

مدیریت کیفیت در داده‌های مکانی مردم گستر: تعیین کیفیت و ارائه آن به کاربر

رویا اسماعیلی^{۱*}، فرید کریمی‌پور^۲، فرزین ناصری^۳، علی اسماعیلی^۳

^۱ دانشجوی دکتری سیستم های اطلاعات مکانی - گروه مهندسی نقشه‌برداری - پردیس دانشکده‌های فنی - دانشگاه تهران
roya.esmaeili@ut.ac.ir

^۲ استادیار گروه مهندسی نقشه‌برداری - پردیس دانشکده‌های فنی - دانشگاه تهران
fkarimipr@ut.ac.ir

^۳ استادیار گروه سنجش از دور - دانشکده مهندسی عمران و نقشه‌برداری - دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته کرمان
naseri@icst.ac.ir
aliesmaeily@kgut.ac.ir

(تاریخ دریافت اردیبهشت ۱۳۹۲، تاریخ تصویب مهر ۱۳۹۲)

چکیده

داده‌های مکانی مردم گستر دائماً توسط کاربران در حال اضافه شدن، تغییر و یا حذف می‌باشند. به طوریکه از یک منطقه داده‌های گوناگون با خصوصیات مختلف تولید می‌شود. سیاست اغلب وب‌گاه‌های مردم گستر، ارائه آخرین نسخه موجود از داده‌ها به کاربر است، چرا که این اعتقاد وجود دارد که کیفیت داده‌ها به علت نظارت تعداد زیادی کاربر رو به بهتر شدن بوده و بنابراین، آخرین نسخه از داده، باکیفیت‌ترین آنها خواهد بود. این در حالی است که کیفیت داده‌های مکانی، دارای پارامترهای مختلفی از جمله میزان کامل بودن، دقت مکانی، سازگاری منطقی و به روز بودن است که ممکن است الزاماً همگی آنها در یک داده جمع نگردد. به عبارت دیگر، داده باکیفیت در کاربردهای مختلف، معانی متفاوتی خواهد داشت. در این مقاله، رویکردی ارائه شده است که در آن، کاربر تعیین کننده نوع پارامتر کیفیت بوده و داده مناسب برای کاربرد خویش را از میان کلیه داده‌های مکانی مردم گستر موجود انتخاب نماید. علاوه بر این، از آنجا که اغلب کاربران سیستم‌های مردم مکانی گستر، افراد عادی با دانش کم نقشه‌ای می‌باشند، به جای کاربرد فراداده برای ارائه پارامترهای کیفیت به کاربران، از روشهای بصری استفاده می‌گردد. رویکرد پیشنهادی، برای یک منطقه مورد مطالعه، پیاده‌سازی گردیده و نتایج مورد ارزیابی قرار گرفته است.

واژگان کلیدی: داده‌های مکانی مردم گستر، ارزیابی کیفیت، ارائه کیفیت

* نویسنده رابط

۱- مقدمه

سیستم اطلاعات مکانی^۴ [۵] و اطلاعات مکانی مردم گستر^۵ (عنوانی که برای اولین بار Goodchild در سال ۲۰۰۷ از آن استفاده کرد) نام برده شده است [۶]. وب-گاههایی نظیر OpenStreetMap (OSM) و Wikimapia که هدف آنها تهیه یک نقشه رایگان، مشارکتی و قابل تصحیح از کل جهان می باشد از جمله این محیطها هستند [۷].

با وجود سودمند بودن داده های مکانی تولید شده توسط مردم و مزایایی که نسبت به داده های رسمی دارند (از جمله جامع بودن، سرعت تهیه بالا و کم هزینه بودن [۸])، اما سوالات بسیاری در مورد کیفیت این داده ها مطرح بوده و تحقیقات بسیاری در این زمینه انجام شده است. تحقیقاتی مانند بررسی کلی دقت داده های مکانی مردم گستر [۹، ۱۰] و یا مطالعات موردی مانند بررسی دقت داده های موجود در وب گاه OpenStreetMap [۱۰-۱۲] از نمونه های تحقیقات انجام شده در این زمینه می-باشند.

این مقاله، با مروری بر فعالیتهای انجام شده در راستای مدیریت کیفیت داده های مکانی مردم گستر، رویکردی را برای تعیین کیفیت و ارائه آن به کاربران داده های مردم گستر، که عموماً افراد عادی هستند، ارائه نموده است. در این راستا در بخش دوم، کلیاتی در خصوص کیفیت داده های مردم گستر، شامل تضمین و تعیین کیفیت و همچنین ارائه کیفیت در این داده ها ارائه شده و رویکردهای موجود در این زمینه معرفی گردیده اند. بخش سوم، به تشریح رویکرد پیشنهادی تحقیق پرداخته است. این رویکرد، برای یک منطقه مورد مطالعه پیاده سازی شده که نتایج آن در بخش چهارم ارائه گردیده است. در نهایت، بخش پنجم، شامل نتیجه گیری و پیشنهاداتی برای تحقیقات آتی در این زمینه می باشد.

۲- کیفیت داده های مکانی مردم گستر

لازمه استفاده درست از هر نوع داده ای، آگاهی از کیفیت آن و اطمینان از مناسب بودن آن برای کاربرد مورد نظر می باشد. داده های مردم گستر نیز از این امر مستثنی نیستند. داده ای که توسط مردم عادی در وب ۲

