه آن عدم استفاده و یا استفاده ناصحیح از این داده‌ها است [۱۴].

مانع موجود بر سر داده‌های VGI این است که برخلاف مجموعه داده‌های مرجع (رسمی) که از یک مجموعه کلمات فنیِ تأییدشده که در سیستم‌های پایگاه داده مکانی استاندارد استفاده می‌کنند، مجموعه داده‌های VGI با استفاده از یک زبان طبیعی[۱۴] (بر خلاف زبان‌هایی همچون زبان کد در کامپیوتر) مورد استفاده توسط انسان‌ها، تولید و ذخیره می‌شوند [۸]. براساس Scheider (۲۰۱۱) اصطلاحات استفاده شده توسط مشارکت‌کنندگان برای توصیف پدیده‌های جغرافیایی مبهم هستند. در نتیجه، بعضی از پروژه‌های VGI همچون OpenStreetMap (OSM)، مشارکت‌کنندگان را ملزم به استفاده از اصطلاحات فنی از قبل تعریف شده می‌کنند [۱۵].

Longhorn و Blakemore (۲۰۰۸) بیان کرده‌اند که ارزش یک مجموعه داده معین برای اشخاص و سازمان‌های مختلف بعلت تفاوت در نیازها، و اهداف متفاوت است. این مورد را به خصوصیات و ویژگی‌های بیرونی[۱۵] اطلاعات جغرافیایی ارجاع داده‌اند و ارزش نسبی داده‌ها را به پتاسیل رفع نیاز و متناسب بودن با اهداف و انتظارات کاربران، مربوط دانسته‌اند [۱۶]. به علاوه، اطلاعات جغرافیایی، ترکیبی از خصوصیات و ویژگی‌های کلی و جزئی عوارض موجود را ارائه می‌دهند، در نتیجه، پردازش این اطلاعات بسیار مهم است، بخصوص در حالتی که بتواند به خوبی نیازهای کاربر را برآورده سازد [۱۷].

۱ Web 2.0
۲ Cloud Computing
۳ Semantic
٤ Geo-referenced data
٥ Web 1.0
٦ up - bottom
۷ bottom – up
۸ User Generated Content (UGC)
۹ Credibility
۱۰ Reliability
۱۱ Quality
۱۲ Volunteered Geographic Information
۱۳ Fitness-for-use
۱٤ Natural Language
۱٥ Extrinsic

بنابراین، با توجه به مطالب بیان‌شده نیاز به بهبود المان "تناسب برای هدف" برای داده‌های VGI احساس می‌شود. سؤال مطرح در این مقاله بدین صورت می‌باشد که: "۱- با چه روشی می‌توان کاربردهای مختلف داده‌ها را در هنگام ذخیره‌سازی در نظر گرفت؟ ۲- روشی که ارائه می‌شود به چه مقدار نسبت به پایگاه داده‌های داوطلبانه موجود (در این حالت، OSM) بهبود ایجاد می‌کند؟". بنابراین: هدف این مقاله، ارائه مدلی مفهومی است که با استفاده از آن بتوان داده‌های ورودی در پایگاه داده داوطلبانه را به صورت درست و بهینه، ذخیره‌سازی کرد. در این راستا از برچسب‌ها برای نشان دادن کاربردهای مختلف داده‌های ورودی و همچنین ذخیره‌سازی آنها استفاده شده است. عوارض جغرافیایی و در نتیجه داده‌های اخذ شده از آن‌ها دارای کاربردهای متفاوتی هستند. با توجه به این موضوع، ممکن است کاربردی متفاوتی از داده در هنگام استفاده، مد نظر کاربران براساس نیازهایشان باشد. همچنین، ممکن است که کاربردهای مختلفی از یک داده توسط کاربران مختلف در تولید و ذخیره‌سازی آن داده، مد نظر باشد. در نتیجه، در این مقاله سعی در اعمال تمهیداتی در فرآیند ذخیره‌سازی این نوع از داده‌ها شده است تا فرصتی برای استفاده بهتر از داده‌ها ایجاد شود. این کار با تعریف مجموعه برچسب‌هایی انجام شده است که بیانگر نقش داده در کاربردهای مختلف هستند. در نتیجه، ذخیره‌سازی داده‌ها بهبود یافته که می‌تواند منجر به پردازش و ارائه بهتر داده‌ها براساس پرسش (نیاز و هدف) کاربر باشد.

در بخش دوم این مقاله، گزیده‌ای از تحقیقات انجام شده در این زمینه ارائه شده و تفاوت کارهای انجام شده با روش معرفی شده بررسی شده است.

در بخش سوم به مبانی نظری تحقیق پرداخته شده است. برای پیاده‌سازی و ارزیابی روش پیشنهادی از الگوریتم اندازه‌گیری و تعیین تشابه معنایی Zhao (۲۰۰۹) [۱۸]، استفاده شده است که در این بخش به معرفی آن پرداخته شده است.

در بخش چهارم به طراحی و معرفی یک مدل مفهومی برای ذخیره‌سازی داده‌های مکانی داوطلبانه (داده‌های خطی) پرداخته شده است. بنابراین، نیازهای مختلف کاربران و همچنین کاربردهای مختلفی که داده‌های تولید شده می‌توانند داشته باشند، تا حد امکان شناسایی شده است. در نتیجه، برای هر داده کاربردهای متفاوتی را می‌توان در نظر داشت ولی ممکن است تنها یک کاربرد خاص از آن داده توسط تولید کننده درنظر گرفته شود، بنابراین سایر کاربردهای آن داده عملاً از دست می‌رود. ولی با استفاده از مدل پیشنهادی سعی شده است تا از بروز این مشکل جلوگیری شود. برای مثال، کاربر یک عارضه از مجموعه راه را بصورت "بزرگراه" ذخیره می‌کند ولی این عارضه دارای کاربردهای دیگری همچون "راه‌شریانی" نیز می‌باشد که اگر در نظر گرفته نشود در پاسخ به پرسشی که در آن مجموعه "راه‌های‌شریانی" نیاز است، ارائه نمی‌شود. حال آنکه با کمک مدل پیشنهادی، سعی شده است که تا حد امکان سایر کاربردهایی که یک داده می‌تواند داشته باشد، در نظر گرفته شده و مانع از بروز مشکلاتی از این قبیل شد.

در بخش پنجم مدل مفهومی پیشنهادی در یک پایگاه داده مکانی پیاده‌سازی شده است. داده جدید به پایگاه داده وارد شده و با توجه به مدل ارائه شده ذخیره می‌شود. برای پیاده‌سازی مدل مفهومی از محیط برنامه‌نویسی C# و کتابخانه‌های ArcObject، و برای مدیریت پایگاه داده از نرم افزار مدیریت پایگاه داده SQLSERVER استفاده شده است. برای اتصال C# به SQLSERVER از LINQ و برای اتصال SQLSERVER به ArcMap از ArcSDE، و در نهایت، از نرم‌افزار ArcMap جهت نمایش خروجی استفاده شده است. داده‌های استفاده شده، مجموعه عوارض خطیِ راه‌های وب‌سایت OSM به عنوان داده‌های جغرافیایی داوطلبانه می‌باشند.

در بخش ششم، کارکرد مدل ارزیابی شده است. بدین منظور، توان پاسخگویی به پرسش‌های کاربران در دو پایگاه داده با نوع ذخیره‌سازی فعلی (موجود) و با اعمال مدل پیشنهادی (جدید)، برای پرسش‌های مختلفی که توسط کاربران انجام شده است، ارزیابی شده و علت خروجی‌های حاصل شده، بررسی و تحلیل شده است. برای این منظور، دو حالت مختلف درنظر گرفته شده است، ۱) پرسش یک کاربر در دو پایگاه داده و ۲) پرسش‌های متفاوت دو کاربر مختلف در دو پایگاه داده. همچنین نتایج پاسخ‌های ارائه شده به گروهی از کاربران، برای ارزیابی بهبود عملکرد پایگاه داده جدید، بررسی شده‌اند.