ظهور اینترنت، موجب گسترش ارتباطات، تولید اطلاعات و به اشتراک گذاشتن آن ها گردید. در ابتدای پیدایش اینترنت، اطلاعات موجود بر روی آن، صرفاً توسط افراد کارشناس تهیه می شد و از طریق صفحات وب، در اختیار کاربران قرار می گرفت. در این گام از اینترنت که "وب ۱" نام دارد، کاربران، مصرف کنندگان صرف بوده و تأثیر چندانی بر روی داده های تولید شده نداشتند [۱]. این در حالی است که به غیر از افراد کارشناس و متخصص، افراد عادی هم توانایی تولید داده را دارند. این امر در گام بعدی توسعه اینترنت (که "وب ۲" نام گذاری شده است) مد نظر قرار گرفت. بدین ترتیب، کاربران توانستند علاوه بر استفاده از داده های موجود، خودشان نیز تولیدکننده داده بر روی اینترنت باشند. کاربران در این دوره، دیگر فقط مصرف کننده نبوده و خود آنها نیز در تولید داده سهیم هستند. به همین دلیل گاهی از وب ۲ با عنوان انتقال از "وب فقط خواندنی" به "وب خواندنی-نوشتنی" یاد می شود [۱]. به وجود آمدن مرحله جدید در اینترنت، امکان استفاده از دانش و داده های مردم عادی، که در برخی موارد تنها منبع داده می باشند را فراهم کرد. این داده های تولید شده توسط عموم، مفهومی را ایجاد کرد که آن را داده های مردم گستر^۱ می نامند [۲]. وب گاه ویکیپدیا از مشهورترین محیط های مردم گستر است که مقالات موجود در آن، توسط عموم تولید و تصحیح می شوند.

پیشرفت های ایجاد شده در زمینه تلفن های همراه، دوربین های رقومی و دستگاه هایی که به سیستم تعیین موقعیت جهانی^۲ مجهز هستند، امکان مشارکت شهروندان در جمع آوری داده های مکانی را فراهم کرده و باعث ایجاد داده های مکانی مردم گستر شد [۳]. تهیه نقشه و تولید اطلاعات مکانی توسط مردم عادی بدون نیاز به آموزش آکادمیک و با استفاده از دانش محلی خود و به اشتراک گذاری آن با سایر افراد، محیطی را فراهم کرد که از آن با عناوینی چون جغرافیای نوین^۳ [۴]، مشارکت مردمی در

^۱ User Generated Content (UGC)

^۲ Global Positioning System (GPS)

^۳ Neogeographic

^۴ Public Participation GIS

^۵ Volunteered Geographic Information (VGI)

به طور کلی، فعالیت‌هایی که سازمان‌های رسمی نقشه‌برداری برای کنترل کیفیت داده‌ها انجام می‌دهند را می‌توان به دو قسمت تقسیم کرد: یک قسمت، فرایندی است که برای کنترل کیفیت در حین تهیه داده‌های مکانی انجام می‌شود؛ و دیگری مربوط به تعیین کیفیت بعد از تهیه داده‌ها می‌باشد که با نمونه‌گیری از داده‌ها و مقایسه آنها با داده‌های مرجع به دست می‌آید و نتایج آن معمولاً به صورت فراداده مستندسازی شده و به همراه داده نگهداری می‌شود [۱۶]. در مورد داده‌های مکانی مردم گستر، فعالیت‌هایی که برای کیفیت داده‌ها صورت می‌گیرد، مشابه روند فوق است که در ادامه به تشریح آنها می‌پردازیم.

۲-۱- تضمین کیفیت

Goodchild فعالیت‌های همزمان با فرایند تولید داده، که با عنوان تضمین کیفیت^۶ از آنها نام می‌برد، را در سه دیدگاه کلی تقسیم کرده است: دیدگاه مردم گستر^۷، دیدگاه اجتماعی^۸ و دیدگاه جغرافیایی^۹ [۱۶].

• دیدگاه مردم گستر

این دیدگاه به توانایی یک گروه از مردم برای معتبر کردن و رفع خطاهای تولید شده توسط یک فرد اشاره دارد [۱۶]. به عنوان مثال، در مورد آتش‌سوزی جنگل، وقتی یک گزارش از آتش‌سوزی دریافت شود نمی‌توان به طور کامل به آن اعتماد کرد، اما اگر تعداد زیادی از افراد از آتش‌سوزی در همان منطقه یا نزدیکی آن گزارش دهند، خبر قابل اعتمادتر می‌شود. به علاوه، داده‌هایی که توسط گروهی از افراد تهیه می‌شود به داده صحیح‌گرایش دارد. اساس آماری این ادعا آن است که هر چقدر اندازه یک نمونه بزرگ‌تر باشد، میانگین آن به به داده واقعی نزدیکتر می‌شود [۹]. در مورد داده‌های مکانی، از آنجا که غالباً مقادیر واقعی در دسترس نیستند، بنابراین مقایسه بین اجماع یک گروه و نمونه در بسیاری از موارد به عنوان داده صحیح فرض می‌شود [۸].

تهیه می‌شود، مسلماً از لحاظ کیفیت، به خوبی داده‌ای که به صورت حرفه‌ای تهیه می‌شود نیست. با گسترش حجم اطلاعات مردم گستر، مسئله ارزیابی کیفیت این داده‌ها و ممانعت از خرابکاری‌ها نیز اهمیت بیشتری پیدا می‌کند [۱۳].

داده‌های مکانی مردم گستر، توسط افراد عادی تهیه می‌شوند و از آنجایی که دید مردم عادی نسبت به جغرافیا و مکان، یک دید مبهم است داده‌های مکانی مردم گستر همواره با ابهام همراه هستند. به همین دلیل گفته می‌شود که داده‌های مکانی مردم گستر، بیشتر بر پایه ادراک هستند تا اندازه‌گیری [۱۴]. برای تعیین کیفیت داده‌هایی که مبهم بوده و برای پایه ادراک می‌باشند، استفاده از مفاهیم کیفیت که در مورد داده‌های رسمی به کار برده می‌شود کارایی کاملی نخواهند داشت. در این راستا، Longueville معتقد است که مفهوم درجه درستی^۱ برای این نوع داده‌ها و بیان این مسئله که داده به چه اندازه و تا چه درجه‌ای دارای یک خصوصیت می‌باشد، به جای بیان آن به صورت صفر و یک مناسب‌تر است [۱۴]. همچنین Metzger و Flagin در نظر گرفتن مفهوم اعتبار^۲ به جای دقت را برای محیط‌های مردم گستر مناسب‌تر می‌دانند؛ زیرا دقت، یک ابزار عینی و اعتبار، یک ابزار غیرعینی است؛ و از این رو، استفاده از مفهوم اعتبار برای محیط‌های مردم گستر که دارای داده‌های کیفی و ذهنی می‌باشند مناسب‌تر است [۱۵]. برای اعتبار، تعریف مشخصی وجود ندارد، اما می‌توان گفت که مفهوم اعتبار، ترکیبی است که حول محور باور کردن^۳ و قبول کردن اطلاعات بر پایه اعتماد^۴ و مهارت^۵ می‌چرخد که می‌تواند به عنوان معیاری برای کیفیت در داده‌های مکانی مردم گستر در نظر گرفته شود. نمونه‌ای از این رویکرد، رده-بندی کاربران و سپس اجازه فعالیت به آنها متناسب با درجه اعتبار و رده‌بندی آنهاست [۱۵]. به عنوان مثال، برای کاربران Wikimapia، رده‌بندی‌هایی بر اساس امتیازات تجربه، میزان مشارکت پیشین، کیفیت کار، وسعت کار و تأیید مدیر سیستم تعریف شده‌است.

۱ Degree of truth

۲ Credibility

۳ Believability

۴ Trustworthiness

۵ Expertise

۶ Quality assurance

۷ Crowd-sourcing approach

۸ Social approach

۹ Geographic approach

• دیدگاه جغرافیایی

در این دیدگاه، مجموعه‌ای از قوانین تعیین می‌شود که مشخص کنند چه داده‌ای کجا می‌تواند قرار بگیرد یا نگیرد. مهمترین اصلی که در مورد داده‌های جغرافیایی وجود دارد این است که داده‌های نزدیک به هم باید با هم سازگار باشند. همچنین اطلاعاتی که مربوط به یک محل می‌باشند نیز با یکدیگر سازگاری داشته و به هم مرتبط باشند [۱۶]. به عنوان مثال، یک کافی شاپ نمی‌تواند در وسط یک اتوبان قرار داشته باشد؛ و یا اگر یک خیابان و رودخانه همدیگر را قطع می‌کنند، باید یک پل در این محل وجود داشته باشد و یا رودخانه از یک کانال زیرزمینی عبور کرده باشد. وب‌گاه OSM ابزارهایی را برای اعمال کنترل‌های اینچنینی بر روی داده‌های این وب‌گاه فراهم کرده است. یکی از ابزارهایی که به صورت اتوماتیک برخی از خطاهای موجود در OSM را کشف می‌کند، وب‌گاه keepright.at می‌باشد. در این وب‌گاه، به عنوان مثال فضاهایی که بسته نشده‌اند، اتوبان‌هایی که اسم ندارند، تونل‌ها و پل‌هایی که مسیری که بر روی آن قرار دارند مشخص نیست و برخی برجسب‌های نادرست که به عوارض داده شده شناسایی شده و با رنگ‌ها و مؤلفه‌های مختلف بر روی نقشه مشخص می‌شوند تا کاربران آنها را تصحیح کنند.

۲-۲- تعیین کیفیت

پس از تهیه داده‌ها، می‌توان کیفیت آنها را تعیین نمود. روش‌های موجود برای تعیین کیفیت داده‌های مکانی مردم گستر را می‌توان به دو دسته "استفاده از فراداده" و "مقایسه با داده‌های زمینی صحیح" تقسیم‌بندی کرد.

• استفاده از فراداده

اولین مکانیسم برای تبادل اطلاعات در مورد کیفیت داده‌ها، استفاده از فراداده می‌باشد [۹]. با وجود سودمند بودن فراداده، هنوز برای بسیاری از پایگاه‌های داده مردم گستر فراداده در دسترس نیست، چون کاربران مجبور به ثبت کردن فراداده نبوده و یا حتی نسبت به مفهوم آن آگاهی ندارند [۱].

با توجه به دیدگاه فوق، وب‌گاه‌هایی مانند OSM و Wikimapia به کاربران خود اجازه تصحیح داده‌ها را می‌دهند. وب‌گاه OSM ابزارهای مختلفی در اختیار کاربران قرار داده تا با استفاده از آنها بتوانند داده‌های موجود در این وب‌گاه را تصحیح کنند که از آن جمله می‌توان به نرم‌افزار رومیزی JOSM و نرم‌افزار آنلاین Potlatch اشاره نمود. به علاوه وب‌گاه OpenStreetBugs (OSB) را برای نمایش خطاهای موجود در نقشه‌های OSM ایجاد کرده است، به طوری که اگر کاربران در یک منطقه، مشکل یا خطایی پیدا کردند که قادر به تصحیح آن نبوده و یا وقت کافی برای تصحیح آن نداشته باشند، آن خطا را بر روی نقشه موجود در این وب‌گاه مشخص می‌کنند تا توسط سایر کاربران تصحیح شود.

با وجود آنکه با استفاده از این رویکرد، بسیاری از خطاهای موجود کشف و تصحیح می‌شوند، اما این دیدگاه فقط در مورد مکان‌هایی که پر جمعیت هستند یا برای مردم جذابیت دارند قابل استفاده است؛ و خطاهای سایر مناطق، به علت نبودن مشاهده کننده و ناظر کافی تصحیح نشده و باقی می‌مانند [۱۶].

• دیدگاه اجتماعی

این دیدگاه، اجتماعی نامیده شده چون اساس آن بر پایه کشف خطاها توسط یک گروه خاص از کاربران که کاربران پیشرفته یا نگهبان‌ها^۱ نامیده می‌شوند قرار دارد. از این رو، این دیدگاه شبیه به چک کردن داده‌ها با استفاده از استانداردها در تهیه داده به صورت رسمی می‌باشد [۱۶].