در نهایت، در بخش هفتم، به بحث و نتیجه‌گیری، و ارائه پیشنهادهایی برای کارهای آینده، پرداخته شده است.

۲- پیشینه تحقیق

نخستین تحقیقات در حوزه کیفیت مجموعه داده‌های VGI و به طور خاص داده‌های OSM، در رابطه با ارزیابی کیفیت کلی مجموعه داده‌های VGI، مقایسه داده‌های VGI با داده‌های مرجع برای یک منطقه مشابه بوده است. هدف این تحقیقات بررسی دقت مکانی[۱] و کامل بودن داده‌ها[۲] بوسیله مقایسه داده‌های OSM با داده‌های مرجع دولتی بوده که در کشورهای مختلفی همچون انگلستان [۱۹]، فرانسه [۲۰]، و آلمان [۲۱،۲۲]، انجام شده است.

اما بعد از رفع نسبی نگرانی‌ها در مورد وجود داده‌های با دقت و قابل استفاده، سؤال موجود این است که چرا بهره‌وری بیشینه از داده‌ها مشاهده نمی‌شود [۱۴]؟

در تحقیق صورت گرفته توسط Mondzech and Sester (۲۰۱۱)، چگونگی مفید بودن داده‌ها برای کاربردها بیان شده است. در پاسخ به این سؤال و در مطالعه موردی بین پایگاه داده OSM و ATKIS با مقایسه نتایج اعمال الگوریتم کوتاهترین مسیر برای یافتن کوتاهترین فاصله میان دو نقطه بیان کرده‌اند که پایگاه داده‌ای که مجموع فاصله بین دو نقطه دارای طول کوتاهتری است برای یافتن کوتاهترین مسیر بین دو نقطه مناسب‌تر است. در واقع در این تحقیق، مقایسه‌ای میان پایگاه داده‌های موجود انجام شده است و تنها مناسب بودن داده‌های موجود را برای یک کاربرد خاص بررسی کرده‌اند [۲۳]. در این تحقیق تنها به این موضوع پرداخته شده است که کدام پایگاه داده و چرا برای کاربردی خاص همچون مسیریابی مناسب است، اما کاری در رابطه با بهبود وضع داده‌های موجود بیان نشده است، در صورتی که در مقاله حاضر بجای یافتن پایگاه داده مناسب برای کاربردی خاص، سعی به غنی کردن داده‌ها شده است.

در تحقیق انجام شده توسط Kalantari و همکاران (۲۰۱۳)، فراداده[۳] به عنوان یک المان کلیدی برای افزایش ارزش داده‌ها و بهبود استفاده از داده‌ها در نظر گرفته شده است. در نتیجه، برای افزایش میزان استفاده از داده‌ها در مواردی که امکان این استفاده وجود دارد، و غنی کردن توصیفات داده‌ها اقدام به ایجاد فراداده برای داده‌های VGI در دو حالت مختلف و اجرای آنها در پایگاه داده Wikimapia کرده‌اند: ۱) به صورت ضمنی[۴] که در آن مشارکت‌کنندگان اطلاعی از مشارکت خود ندارند، و ۲) به صورت صریح[۵] که در آن مشارکت‌کنندگان به صورت آگاهانه تلاش به ایجاد و بهبود فراداده برای داده‌ها می‌کنند [۲۴]. تفاوت مقاله حاضر با این تحقیق، بررسی جامع‌تر کاربردهایِ داده‌ها می‌باشد. زیرا در رابطه با استفاده از خود کاربران برای ذخیره کاربردها بازهم ممکن است کاربردی از داده‌ها در نظر گرفته نشده باشد و توسط کاربر جدید در زمان دیگر درخواست شود؛ زیرا تمرکز اصلی بر روی بررسی کاربردهای مختلف داده نبوده است.

Xiang Zhang (۲۰۱۵) به بررسی تأثیر متفاوت بودن برچسب عوارض که نشان‌دهنده کاربرد داده‌ها هستند در هنگام ذخیره‌سازی داده‌ها پرداخته است. در این تحقیق بیان شده است که کاربران محدود به انتخاب بخشی از کاربردهای عوارض هستند، زیرا در سایت OSM تنها مجموعه‌ای از برچسب‌های محدود مورد استفاده می‌باشند. ایشان در کل مجموعه خطاهای ممکن در رابطه با برچسب‌ها را به صورت زیر معرفی کرده‌اند: نزدن برچسب، برچسب‌های ناسازگار، برچسب‌های غلط، محدودیت‌های بیان نشده یا به غلط بیان شده؛ از جمله تقاطع، و در نهایت یک طرفه و یا دو طرفه بودن مسیر که به اشتباه وارد شده‌اند. مجموعه این عوامل باعث بوجود آمدن ناسازگاری در داده‌ها شده که در استفاده باعث بروز مشکل می‌شوند. در نتیجه، روشی برای شناسایی و رفع خطای اینگونه داده‌ها ارائه کرده است [۲۵]. در این روش برچسب‌ها یا خصوصیات توصیفی که یا غلط بوده و یا به صورت ناسازگار برای یک عارضه (راه) ایجاد شده‌اند شناسایی شده است. در واقع ممکن است کاربران برای قسمت‌های مختلف یک راه نوع‌های "trunk" و "primary" را ذخیره کرده باشند که برای بیان کاربردهای مختلفی که راه می‌تواند داشته باشد، برای هر قطعه از آن راه بیان شده است. نتیجه اینکه، این کار باعث بروز ابهام در مورد این داده‌ها می‌شود. در این مقاله تنها بررسی و اصلاح این مشکلات پرداخته شده و سعی شده است که مفاهیم استفاده شده برای ذخیره نوع عارضه برای یک

۱ Positional Accuracy
۲ Data Completeness
۳ Metadata
٤ Implicit Model
٥ Explicit Model

عارضه خاص در قطعات مختلف آن عارضه یکسان باشند. در صورتی که بازهم نسبت به در نظر گرفتن کاربردهای مختلف داده‌ها تلاشی انجام نشده است.

Devillers و Arnaud (۲۰۱۵) یک سیستم توصیه‌گر برچسب برای وب‌سایت OSM با نام OSMantic, طراحی کرده‌اند که به صورت خودکار برچسب مناسب و مرتبط با عارضه ورودی را به کاربر در لحظه ورود داده پیشنهاد می‌دهد [۲۶]. کاری که در این تحقیق انجام شده است در واقع هدایت کاربر به استفاده از برچسب‌های درست از میان برچسب‌های موجود در پایگاه داده OSM می‌باشد. حال آنکه، حتی اگر برچسب‌های استفاده شده از جانب کاربران به درستی انتخاب شوند، بعلت محدودیت برچسب‌ها به یک کاربرد، سایر کاربردهای داده‌ها در نظر گرفته نمی‌شود.

تحقیقاتی نیز به بررسی تناسب داده برای هدف از جنبه دیگری پرداخته‌اند. از جمله آنها می‌توان به بررسی‌های انجام‌شده توسط واحدی (۱۳۹۳)، Bordogna و همکاران (۲۰۱۳) اشاره کرد. در اینگونه تحقیقات با استفاده از مفاهیم کیفیت درونی، کیفیت بیرونی و کیفیت عملی همراه با موارد دیگری همچون اعمال متغییرهای زبانی در تصمیم‌گیری و استفاده از عملگرهای تصمیم-گیری همچون OWA[1] روش‌هایی پیشنهاد کرده‌اند تا کاربران بدون تخصص هم بتوانند یک برآورد حدودی از کیفیت داده‌های VGI داشته باشند [۲۷،۲۸]. اما بازهم به استفاده از داده‌های موجود توجه شده است بدون اینکه روندی برای ثبت و ذخیره بهتر داده‌های داوطلبانه معرفی گردد.