در وب‌گاه‌های مردم گستر، برخی از کاربران خرابکاری می‌کنند، یعنی برخی از داده‌ها را بدون دلیل حذف کرده یا تغییر می‌دهند. در وب‌گاه Wikimapia، کاربران سطح بالا می‌توانند به فردی که این خرابکاری‌ها را انجام می‌دهد پیغام هشدار دهند و اگر همچنان به فعالیت خود ادامه داد، وی را محدود^۲ کنند. در وب‌گاه OSM هم یک دسته کاربر تحت عنوان کاربران ماهر^۳ وجود دارند که کار حل اختلافات و خرابکاری‌ها را بر عهده دارند [۱۶].

^۱ Gate-keeper

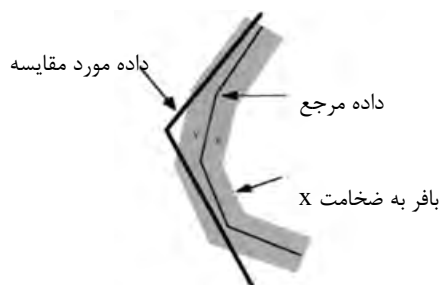
^۲ Banned

^۳ Data Working Group (DWG)

بسیاری موارد، امکان دسترسی به داده‌های زمینی صحیح وجود ندارد [۱۷].

در بیشتر فعالیت‌های انجام شده برای تعیین صحت مکانی و کامل بودن داده‌ها، از روش مقایسه با داده‌های زمینی صحیح استفاده شده است. در سال ۲۰۰۹، Kounadi به ارزیابی کیفیت داده‌های OSM در مقایسه با داده‌های HMGS (یک سرویس کارتوگرافی رسمی در یونان) پرداخت و همپوشانی ۸۰ درصد بین داده‌ها را نتیجه گرفت، اما کامل بودن اسامی و صحت نوع آنها در OSM فقط ۳۰ درصد بوده است [۱۸]. در سال ۲۰۱۰، Haklay صحت مکانی و کامل بودن داده‌های OSM را با داده‌های Ordnance Survey مقایسه نمود. نتایج این بررسی نشان داد که داده‌های OSM به طور نسبی صحیح بوده و با خطای در حدود ۶ متر با مکانهای Ordnance Survey قرار دارند. همچنین در حدود ۲۴ درصد از داده‌های OSM فاقد اطلاعات توصیفی بودند.

در هر دو مثال فوق، برای ارزیابی صحت مکانی از روش Goodchild و Hunter که در سال ۱۹۹۷ ارائه شده، استفاده گردید [۱۹]. در این روش، به دو مجموعه داده نیاز است: یکی به عنوان داده صحیح که در شکل ۱ با خط باریک‌تر نشان داده شده و دیگری داده مورد مقایسه که با خط ضخیم‌تر ترسیم شده است. سپس یک بافر به ضخامت x در اطراف داده صحیح ترسیم شده و مقداری از داده مورد مقایسه که داخل این بافر قرار می‌گیرد محاسبه می‌شود. میزان صحت نتیجه، بستگی به مقدار x دارد.



شکل ۱- مثالی از روش بافر که در آن یک بافر به ضخامت x در اطراف داده مرجع ترسیم می‌شود و درصد شی مورد مقایسه واقع در این محدوده محاسبه می‌گردد [۱۹].

۲-۳- ارائه کیفیت به کاربران

برای اینکه داده‌هایی که توسط مردم تهیه می‌شود قابل استفاده در کاربردهای مختلف باشد، نیاز به پاکسازی

در دهه ۸۰ استانداردهایی برای تهیه فراداده تعریف شد که با گذشت بیش از ۲۰ سال از تعیین آن‌ها، دیگر بخوبی پاسخگوی نیازهای وب مکانی نیستند. یکی از مسائلی که برای وب مکانی اهمیت دارد این است که علاوه بر دانستن دقت یک مجموعه داده به صورت یک و تنها، دقت آن را به صورت نسبی با سایر مجموعه داده‌ها بدانیم. به عنوان مثال، استفاده از یک تصویر Google Earth که به صورت اشتباه زمین مرجع شده است در کاربردی که همه تصاویر موجود در آن با همان تصویر، زمین مرجع شده‌اند قابل قبول است؛ اما اگر هر کدام از تصاویر به صورت مستقل زمین مرجع شده باشند، استفاده از آن تصویر قابل قبول نیست. به همین دلیل Goodchild مفهوم فراداده دوتایی^۱ را مطرح می‌کند که میزان تناسب دو مجموعه داده را برای در کنار هم بودن توصیف می‌کند. این فراداده به مرکزیت کاربر تعریف می‌شود و قابل تطبیق با ماهیت ارزیابی کیفیت داده‌های مردم گستر می‌باشد [۹].

برای تعیین کیفیت با استفاده از فراداده، Ciepluch بیان می‌کند که با مشخص شدن سه عامل می‌توان کیفیت داده‌هایی که توسط کاربران در OSM تهیه می‌شوند را مشخص کرد: (۱) میزان مهارت و تجربه آن‌ها در زمینه GIS؛ (۲) میزان مشارکت قبلی؛ و (۳) میزان استفاده کاربر از اصولی که برای تهیه فراداده وضع شده‌اند [۱۷]. در محیط مردم گستر که Longueville برای جمع‌آوری اطلاعات در مورد یک محیط طبیعی در سوئد فراهم کرده بود، برای برطرف کردن ابهام‌های ایجاد شده از جمع‌آوری دو نوع فراداده استفاده کرد: (۱) ارزیابی کاربران از داده‌های خود که بر این اساس عددی از ۰ تا ۵ را به کار خود اختصاص می‌دادند؛ (۲) فراداده‌ای که به صورت اتوماتیک توسط محیط مردم گستر محیطی^۲ ذخیره می‌شد [۱۴].

• مقایسه با داده‌های زمینی صحیح

یک روش برای تعیین کیفیت داده مکانی تهیه شده، مقایسه آن با داده زمینی صحیح است. مشکل اساسی در این روش، انتخاب مجموعه داده‌ها است به طوری که داده‌های تهیه شده با داده‌های مناسب مقایسه شوند [۱۰]. به علاوه، در مورد داده‌های مکانی مردم گستر، در

^۱ Binary metadata

^۲ Environmental VGI (eVGI)

دارند. پاکسازی داده شامل ادغام کردن، حذف کردن و عملیات تصحیح بر روی عوارض جغرافیایی برای تبدیل آن‌ها به یک فرم یکپارچه و موثر می‌باشد. عملیات حذف و ترکیب می‌تواند به صورت اتوماتیک انجام شود، اما برای عملیات تصحیح به نظر افراد نیاز است [۲۰]. همانطور که گفته شد وب‌گاههایی مانند OSM و Wikimapia امکان تصحیح داده‌ها را برای کاربران فراهم کرده اند.

کاربران برای استفاده از داده‌های مکانی باید از کیفیت آنها آگاه باشند تا بتوانند میزان مناسب بودن آنها برای کاربرد مورد نظرشان را بررسی نموده و از بین داده‌های موجود، مناسب‌ترین داده را انتخاب کنند [۲۱]. در محیط‌های مکانی مردم گستر، به علت مشارکت تعداد زیادی از داوطلبان، داده‌های مکانی زیادی بدست می‌آید و در نتیجه از یک منطقه، داده‌های مکانی با خصوصیات و ویژگی‌های مختلف ایجاد می‌گردد. اما در این وب‌گاه‌ها، به علت عدم آگاهی مردم از کیفیت داده‌های مکانی، خود وب‌گاه‌ها تعیین کننده داده نهایی جهت نمایش و ارائه به کاربران هستند. در بسیاری موارد، آخرین داده به عنوان بهترین داده موجود به کاربر نمایش داده می‌شود. اساس این امر بر این اصل استوار است که آخرین داده، به روزتر است و به علت نظارت تعداد زیادی از افراد، این داده‌ها روز به روز در حال بهتر شدن می‌باشند. اما همان طور که پیشتر بیان شد، کیفیت داده مکانی پارامترهای مختلفی دارد که در کاربردهای مختلف، برخی از پارامترها از سایر پارامترها اهمیت بیشتری خواهند داشت. به عنوان مثال، ممکن است در یک کاربرد، داده با دقت مکانی بالا اهمیت بیشتری داشته باشد، حال آنکه در کاربرد دیگر، میزان کامل بودن، مهمتر باشد. بنابراین امکان انتخاب به کاربران بر اساس معیارهایشان، بهتر از تصمیم‌گیری به جای آن‌ها خواهد بود.

برای ارائه کیفیت داده‌ها به کاربران، تهیه کنندگان داده‌های مکانی، معمولاً فراداده‌هایی را تهیه و به فرمت‌های استاندارد به کاربران ارائه می‌کنند و در آن جنبه‌های مختلف داده را توصیف می‌نمایند تا به کاربران در انتخاب داده مناسب کمک کنند. اما در مورد داده‌های مردم گستر، ممکن است فراداده‌ای در دسترس نباشد. از سوی دیگر، فراداده به صورت مستقل از داده ذخیره می‌شود، به همین دلیل اگر داده دچار تغییر شود این تغییرات ممکن است به فراداده اعمال نگردد؛ به عبارت دیگر، فراداده‌ها

حالت استاتیک دارند و برای فعالیت‌های دینامیک مناسب نمی‌باشند. به علاوه فراداده‌ها معمولاً توصیفات تکنیکی هستند که برای افراد حرفه‌ای و کارشناس تهیه شده‌اند و استفاده از کلمات و اصطلاحات فنی به کار برده شده در آن‌ها، برای افراد عادی کار دشواری است. در نتیجه، فراداده‌های مکانی که با عبارات تخصصی توصیف شده‌اند، اغلب توسط کاربران عادی مورد استفاده قرار نمی‌گیرند [۲۱].

۳- رویکرد پیشنهادی

با توجه به موارد گفته شده در بخش‌های قبل، در این بخش به بیان رویکرد پیشنهادی این مقاله برای مدیریت کیفیت در داده‌های مکانی مردم گستر (شامل تعیین کیفیت و ارائه آن به کاربر) می‌پردازیم.

در رویکرد پیشنهادی، اولاً هیچ داده‌ای حذف نمی‌شود، بلکه پارامترهای مختلف کیفیت آن‌ها تعیین و ثبت می‌گردد. سپس این پارامترهای کیفیت، با استفاده از مؤلفه‌های بصری مانند تغییر رنگ و یا ضخامت خطوط به کاربران نمایش داده می‌شود. به عنوان مثال، هرچقدر یک داده کاملتر باشد، خطوط آن پررنگ‌تر یا ضخیم‌تر نمایش داده می‌گردد. در نهایت، کاربران از بین داده‌های موجود، و براساس پارامترهای کیفیت مورد نظر داده مناسب برای کاربرد خود را انتخاب خواهند کرد. در ادامه به تشریح نحوه تعیین کیفیت و ارائه آن به کاربر در رویکرد پیشنهادی می‌پردازیم.

۳-۱- تعیین کیفیت داده‌ها

با توجه به امکان عدم دسترسی به داده زمینی صحیح در اغلب داده‌های مکانی مردم گستر از یک سو، و وجود نسخه‌های مختلف از یک داده مشابه که توسط کاربران مختلف جمع‌آوری شده از سوی دیگر، پیشنهاد این مقاله برای تعیین کیفیت داده‌ها، مقایسه آنها با یکدیگر می‌باشد. در این بخش، نحوه بکارگیری این رویکرد برای دو پارامتر دقت مکانی و کامل بودن شرح داده می‌شود:

• کامل بودن داده‌ها

برای تعیین میزان کامل بودن داده‌ها به شرح زیر عمل می‌شود:

۳-۲- ارائه کیفیت به کاربران و حق انتخاب به

آنها

در محیط‌های مکانی مردم گستر، هر یک از داده‌ها می‌توانند یک ویژگی مثبت داشته باشند که آنها را برای یک کاربرد خاص مناسب‌تر می‌کند. مثلاً ممکن است یک کاربر، داده‌ای کاملتر با دقت مکانی پایین‌تر و کاربر دیگر، داده‌ای ناقص‌تر اما با دقت مکانی بالاتر تهیه کرده باشد. همانطور که پیشتر شرح داده شد، در محیط‌های مکانی مردم گستر امکان انتخاب از بین این داده‌ها بر اساس خصوصیات آنها به صورتی که بتوان از بین آنها داده کامل‌تر، به روزتر یا با دقت مکانی بالاتر را انتخاب کرد وجود نداشته و عموماً مدیران این محیط‌ها، یک داده را به عنوان بهترین داده ارائه می‌کنند.