با مقایسه تحقیقات بررسی شده به نظر می‌رسد که بهبود کیفیت داده‌های داوطلبانه در راستای افزایش استفاده از داده‌ها تنها برای وضعیت موجود داده‌ها انجام شده است. حال آنکه در این تحقیق تلاش به ایجاد روشی مناسب برای ذخیره اصولی و دقیقتر داده‌ها با در نظر گرفتن کاربردهای متفاوتشان شده است. از اهداف این تحقیق، فرآهم آوردن فرصتی برای بیشینه کردن میزان بهره‌وری از داده‌ها در هنگام استفاده می‌باشد که این امر با درنظر گرفتن کاربردها در لحظه ذخیره‌سازی داده‌ها در این تحقیق پیشنهاد شده است.

۳- مبانی نظری تحقیق

مدل‌های مختلفی برای تعیین میزان تشابه داده‌ها ارائه شده‌اند. این روش‌ها را می‌توان به ۵ گروه اصلی براساس تشابه مفاهیم دسته بندی کرد [۲۹] : ۱) مدل هندسی، ۲) مدل ویژگی، ۳) مدل ترازبندی، ۴) مدل شبکه، و ۵) مدل انتقال.

این پنج دسته منحصر بفرد نبوده و می‌بایست به صورت ترکیبی برای اندازه‌گیری تشابه بین دو مفهوم استفاده شوند. با توجه به نوع مدل کردن داده‌ها، به دو طریق می‌توان میزان مشابه بودن ساختارهای معنایی را بررسی کرد: ۱) اندازه‌گیری میزان شباهت هندسی و موقعیت مکانی، و ۲) اندازه‌گیری میزان مشابهت ویژگی‌های توصیفی براساس برچسب‌های عارضه. برای این کار از الگوریتم Zhao [۱۸]، برای محاسبه P-Rank برای محاسبه امتیاز تشابه معنایی بین عوارض استفاده شده است. این الگوریتم برعکس سایر الگوریتم‌ها همانند SimRank، الگوریتم بازگشتی براساس لینک‌های ورودی و خروجی بوده و در آن، دو هستنده مشابه هستند اگر: ۱- آن‌ها توسط هستنده مشابه مرجع شده باشند، و ۲- آن‌ها به هستنده مشابه مرجع‌دهی کنند. در این الگوریتم تشابه‌معنایی به صورت $R(a,b) \in [0,1]$ بین دو ورتکس (a,b) از شبکه G در حالتی که $a, b \in G$ باشند در قالب معادله ۱، به صورت زیر نشان داده می‌شود [۱۸].

$$R_{k+1}(a,b) = \lambda \times \frac{C}{|I(a)||I(b)|} \sum_{i=1}^{|I(a)|} \sum_{j=1}^{|I(b)|} R_k\left(I_i(a), I_j(b)\right) + (1-\lambda) \times \frac{C}{|O(a)||O(b)|} \sum_{i=1}^{|O(a)|} \sum_{j=1}^{|O(b)|} R_k\left(O_i(a), O_j(b)\right) \quad (۱)$$

جاییکه:

مقدار λ و همچنین C پارمتر دمپینگ[2]، مقادیری ثابت بوده که بیانگر جهت لینک بوده و در بازه‌ی [۱و۰] می‌باشند. پارامترهای I و O به ترتیب بیانگر لینک ورودی و خروجی می‌باشند. امتیاز تشابه P-Rank بین دو ورتکس a و b در تکرار kام است. این مقدار در هر تکرار به روز می‌شود تا به بیشینه مقدار تکرارها دست یافت. مقادیر خروجی بین صفر (نبود تشابه معنایی) و یک (تشابه کامل) متغیر است. شکل ۱، شبه کد الگوریتم اندازه‌گیری تشابه معنایی Zhao را نشان می‌دهد [۱۸].

[1] Ordered Weighted Averaging

[2] Damping factor

Algorithm 1: P-Rank (G, λ, C, k)

Input : An IN G, the relative weight λ, the damping factor C, the iteration number k
Output: P-Rank score $s(a,b)$, $\forall\ a,b \in G$

```
foreach a ∈ G do                                   /* Initialization */
    foreach b ∈ G do
        if a == b then R(a,b) = 1
        else R(a,b) = 0
while (k > 0) do                                   /* Iteration */
    k ⟵ k − 1
    foreach a ∈ G do
        foreach b ∈ G do
            in ⟵ 0
            foreach i_a ∈ I(a) do
                foreach i_b ∈ I(b) do
                    in ⟵ in + R(i_a, i_b)
            R*(a,b) ⟵ λ * C*in / (|I(a)||I(b)|)
            out ⟵ 0
            foreach o_a ∈ O(a) do
                foreach o_b ∈ O(b) do
                    out ⟵ out + R(o_a, o_b)
            R*(a,b) += (1 − λ) * C*out / (|O(a)||O(b)|)
    foreach a ∈ G do                               /* Update */
        foreach b ∈ G do
            R(a,b) = R*(a,b)
return R(*,*)
```

شکل۱- شبه کد الگوریتم اندازه‌گیری تشابه معنایی Zhao [۱۸]

۴- روش پیشنهادی

روشی که در اینجا مطرح می‌گردد، مدلی مفهومی است که در آن، ذخیره‌سازی داده‌ها براساس کاربردهای مختلفی که دارند، انجام می‌پذیرد. این مدل از مدل‌های پیشنهادی و استفاده شده در زمینه CityGML [۳۰]، الهام گرفته است و براساس این تحقیق تغییر و توسعه یافته است. بنابراین، کاربردهای مختلف داده تعیین می‌شود و در نتیجه با وارد کردن هر کدام از کاربردهای داده که مد نظر کاربر در لحظه ورود آن داده است، سایر کاربردهای آن داده نیز در نظر گرفته شده و داده ورودی به صورت غنی‌تر ذخیره می‌شود. این تحقیق شامل مراحل زیر است:

۱- شناسایی و طبقه‌بندی کاربردهای مختلف هر نوع عارضه در مجموعه عوارض خطیِ راه‌ها

۲- اعمال برچسب‌های مرتبط با هر کاربرد

۳- پیاده‌سازی و ارزیابی

همچنین در هنگام ورود داده، می‌بایست موجود بودن و یا نبودن عارضه از قبل بررسی شود. برای این منظور از الگوریتم Zhao [۱۸]، استفاده شده است. این کار با بررسی خواص توصیفی عوارض انجام می‌شود. همچنین برای افزایش دقت از موقعیت مکانی عوارض نیز استفاده می‌شود.