پیشنهاد این مقاله، ارائه حق انتخاب به کاربران برای انتخاب داده مناسب از میان همه داده‌های موجود می‌باشد. این امر، با استفاده از مؤلفه‌های بصری قابل درک برای کاربر عادی به منظور ارائه پارامترهای کیفیت محقق می‌گردد که از آن جمله می‌توان به موارد زیر اشاره نمود:

- رنگ: مثلاً داده‌های با دقت بالا با رنگ سبز، داده‌های با دقت متوسط با رنگ زرد و داده‌های با کیفیت پایین با رنگ قرمز نمایش داده شوند.
- طیف رنگی: داده‌ها از کیفیت زیاد به کم، به صورت پرنرنگ به کم‌رنگ نمایش داده شوند.
- ضخامت خطوط: ضخامت خط مورد استفاده برای نمایش عوارض، بر اساس میزان کیفیت آنها تغییر نماید.
- ترکیب مؤلفه‌های فوق: به عنوان مثال نمایش داده با کیفیت‌تر، به صورت پرنرنگ‌تر و ضخیم‌تر.

۴- پیاده‌سازی

رویکرد پیشنهادی، در بخشی از محوطه دانشگاه تحصیلات تکمیلی کرمان به عنوان منطقه مورد مطالعه پیاده‌سازی گردید (شکل ۲). برای این منظور، ۱۶ نقشه به روشهای مختلف توسط افراد متفاوت از این محدوده تهیه شد (شکل ۳):

۱. محاسبه اجتماع کلیه داده‌های موجود که نتیجه آن به عنوان "داده کامل" منظور می‌گردد.
 ۲. تعیین برآوردی از حجم داده‌های موجود در هر داده و همچنین داده کامل به دست آمده در مرحله ۱. این برآورد برای داده‌های نقطه‌ای می‌تواند تعداد نقاط و برای داده‌های خطی، مجموع طول خطوط موجود باشد.
 ۳. تعیین میزان کامل بودن هر داده از طریق مقایسه برآوردهای به دست آمده در مرحله ۲ برای آن داده و داده کامل.
- عدد فوق، همواره بین صفر و ۱ قرار دارد. هر چقدر داده کامل‌تر باشد این عدد به یک نزدیک‌تر خواهد بود.

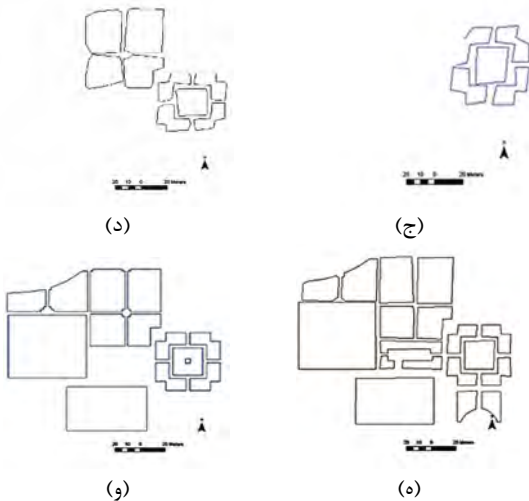
• دقت مکانی

برای تعیین دقت مکانی داده‌ها، روش مقایسه بین داده‌های موجود به صورت زیر پیشنهاد می‌شود:

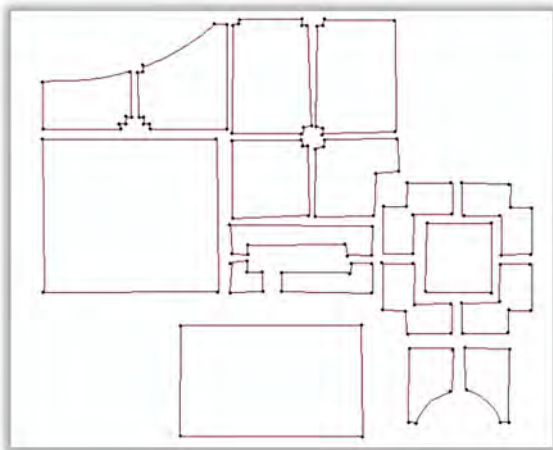
۱. درخواست از تهیه کننده داده برای ارائه اطلاعاتی در خصوص روش جمع آوری داده، دقت وسیله مورد استفاده، و همچنین اطلاعاتی مربوط به سن، جنسیت و بومی بودن یا نبودن
۲. تعیین دقت اولیه برای هر داده بر اساس اطلاعات ارائه شده توسط کاربران. به عنوان مثال، برای داده‌ای که به روش زمینی برداشت شده نسبت به داده‌ای که با GPS تهیه گردیده، دقت بیشتری منظور نمود.
۳. تعیین عوارض مشترک بین داده‌های مختلف
۴. محاسبه میانگین وزن دار برای مختصات عوارض مشترک و تعیین انحراف معیار. وزن هر عارضه در هر نقشه، متناسب با دقت آن نقشه می‌باشد.
۵. حذف عوارضی که فاصله آنها تا میانگین از یک حد آستانه بیشتر باشد؛ سپس تکرار مراحل ۴ و ۵ تا زمانی که هیچ عارضه‌ای حذف نشود.
۶. منظور کردن میانگین نهایی به عنوان مختصات نقاط عوارض
۷. تعیین دقت مکانی هر نقشه بر اساس اختلاف مختصات عوارض موجود در آن با مختصات‌های به دست آمده در مرحله ۶.



شکل ۲- تصویر Google Earth از منطقه مورد مطالعه



شکل ۳- نمونه‌ای از نقشه‌های تهیه شده به روش‌های مختلف: (الف) قدم اتالونه؛ (ب) مترکشی؛ (ج) مارک کردن؛ (د) ردیابی؛ (ه) رقومی-ساز؛ (و) نقشه‌برداری زمینی



شکل ۴- کاملترین نقشه بدست آمده از اجتماع نقشه‌های تهیه شده از منطقه. نقاط مشخص شده برای تعیین تعداد کل رؤوس، شمارش شدند

• دقت مکانی

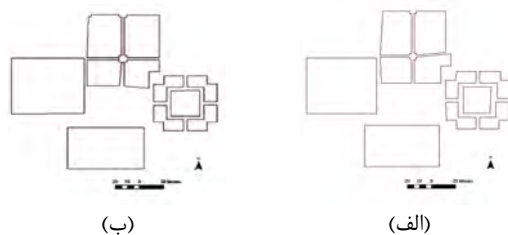
۱. برای تعیین دقت مکانی هر نقشه، مراحل زیر انجام شد:
 - تعیین دقت اولیه هر کدام از نقشه‌ها بر اساس روش تهیه نقشه (جدول ۱). برای محاسبه مقدار خطای هر روش، یک طول معین، ۵۰ بار به آن روش اندازه-گیری شده و انحراف معیار نتایج به عنوان میزان خطای آن روش در نظر گرفته شد.
 ۲. شناسایی نقاط مشترک کلیه نقشه‌ها. در مجموع ۲۳ نقطه مشترک شناسایی شد.
 ۳. تعیین اندازه‌گیری‌های اشتباه با میانگین‌گیری وزندار شرح داده شده در بخش ۳-۱. با فرض توزیع نرمال

- ۱) سه نقشه با اندازه‌گیری طول‌ها و فواصل با قدم اتالونه
 - ۲) دو نقشه با اندازه‌گیری طول‌ها و فواصل با مترکشی
 - ۳) دو نقشه به وسیله GPS به روش مارک کردن رؤوس عوارض
 - ۴) چهار نقشه به وسیله GPS به روش ردیابی (tracking) عوارض
 - ۵) سه نقشه از طریق رقومی کردن عوارض بر روی تصویر ماهواره‌ای Google Earth
 - ۶) دو نقشه به روش نقشه‌برداری زمینی
- همچنین از افراد خواسته شد تا تنها بخشی از عوارض را برداشت کنند تا داده‌هایی با درجات کامل بودن متفاوت در اختیار داشته باشیم.

۴-۱- تعیین کیفیت داده‌ها

• میزان کامل بودن

برای تعیین میزان کامل بودن داده‌های موجود در هر نقشه، ابتدا اجتماع عوارضی که در همه نقشه‌ها وجود داشت در یک نقشه به عنوان نقشه کامل ترسیم شد (شکل ۴) و تعداد رؤوس موجود در این نقشه کامل و همچنین تعداد رؤوس موجود در هر کدام از نقشه‌ها شمارش گردید. با تقسیم تعداد رؤوس هر نقشه بر تعداد کل رؤوس بدست آمده، میزان کامل بودن هر نقشه قابل محاسبه است.

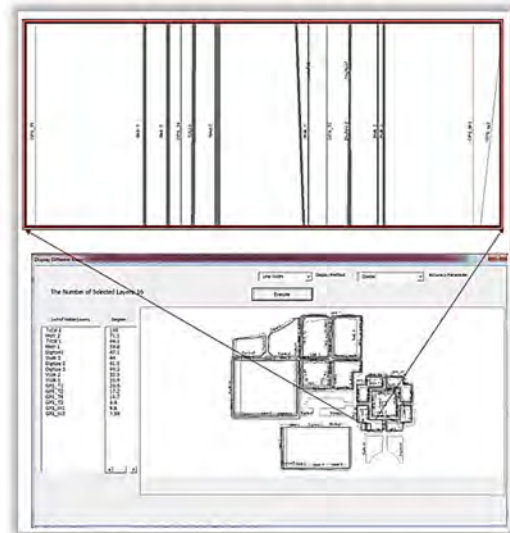


که حداقل دقت مکانی آنها ۵۰ سانتیمتر و میزان کامل بودن آنها بیشتر از ۶۰ درصد است، این اعداد را به سیستم معرفی نموده و نقشه‌هایی که این شروط حداقلی را دارند تعیین می‌شوند. در نهایت، می‌توان برای مقایسه نقشه‌های دارای شرایط مورد نظر، پارامترهای کیفیت آنها را با مؤلفه‌های بصری مختلف نمایش داد.

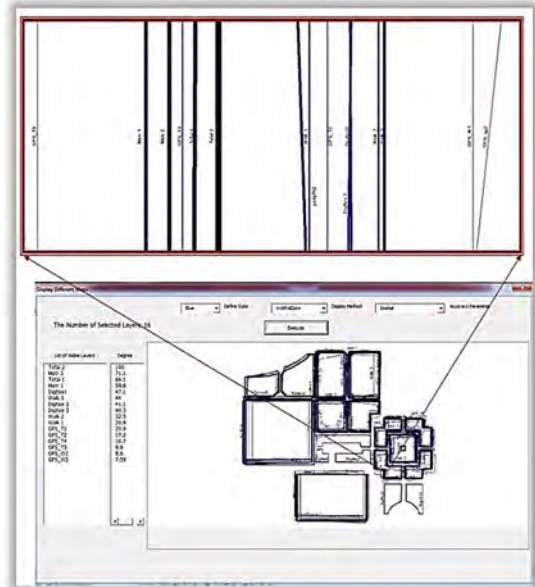
۵- نتیجه گیری

در این مقاله به بررسی مدیریت کیفیت در داده‌های مکانی مردم گستر پرداخته شد. با وجود سودمند بودن داده‌های مکانی تولید شده توسط مردم و مزایایی که نسبت به داده‌های رسمی دارند (از جمله جامع بودن، سرعت تهیه بالا و کم هزینه بودن)، اما سوالات بسیاری در مورد کیفیت این داده‌ها مطرح بوده و تحقیقات بسیاری در این زمینه انجام شده است.