۴-۱- طراحی مدل مفهومی ارتباط داده‌ها

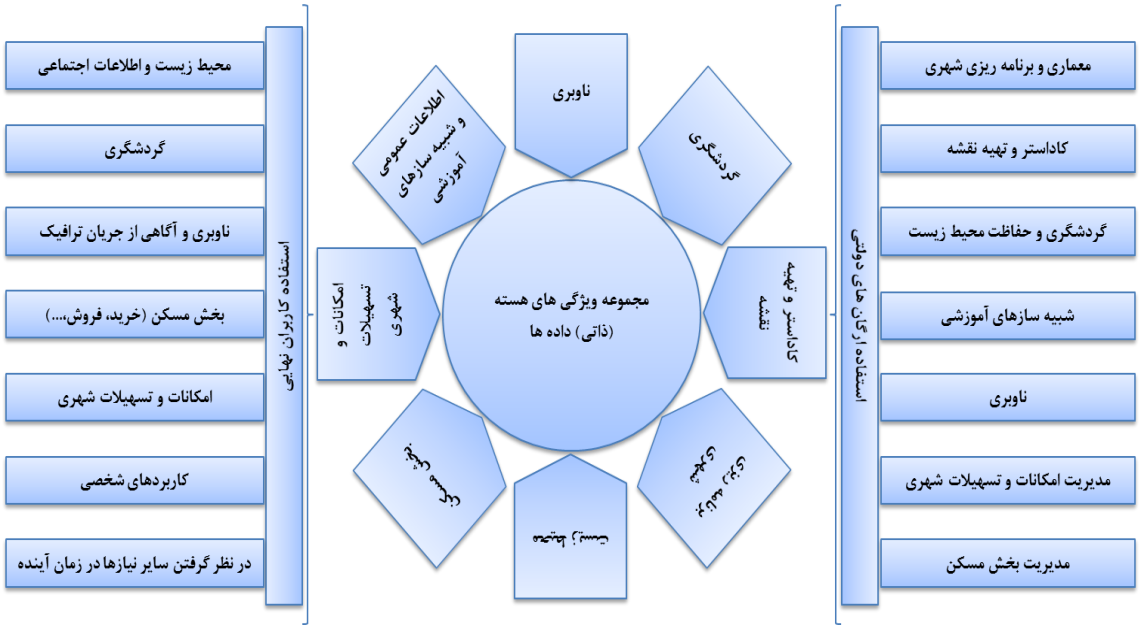
پس از انتخاب عوارض خطی از میان عوارض موجود و ورودی (جدید) در پایگاه داده، در اولین گام همانند شمای ارائه‌شده در شکل ۲، می‌بایست کاربردهای مختلف هر یک از عوارض موجود در دسته راه درنظر گرفته شود. بنابراین، ابتدا داده‌ها از نظر ویژگی‌های ذاتی که در واقع بیانگر نوع عارضه می‌باشد در هسته این مدل قرار می‌گیرند. در حالت کلی سه نوع عارضه نقطه‌ای، خطی، و پلیگونی را می‌توان برای ویژگی‌های هسته در نظر گرفت. سپس کاربردها در دو دسته کلی استفاده در ارگان‌های دولتی، و استفاده شخصی دسته بندی می‌شوند. در نتیجه، پس از تعیین نوع داده ورودی، هر یک از کاربردهایشان که در مدل در نظر گرفته شده است بررسی شده و برچسب‌های متناظر برای هر کاربرد تعیین می‌شوند. در واقع این کار به این معنی است یک المان خطی ممکن است دارای کاربردهای مختلفی باشد. برای مثال یک جاده در پایگاه داده‌های مربوط به وزارت راه ممکن است با عنوان (کاربرد) "بزرگراه" درنظر گرفته شود درحالی که همان عارضه در شهرداری به عنوان (کاربرد) "راه شریانی" ذخیره می‌شود. این مسئله در هنگام بازیابی عارضه براساس کاربرد آن و استفاده از آن می‌تواند مشکلاتی همچون ناتوانی سیستم در پاسخگویی به پرسش کاربران را درپی داشته باشد.

برای مجموعه داده‌ها، دو سطح استفاده در نظر گرفته شده است. حالت اول، استفاده دولتی و سازمانی و حالت دوم، استفاده شخصی توسط کاربران نهایی[۱]. با توجه به شکل ۲، کاربردهای مختلفی که یک داده می‌تواند داشته باشد، تا حد امکان شناسایی شده‌اند. در گروه اول که مربوط به مدیریت کلان و ارگان‌های دولتی است، کاربردهای کاداستر و تهیه نقشه، گردشگری، شبیه‌سازی‌های آموزشی (بخصوص در بخش مدیریت بحران)، ناوبری، مدیریت امکانات و تسهیلات شهری، مدیریت بخش مسکن، معماری و برنامه‌ریزی شهری، و حفاظت محیط زیست شناسایی شده‌اند. در گروه دوم نیز که در مورد بخش مربوط به استفاده کاربران نهایی است، به صورت مشابه، کاربردهای ناوبری، گردشگری، اطلاعات عمومی (همانند هزینه)، جریان ترافیک، امکانات و تسهیلات شهری، بخش مسکن، اطلاعات توزیع جمعیت و علایق، و محیط زیست شناسایی شده است. البته در این دسته، با توجه به اینکه نیازهای کاربران متغیر بوده و در نتیجه ممکن است که تعاریف جدیدی از کاربرد داده‌ها در آن نیازها ایجاد شود، باید امکان تعبیه دسته‌های جدید

۱ End-user

درنظر گرفته شود تا بتوان وضعیتی انعطاف‌پذیر برای افزایش ارزش داده‌ها را در نظر گرفت.

به عنوان نتیجه، همانند مثال یاد شده برای راه، داده ورودی براساس کاربردهای مختلفی که دارد، می‌تواند برچسب‌های متفاوتی داشته باشد. در نتیجه، هر کاربر می‌تواند براساس کاربردی که مد نظر دارد، داده مورد نظر را بازیابی کند که نتیجه آن استفاده بهتر از داده‌ها برای بازیابی و ارائه به کاربر است.

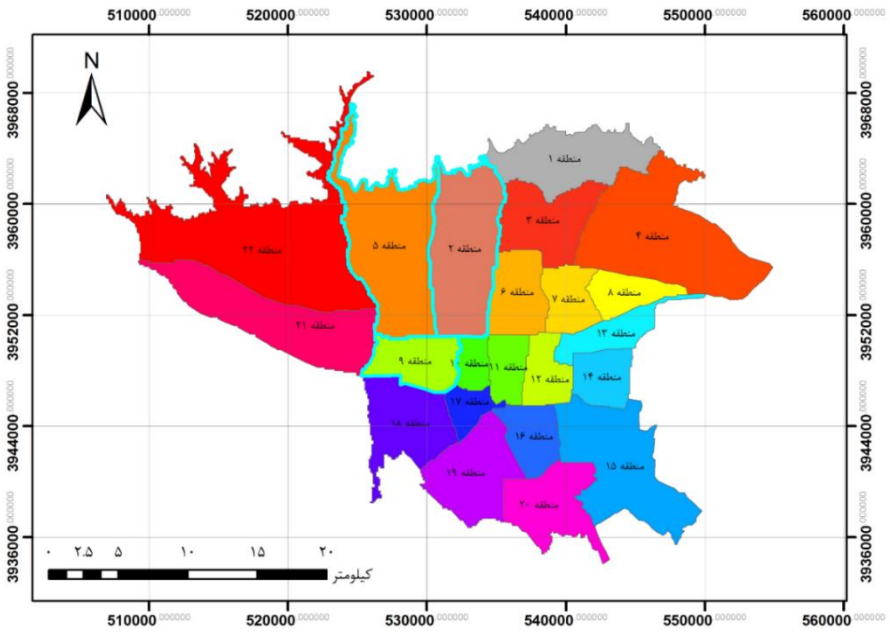


شکل۲- دسته‌بندی کاربردهای مختلف داده‌ها، توسعه یافته براساس [۳۰]

۵- پیاده‌سازی

همانند شکل ۳، منطقه مورد مطالعه در این تحقیق، منطقه‌های ۲، ۵، و ۹ شهرداریِ شهر تهران می‌باشد. داده‌های شبکه راه‌ها و معابر تهران برای سایت OSM به عنوان مجموعه داده داوطلبانه مورد استفاده این تحقیق هستند.

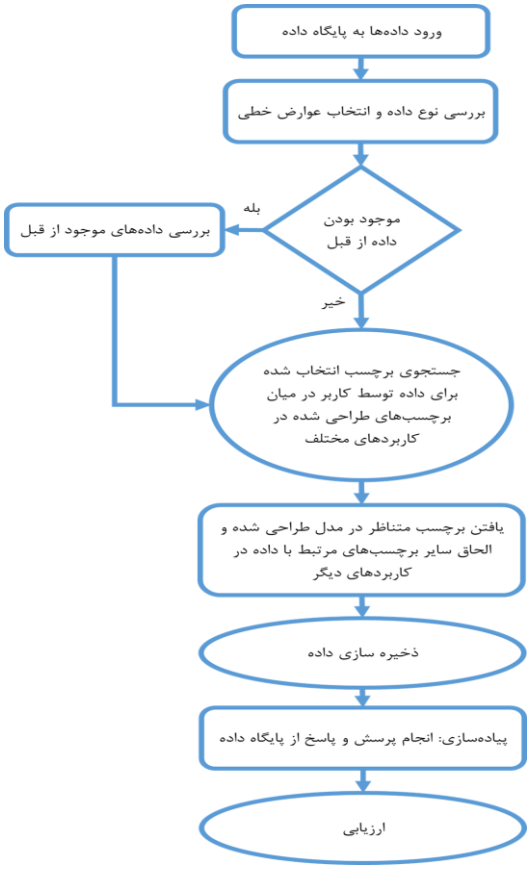
استفاده از داده‌های سایت OSM بدلیل شناخته شدن این سایت به عنوان یکی از نمونه‌های موفق پروژه‌های داوطلبانه است. در ضمن، داده‌های این سایت به صورت کاملاً رایگان و به راحتی قابل دریافت هستند. داده‌های استفاده شده منطبق با آخرین بروزرسانی‌های انجام شده می‌باشند.



شکل۳- مناطق شهرداریِ شهر تهران - مناطق مورد مطالعه ۲، ۵، و ۹

برای پیاده‌سازی مدل مفهومی از محیط برنامه نویسی C# و کتابخانه‌های ArcObject، و برای مدیریت پایگاه داده از نرم افزار مدیریت پایگاه داده SQLSERVER استفاده شده است. برای اتصال C# به SQLSERVER از LINQ و برای اتصال SQLSERVER به ArcMap از ArcSDE، و در نهایت، از نرم‌افزار ArcMap جهت نمایش خروجی استفاده شده است.

برای ارزیابی مدل طراحی شده از مجموعه‌ای داده‌های خطیِ راه‌ها و ویژگی‌های توصیفی مربوط به هرکدام برای ذخیره مناسب آنها در هنگام ورود به پایگاه داده استفاده شده است. سپس به بررسی این مهم پرداخته شده است که با وجود تفاوت ایجاد شده در هنگام ذخیره‌سازی داده، میزان بهبودی در پاسخ‌ها و نتایج ارائه شده به پرسش‌های مختلف به چه میزان بوده است. فلوچارت پیاده‌سازی مدل ارائه شده بصورت شکل ۴ می‌باشد.



شکل۴- فلوچارت پیاده‌سازی و ارزیابی مدل پیشنهادی

در ابتدا، داده‌های ورودی به پایگاه داده از نظر نوع داده (نقطه، خط، پلیگون) بررسی می‌شوند. در این ارزیابی تنها از مجموعه داده خطیِ راه‌ها برای ارزیابی مدل استفاده شده است. با توجه به قابلیت ArcObject در تشخیص هندسه عوارض و اینکه کاربران داده‌های ورودی را با هندسه مشخص و از پیش تعریف شده به سیستم نرم‌افزاریِ OSM معرفی می‌کنند، مشکلی در رابطه با تشخیص هندسه عوارض وجود نداشته و در نتیجه تنها عوارض خطی انتخاب می‌شوند. شبه کد الگوریتم اجرا شده برای انتخاب عارضه و ذخیره‌سازی آن براساس مدل مفهومی پیشنهادی، مطابق شکل ۵ می‌باشد.

Algorithm 2: Data Entry & Labeling

Input : Entered Data by Users, Existing OSM Database
Output: OSM Database with new Data labeling

```
1  foreach a∈ID do
2    if a=="New Feature"
3      if a=="line"do
4        while (a_(i)≠ϕ)
5          if a_(i)={every word which describes the attribute of Road}
6            b_(j)=a_(i)
7        while (b_(j)≠ϕ)
8          if Lables_(i)=b_(j)
9            b_(j)=find {core label, other labels}
10       while (b_(j)≠ϕ)
11         a=a+b_(j)
12   if a=="Existing Feature"
13       while (a_(i)≠ϕ)
14         if a_(i)={every word which describes the attribute of Road}
15           if Lables_(i)=a_(i)
16           a_(i)=find {core label, other labels}
17           a=a+other labels
```

شکل۵- شبه کد الگوریتم ورود و برچسب گذاری داده‌ها

در الگوریتم شکل ۵، پس از انتخاب عارضه خطی، مجموعه برچسب‌هایی که با توجه به مدل طراحی شده برای عوارض خطی راه در کاربردهای مختلف استخراج شده‌اند، در نظر گرفته می‌شوند. سپس، برچسبی که برای داده از طرف کاربر به عنوان توصیفات عارضه برای ذخیره‌سازی در نظر گرفته شده است شناسایی شده و با برچسب‌های طراحی شده مقایسه می‌شود و برچسب متناظر استخراج می‌شود. این برچسب متناظر در واقع کلید یافتن سایر برچسب‌های مرتبط با دیگر کاربردهای داده می‌باشد. در نتیجه، در مرحله بعدی سایر کاربردهای داده در نظر گرفته شده و برچسب متناسب با هر کاربرد به داده نسبت داده می‌شود. در نهایت داده‌ای ذخیره می‌شود که علاوه بر داشتن اطلاعاتی که توسط کاربر وارد می‌شود، اطلاعات مربوط به سایر کاربردها را نیز دارد. در این حالت داده ورودی غنی‌تر شده و بنابراین در پاسخ به پرسش‌هایی که جنبه‌ها و کاربردهای مختلفی از داده را مد نظر دارند به درستی عمل می‌کند. در صورت موجود بودن داده از قبل، تنها به غنی سازی اطلاعات توصیفی و اختصاص برچسب‌هایی که داده می‌تواند داشته باشد ولی ممکن است نداشته باشد، پرداخته می‌شود؛ با این کار به تدریج اطلاعاتی

که از قبل در پایگاه داده VGI وجود دارند به مرور زمان بهبود می‌یابند. در صورت موجود نبودن نیز همین روند انجام می‌شود. تنها دلیل بررسی موجود بودن یا نبودن داده، جلوگیری از ایجاد افزونگی داده است.

۶- ارزیابی

تا به اینجا پایگاه داده جدیدی ایجاد شده که عوارض ذخیره شده در آن از نظر اطلاعات توصیفی و شناسایی کاربردهای مختلفی که دارند، به صورت بهینه ذخیره شده‌اند. حال بمنظور ارزیابی مدل طراحی شده در پایگاه داده جدید، پرسش‌های کاربران مختلف در کاربردهای متفاوتی که مد نظر آنهاست، مطرح شده، و پاسخ‌های ارائه شده توسط پایگاه داده‌های ۱) موجود و ۲) جدید (تغییریافته به واسطه روش پیشنهادی) مورد ارزیابی قرار گرفته و نتایج بررسی می‌شوند. طرز عملکرد هر دو پایگاه داده برای پاسخ به پرسش کاربران مطابق شبه کد نشان داده شده در شکل ۶ است. اما نوآوری مدل طراحی شده در آنجاست که پایگاه داده‌ای که با این مدل عمل می‌کند منبع غنی‌تری از اطلاعات را بواسطه ذخیره بهتر داده‌ها برای پاسخ دهی در اختیار دارد.