رویکرد پیشنهادی این تحقیق برای تعیین کیفیت در سیستم‌های مکانی مردم گستر، مقایسه داده‌های موجود با یکدیگر است. همچنین، برای ارائه پارامترهای کیفیت به کاربران از مؤلفه‌های بصری قابل درک برای کاربران عادی استفاده گردید. پیاده‌سازی این رویکرد برای یک منطقه مورد مطالعه و ارائه نتایج آن در یک محیط برنامه‌نویسی شده، بیانگر کارایی این روش در برآورده نمودن هر چه بیشتر نیازهای کاربران برای انتخاب داده مناسب می‌باشد. این مقاله، نحوه استفاده از رویکرد پیشنهادی را صرفاً برای پارامترهای دقت مکانی و کامل بودن داده‌ها بررسی نموده است. تعیین کیفیت سایر پارامترهای کیفیت مکانی داده‌ها بدون نیاز به مقایسه با داده‌های صحیح زمینی می‌تواند موضوع تحقیقات بعدی باشد. به علاوه، تاثیر برخی پارامترها بر کیفیت داده‌های مکانی مردم گستر نظیر میزان تحصیلات، سن، جنسیت و حتی شغل فرد تهیه کننده در تعیین پارامترهای کیفیت، می‌تواند به عنوان گام بعدی این تحقیق مد نظر قرار گیرد.



شکل ۶- نمایش میزان دقت مکانی نقشه‌ها با استفاده از تغییر ضخامت خط



شکل ۷- نمایش میزان دقت مکانی نقشه‌ها با استفاده از تغییر درجه روشنی رنگ آبی و ضخامت خط

۴-۲-۲- انتخاب بین نقشه‌های موجود بر اساس محدوده کیفیت مشخص شده

در این نوع انتخاب، داده‌هایی که پارامترهای کیفیت آنها دارای شرایط خاص تعیین شده توسط کاربر هستند مشخص می‌گردند. به عنوان مثال، برای تعیین داده‌هایی

- [1] Cooper, A., S. Coetzee, and D. Kourie, (2012), "Volunteered Geographical Information – The Challenges", *PoPositionIT*. p. 34-38.
- [2] Baeza-Yates, R., (2009), "User Generated Content: How Good Is It?", 3rd Workshop on Information Credibility on the Web, Madrid, Spain.
- [3] Haklay, M., A. Singleton, and C. Parker, (2008), "Web Mapping 2.0: The Neogeography of the GeoWeb", *Geography Compass*, 2, 6. p. 2011-2039.
- [4] Elwood, S., (2008), "Volunteered Geographic Information: Key Questions, Concepts and Methods to Guide Emerging Research and Practice", *GeoJournal*, 72. p. 133-135.
- [5] Bugs, G., et al., (2010), "An Assessment of Public Participation GIS and Web2.0 Technologies in Urban Planning Practice in Canela, Brazil", *Cities* 27, 3. p. 172-181.
- [6] Goodchild, M.F., (2007), "Citizens as Sensors: Web 2.0 and The Volunteering of Geographic Information", *GeoFocus*, 7. p. 8-10.
- [7] Haklay, M., P. Weber, (2008), "OpenStreetMap: User-Generated Street Maps", *Pervasive Computing*, 7, 4. p. 12-18.
- [8] Qian, X.a.L.D., (2009), "Data Cleaning Approaches in Web2.0 VGI Application", *Geoinformatics*, 2009 17th International Conference.
- [9] Goodchild, M.F., (2008), "Spatial Accuracy 2.0 ", 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences.
- [10] Haklay, M., (2010), "How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets", *Environment and Planning B: Planning and Design*, 37, 4. p. 682-703.
- [11] Kounadi, O., (2009), "Assessing The Quality of OpenStreetMap Data", University College of London.
- [12] Ather, A., (2009), " A Quality Analysis of OpenStreetMap Data", University College London.
- [13] Chatterjee, K., L.d. Alfaro, and I. Pye, (2008), "Robust Content-Driven Reputation", 1st ACM workshop on Workshop on AISec, Alexandria, Virginia, USA.
- [14] Longueville, B.D., N. Ostländer, and C. Keskitalo, (2010), "Addressing Vagueness in Volunteered Geographic Information (VGI) – A Case Study", *International Journal of Spatial Data Infrastructures Research*, 5.
- [15] Flanagan, A. and M. Metzger, (2008), "The Credibility of Volunteered Geographic Information", *GeoJournal*, 72, 3. p. 137-148.
- [16] Goodchild, M.F. and L. Li, (2012), "Assuring the Quality of Volunteered Geographic Information", *Spatial Statistics*, 1. p. 110-120.
- [17] Ciepluch, B., P. Mooney, and A.C. Winstanley, (2011), "Building Generic Quality Indicators for OpenStreetMap", 19th annual GIS Research UK (GISRUk).
- [18] Kounadi, O., (2009), "Assessing the Quality of OpenStreetMap data", University College of London.
- [19] Goodchild, M.F. and G.J. Hunter, (1997), "A Simple Positional Accuracy Measure for Linear Features", *International Journal of Geographical Information Science*, 11, 3. p. 299-306.
- [20] Qian, X. and L. Di, (2009), "Data Cleaning Approaches in Web2.0 VGI Application", *Geoinformatics 2009*, 17th International Conference .
- [21] Devillers, R., et al., (2002), "Spatial Data Quality: From Metadata to Quality Indicators and Contextual End-User Manual", *OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management*.