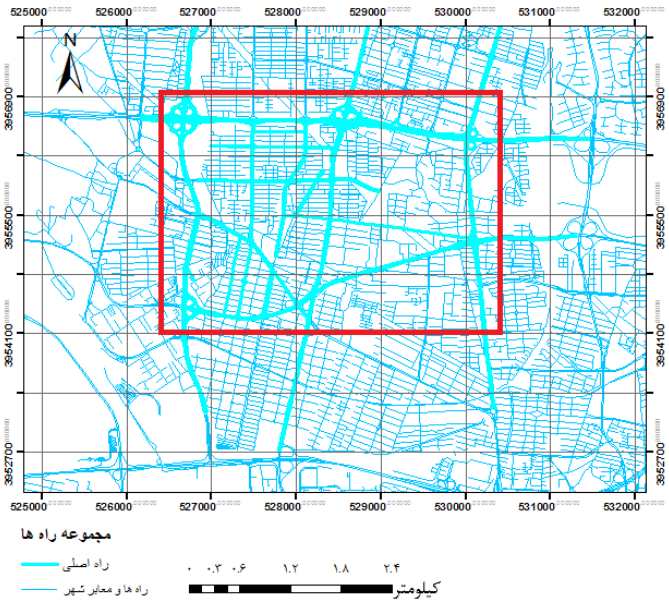
```
Algorithm 3: Answering to Query

    Input : Query from Users
    Output: Answers to Queries

1   Selecting the required data type
2   Selecting the suitable data by their labels
3         Searching among objects with matched data type and labels
4   Selecting the application
5   using appropriate data for answering the query
6   retrieving the answers
```

شکل۶- شبه کد الگوریتم پاسخ به پرسش کاربران در دو پایگاه داده

شکل ۷، پاسخ ارائه شده به پرسشی یکسانی که توسط یک کاربر مطرح شده است را در پایگاه داده جدید نشان می‌دهد. در این پرسش، در یک محدوده مشخص؛ منطقه غرب تهران محدود شده توسط بزرگراه اشرفی از شرق، بزرگراه باکری از غرب، بزرگراه همت از شمال و بزرگراه حکیم از جنوب، کاربر به دنبال خانه‌هایی در محدوده راه‌های اصلی بوده تا بتواند سطح دسترسی خوبی داشته باشد اما در مجموعه داده‌های OSM راه‌ها دارای برچسب‌هایی هستند که از مجموعه برچسب‌های از پیش تعریف شده توسط OSM انتخاب شده‌اند، مانند، Primary, Secondary, Residential (شکل ۸)، بعلاوه تنها بخشی از راه‌ها همچون بزرگراه‌ها از پیشوند نوع راه در فیلد نام عارضه بهره برده‌اند. در نتیجه خروجی توسط پایگاه داده موجود (OSM) برگردانده نمی‌شود. با استفاده از تعاریف مختلف راه در استانداردهای وزارت راه، و شهرداری، کاربردهای راهِ مورد نیاز در این پرسش و برچسب‌های متناسب برای هر کاربرد استخراج شده و کاربر توانسته در پایگاه داده بروز شده به مقصود خود برسد که همانا شناسایی راه‌های اصلی به منظور تصمیم‌گیری در مورد انتخاب خانه مناسب است. در این مورد فرض شده است که کاربر هیچگونه دانش خارجی نسبت اصلی یا فرعی بودن راه‌ها براساس نامشان ندارد. در شکل ۷ همانطور که نمایش داده شده است، می‌توان انتخاب صحیح راه‌های اصلی و ارائه آنها به کاربر را مشاهده کرد.



شکل۷- خروجی پایگاه داده جدید در نمایش راه‌های اصلی در محدوده مورد نظر (کادر قرمز)

Table

راه ها و معابر شهر

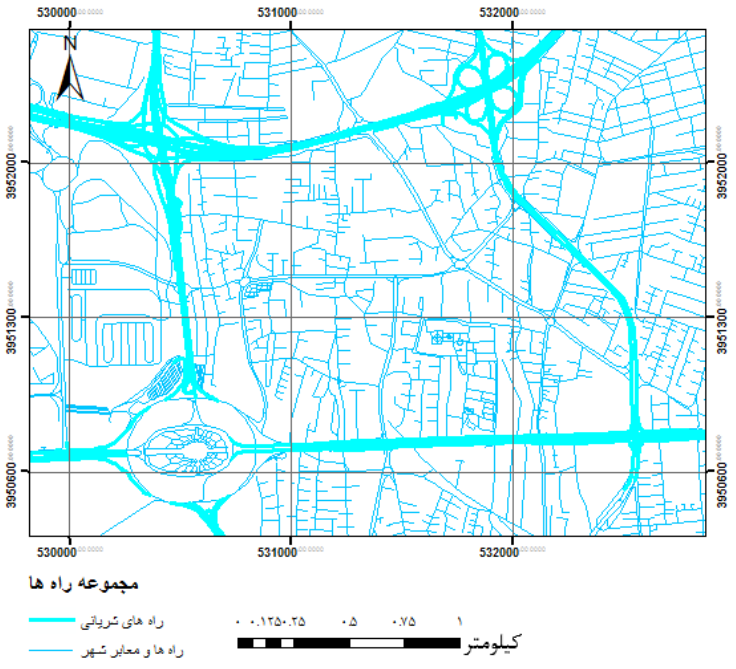
FID	Shape	osm_id	name	ref	type	oneway	bridge	tunnel	maxspeed
14	Polyline	4294716			secondary	1	0	0	0
17	Polyline	4294743	بزرگراه محمدعلی جناح		trunk	1	0	0	0
18	Polyline	4294746	شهید صارمی		secondary	0	0	0	0
23	Polyline	4294977	بزرگراه شهید ستاری		trunk	1	0	0	80
70	Polyline	4297325			secondary	1	0	0	0
71	Polyline	4297326			trunk_link	1	0	0	0
81	Polyline	4299543	بلوار جنت‌آباد		primary	1	0	0	0
86	Polyline	4299686	بلوار لاله		secondary	1	0	0	0
104	Polyline	4304594			trunk_link	1	0	0	0
105	Polyline	4304609			trunk_link	1	0	0	0
106	Polyline	4304611	بلوار ابوذر		secondary	1	0	0	0
107	Polyline	4304614	بزرگراه شهید حکیم		trunk	1	0	0	0
108	Polyline	4304615	بزرگراه شهید حکیم		trunk	1	0	0	0
124	Polyline	4313919			secondary_link	1	0	0	0
169	Polyline	4320096			trunk_link	1	0	0	0
170	Polyline	4320097			trunk_link	1	0	0	0
188	Polyline	4332766			trunk_link	1	0	0	0
190	Polyline	4332797	بزرگراه شهید حکیم		trunk	1	0	0	0
192	Polyline	4332799	بزرگراه شهید حکیم		trunk	1	0	0	0
194	Polyline	4334453			motorway_link	1	0	0	0
291	Polyline	4354806	کاشانی		trunk	1	0	0	0
292	Polyline	4354812			trunk_link	1	0	0	0
293	Polyline	4354813	هشت‌متری		residential	1	0	0	0
325	Polyline	4368251	میدان صادقیه		trunk	0	0	0	0
408	Polyline	4433434	بلوار هزارداران		primary	1	0	0	0

0 (1582 out of 39872 Selected)

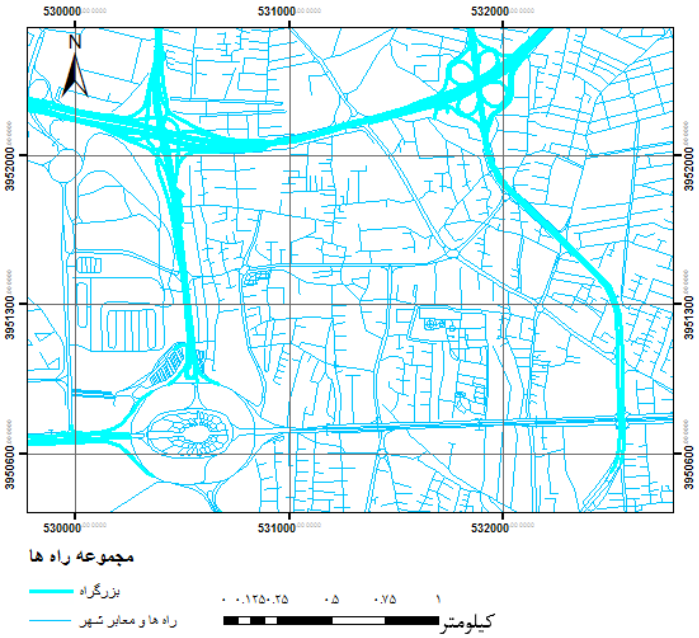
راه ها و معابر شهر

شکل۸- ذخیره نوع داده‌ها در پایگاه داده OSM

پاسخ‌های ارائه شده به پرسش‌های متفاوتی که توسط کاربران مختلف در کاربردهای متفاوت مطرح شده است نیز می‌تواند حاوی اطلاعات یکسان ولی با هدفی متفاوت باشد. برای مثال دو کاربر "الف" و "ب" به ترتیب نیازمند اطلاعاتی در کاربردهای طراحی و برنامه ریزی شهری و جریان ترافیکی هستند. کاربر "الف" نیاز به اطلاعات "راه‌های شریانی" داشته و کاربر "ب" به اطلاعاتی نیاز داشته که مجموعه "بزرگ‌راه‌ها" را در اختیار وی قرار دهد. با توجه به موقعیت کاربر در میدان آزادی، همانطور که شکل‌های ۹ و ۱۰ نشان می‌دهند، بزرگراهایی همچون بزرگراه "شیخ فضل‌الله نوری" توسط پایگاه داده جدید به هردو کاربر ارائه شده است اما در کاربردهایی متفاوت. همچنین خیابان‌هایی همچون خیابان آزادی که تنها یک کاربرد را پشتیبانی می‌کنند، تنها به کاربر "الف" ارائه شده‌اند که با توجه به نیاز کاربر منطقی می‌باشد. این در حالی است که پایگاه داده موجود در پاسخ‌گویی به این پرسش‌ها ناتوان است.



شکل۹- خروجی پایگاه داده جدید به پرسش کاربر الف - مجموعه راه‌های شریانی



شکل ۱۰- خروجی پایگاه داده جدید به پرسش کاربر ب - مجموعه بزرگراه‌ها

همان‌طور که در قسمت پیاده‌سازی بیان شد، علت بهبود پاسخ‌های ارائه‌شده به کاربر را می‌توان به علت برچسب‌گذاری صحیح داده در هنگام ذخیره‌سازی برشمرد. همان‌طور که جداول ۱ و ۲ نشان می‌دهند: داده‌های ذخیره‌شده در پایگاه داده جدید، کاربردهای بیشتری از داده‌ها را در نظر گرفته و بنابراین می‌تواند از داده‌های مناسب بیشتری برای پاسخ به پرسش کاربر استفاده کند. این در حالی است که به علت عدم توجه به کاربردهای مختلف داده‌ها در پایگاه داده‌های معمول، نمی‌توان به پرسش‌های مطرح‌شده از جانب کاربر به‌درستی پاسخ داد. المان کلیدی موجود در مدل پیشنهادی این است که نقش داده‌ها در کاربردهای متفاوت به‌طور هم‌زمان در ذخیره‌سازی آن در نظر گرفته می‌شود، که نتیجه آن، استفاده بهتر از داده‌ها است. درحالی‌که در تحقیقات پیشینِ صورت گرفته و مطرح‌شده به استفاده بهتر از داده‌های موجود با همان وضعیت قبلی (تنها در نظر گرفتن یک کاربرد) تلاش شده است.

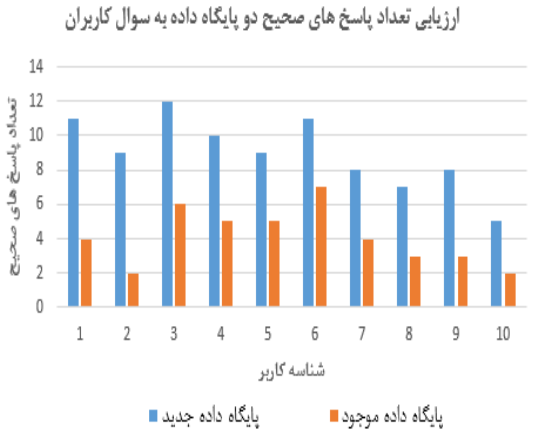
جدول۱- مثالی از شناسایی و دسته‌بندی داده ورودی به پایگاه داده

	نوع عارضه	برچسب تولید کننده و ذخیره‌سازی در پایگاه داده موجود	پایگاه داده جدید	
			برچسب	کاربرد
داوطلب ۱	خطی	Primiary	بزرگراه	ناوبری و جریان ترافیکی
داوطلب ۲	خطی	Highway	راه شریانی	طراحی و برنامه ریزی شهری

جدول۲- مثالی از نحوه بازیابی داده براساس پرسش و پاسخ کاربر از پایگاه داده

	نوع عارضه خواسته شده	دسته مد نظر کاربر	دسته بازیابی شده (پایگاه داده موجود)	دسته بازیابی شده (پایگاه داده جدید)
داوطلب ۱	خطی	ناوبری و جریان ترافیکی	بازیابی بدون در نظر گرفتن کاربرد	ناوبری و جریان ترافیکی
داوطلب ۲	خطی	طراحی و برنامه ریزی شهری	بازیابی بدون در نظر گرفتن کاربرد	طراحی و برنامه ریزی شهری
داوطلب ۳	خطی	ناوبری و جریان ترافیکی	عدم ارائه اطلاعات (نداشتن برچسب مناسب توسط داده)	ناوبری و جریان ترافیکی

همچنین، یک گروه ۱۰ نفره از کاربران در نظر گرفته شده است و هر کاربر، ۱۵ پرسش را در هرکدام از پایگاه داده‌های موجود و جدید بیان کرده است. پاسخ‌های ارائه شده توسط پایگاه داده‌های موجود و جدید به پرسش‌های مطرح شده از جانب کاربران بررسی شده و درستی پاسخ‌های ارائه شده به آنها مورد ارزیابی قرار گرفته که به طور خلاصه در شکل ۱۱ نشان داده شده است. با توجه به شکل ۱۱، پایگاه داده جدید توانسته عملکرد بهتری را از خود نشان دهد. لازم به ذکر است، تمامی پاسخ‌های داده شده به کاربران توسط پایگاه داده جدید درست نبوده است که می‌تواند به علت محدودیتی باشد که در قسمت نتایج به آن پرداخته می‌شود.



شکل۱۱- نتایج ارزیابی کاربران در مورد پاسخ‌های ارائه شده توسط دو پایگاه داده موجود و جدید

۷- بحث، نتیجه‌گیری و پیشنهاد

رشد محبوبیت پروژه‌های VGI همچون OpenStreetMap، روند تولید، انتشار، و استفاده داده‌های جغرافیایی را از حالت سنتی تغییر داده است. اگرچه این داده‌های رایگان و به هنگام می‌توانند اغلب جایگزین ارزشمندی برای مجموعه داده‌های مرجع باشد، ولی، ارزیابی و بهبود کیفیت مجموعه داده‌های VGI به عنوان یک چالش مهم باقی می‌ماند. این مقاله به بررسی و طراحی مدلی مفهومی برای بهبود ذخیره‌سازی داده‌های VGI بمنظور افزایش کارایی داده‌ها، پرداخته است. کاربردهای مختلفی برای داده مشخص شده و در هنگام ذخیره‌سازی داده در نظر گرفته می‌شوند. با این کار، بازیابی اطلاعات متناسب با هر کاربرد در پاسخ به پرسش صورت گرفته از پایگاه داده انجام می‌شود.

نوآوری این تحقیق در بررسی داده‌ها در کاربردهای مختلف و ذخیره‌سازی آن به همراه داده به منظور افزایش کارایی داده‌ها می‌باشد. در نظر گرفتن کاربردهای مختلف بوسیله برچسب‌هایی که برای داده‌ها در نظر گرفته شده است، اعمال شد. ارزیابی مدل طراحی شده با وارد کردن مجموعه عوارض خطی در پایگاه داده انجام شده و اینکه نحوه ذخیره‌سازی با استفاده از مدل ارائه شده چه مقدار بهبود در پرسش و پاسخ‌های مکانی و ارائه اطلاعات به کاربران، ایجاد می‌کند. در حالیکه سایر تحقیقاتی که در پیشینه تحقیق به آنها اشاره شد، تنها هدف به انتخاب درست برچسب‌ها از میان برچسب‌های مورد استفاده داشتند، اما در این مقاله به افزایش طیف برچسب‌های مورد استفاده برای غنی کردن داده‌ها پرداخته شده است. بهبودی که نوآوری این تحقیق در آن سهیم است، افزایش توانایی پایگاه داده در پاسخ به پرسش‌های کاربران است. البته نقصی که وجود داشته و نیازمند تحقیقاتی در این زمینه است شناسایی هرچه کاملتر کاربردهای داده‌ها در سه دسته کلی عوارض نقطه‌ای، خطی، و پلیگونی برای هرچه بهتر شدن این مدل می‌باشد. از جمله محدودیت-های این مدل اشتراک بعضی از کاربردها در دو دسته کلی مطرح شده برای کاربردهای داده است؛ زیرا کاربردهایی وجود دارد که در هر دو سطح استفاده دولتی و شخصی وجود دارد، و همچنین وجود اشتراک که در میان زیردسته‌های هرکدام از این کاربردها ظاهر می‌شود که نیاز به تحقیقات بیشتر برای رفع این مشکل احساس می‌شود.

همچنین، برای کارهای آتی، با توجه به موقعیت استفاده از این مدل می‌توان هم در مرحله پردازش داده‌های موجود و هم در مرحله ورود داده از آن استفاده کرد که در مورد استفاده دوم، بررسی امکان استفاده از این مفهوم برای تولید یک فیلتر خودکار در زمان ورود داده پیشنهاد می‌شود.

مراجع

[1] Cooper, A., Coetzee, S., and Kourie, D., (2012). "Volunteered Geographical information – The Challenges." PoPositionIT, pp. 34-38.

[2] Baeza-Yates, R., (2009). "User generated content: how good is it?" In Proceedings of the 3rd workshop on Information credibility on the web, pp. 1-2, ACM.

[3] Haklay, M., Singleton, A., and Parker, C., (2008). "Web mapping 2.0: The neogeography of the GeoWeb." Geography Compass, 2(6), pp. 2011-2039.

[4] Goodchild M. F., (2007). "Citizens as sensors: the world of volunteered geography." GeoJournal, 69 (4), pp. 211-221.

[5] Craglia, M., (2007). "Volunteered Geographic Information and Spatial Data Infrastructures: when do parallel lines converge?" Presented at the Workshop on Volunteered Geographic Information, Santa Barbara, CA.

[6] Zielstra, D., Hochmair, H. H., and Neis, P., (2013). "Assessing the effect of data imports on the completeness of OpenStreetMap–a United States case study." Transactions in GIS, 17(3), pp. 315-334.

[7] Bishr, M., and Mantelas, L., (2008). "A trust and reputation model for filtering and classifying knowledge about urban growth." GeoJournal, 72(3-4), pp. 229-237.

[8] Elwood, S., (2008). "Volunteered geographic information: Future research directions motivated by critical, participatory, and feminist GIS." GeoJournal, 72(3), pp. 173–183.

[9] Flanagin, A., and Metzger, M., (2008). "The credibility of volunteered geographic information." GeoJournal, 72(3), pp. 137–148.

[10] Gouveia, C., and Fonseca, A., (2008). "New approaches to environmental monitoring: The use of ICT to explore volunteer geographic information." GeoJournal, 72(3), pp. 185–197.

[11] Jokar Arsanjani, J., Barron, C., Bakillah, M., and Helbich, M., (2013). "Assessing the quality of OpenStreetMap contributors together with their contributions." Proceedings of the 16th AGILE Conference, Leuven, Belgium.

[12] Sieber, R., (2007). "Geoweb for social change." Position Paper Presented at the Workshop on Volunteered Geographic Information, Santa Barbara, CA, December 13–14, 2007, 3pp.

[13] De Longueville, B., Ostlander, N., and Keskitalo, C., (2009). "Addressing vagueness in volunteered geographic information (VGI)—A case study." International Journal of Spatial Data Infrastructures Research, Special Issue GSDI-11.

[14] Crompvoets, J., de Man, E., and Macharis, C., (2010). "Value of spatial data: Networked performance beyond economic rhetoric." International Journal of Spatial Data Infrastructures Research, 5, pp. 96–119.

[15] Scheider, S., Keßler, C., Ortmann, J., Devaraju, A., Trame, J., Kauppinen, T., and Kuhn, W., (2011). "Semantic referencing of geosensor data and volunteered geographic information." In Geospatial semantics and the semantic web, Springer US, pp. 27-59.

[16] Longhorn, R., and Blakemore, M., (2007). "Geographic information: Value, pricing, production and consumption." Boca Raton: CRC Press.

[17] Craglia, M., and Nowak, J., (2006). "Report of international workshop on spatial data infrastructures: Cost-benefit/return on investment: Assessing the impacts of spatial data infrastructures, European Commission, Directorate General Joint Research Centre (Technical report)." Ispra: Institute for Environment and Sustainability.

[18] Zhao, P., Han, J., and Sun, Y., (2009). "P-Rank: a comprehensive structural similarity measure over information networks." In Proceedings of the 18th ACM conference on Information and knowledge management, pp. 553-562. ACM.

[19] Haklay, M., Basiouka, S., Antoniou, V., and Ather, A., (2010). "How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information." The Cartographic Journal, 47(4), pp. 315-322.

[20] Girres, J. F., and Touya, G., (2010). "Quality assessment of the French OpenStreetMap dataset." Transactions in GIS, 14(4), pp. 435-459.

[21] Ludwig, I., Voss, A., and Krause-Traudes, M., (2011). "A Comparison of the Street Networks of Navteq and OSM in Germany." In Advancing Geoinformation Science for a Changing World, Springer Berlin Heidelberg, pp. 65-84.

[22] Zielstra, D., Zipf, A., (2010). “A Comparative Study of Proprietary Geodata and VGI for Germany.” In 13th AGILE International Conference on Geographic Information Science. Guimaraes, Portugal. 10-14 May 2010.

[23] Mondzech, J., and Sester, M., (2011). “Quality analysis of OpenStreetMap data based on application needs.” Cartographica: The International Journal for Geographic Information and Geovisualization, 46(2), pp. 115-125.

[24] Kalantari M., Rajabifard A., Olfat H., and Williamson I., (2013). “Crowd-sourcing metadata for VGI.” AGILE 2013 – Leuven, May 14-17, 2013.

[25] Zhang, X., and Ai, T., (2015). “How to Model Roads in OpenStreetMap? A Method for Evaluating the Fitness-for-Use of the Network for Navigation.” In Advances in Spatial Data Handling and Analysis, Springer International Publishing, pp. 143-162.

[26] Vandecasteele, A., and Devillers, R., (2015). “Improving Volunteered Geographic Information Quality Using a Tag Recommender System: The Case of OpenStreetMap.” In OpenStreetMap in GIScience, Springer International Publishing, pp. 59-80.

[27] Vahedi, B., Alesheikh, A. A., Honarparvar, S., (2014). “Quantitative Assessment of Pragmatic Quality of Volunteered Geographic Information Using Fuzzy Linguistic Quantifiers and OWA Operator.” JGST; 3 (4) : pp. 65-76

[28] Bordogna, G., Carrara, P., Criscuolo, L., Pepe, M., and Rampini, A., (2014). “A linguistic decision making approach to assess the quality of volunteer geographic information for citizen science.” Information Sciences, 258, pp. 312-327.

[29] Schwering, A., (2008). “Approaches to Semantic Similarity Measurement for Geo-Spatial Data: A Survey.” Transactions in GIS, 12(1), pp. 5-29.

[30] Kolbe, T. H., (2009). “Representing and exchanging 3D city models with CityGML.” In 3D geo-information sciences. Springer Berlin Heidelberg, pp. 15-31.