اولین کنفرانس مهندسی فناوری اطلاعات مکانی
دانشگاه صنعتی خواجه نصیرالدین طوسی
دانشکده مهندسی نقشه‌برداری
۲۹ و ۳۰ دی ماه ۱۳۹۴

The 1st National Conference on
Geospatial Information Technology
K.N.Toosi University of Technology
Faculty of Geomatics Engineering
19 - 20 January 2016

ارزیابی دقت معنایی داده‌های مکانی داوطلبانه
(مطالعه موردی شهر تهران، منطقه ۵)

محمد عشقی[1*]، علی‌اصغر آل‌شیخ[2]

۱- دانشجوی کارشناسی ارشد سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی
۲- **استاد گروه سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی**
(عضو قطب علمی مهندسی فناوری اطلاعات مکانی)

چکیده :

رشد تعداد کاربران و حجم بالای اطلاعات سایت OpenStreetMap نشان از موفقیت پروژه‌های داوطلبانه در جذب قشرهای مختلفی از مردم در سراسر جهان می‌باشد. حجم بالای اطلاعات تولید شده توسط مردم عادی که به صورت روزانه انجام می‌شود چالش‌هایی را در رابطه با اطمینان به کیفیت اینگونه از داده‌ها ایجاد کرده است. پایگاه داده‌های داوطلبانه دارای جنبه‌های مثبتی بوده که از میان آن‌ها می‌توان به "در دسترس بودن اطلاعات" و "توانایی به‌روزرسانی سریع" اشاره کرد؛ با این حال کیفیت این داده‌ها به‌عنوان یک چالش مهم می‌باشد. در این مقاله، از میان عناصر مرتبط با کیفیت داده‌های مکانی داوطلبانه، سعی در بررسی دقت معنایی عوارض خطیِ راه (خیابان-ها) داده‌های داوطلبانه منطقه ۵ شهرداریِ شهر تهران شده است. علت این امر تفاوت در استانداردهایی است که مجموعه داده‌های مختلف از جمله مرجع و داوطلبانه در تعریف نوع و برچسب‌گذاری داده‌های مورد استفاده دارند. در این مقاله، برچسب‌های ارائه شده برای "نوع داده" در مجموعه داده مرجع و داوطلبانه بررسی شده و برچسب‌هایی که دارای بالاترین شباهت باشند به عنوان متناظر شناخته شده و از میزان این شباهت‌ها دقت معنایی عوارض خطیِ راه داده‌های داوطلبانه ارائه و ارزیابی شده است. نتایج نشان می‌دهد که از میان طیف گسترده‌ای از برچسب‌های تعریف شده از قبل برای هرکدام از پایگاه داده‌های مرجع و داوطلبانه، تنها تعدادی از برچسب‌ها در منطقه مطالعاتی متناظر یکدیگر هستند، برای مثال برچسب‌های تندراه و Trunk با ۶۰ عارضه متناظر (درصد تشابه ۸۵/۷ ٪) به عنوان برچسب-های مشابه استخراج شده‌اند.

واژه‌های کلیدی : سیستم اطلاعات مکانی، اطلاعات جغرافیایی داوطلبانه، کیفیت، دقت معنایی.

نویسنده مکاتبه کننده : محمد عشقی
آدرس پستی : تهران، فلکه ۲ صادقیه، بلوار فردوس شرق، خیابان ولیعصر، کوچه ۱۲ متری هفتم، پلاک ۴۴، واحد ۹. کد پستی: ۱۴۸۱۶۱۳۸۴۱
تلفن : ۰۹۱۲۸۴۴۹۶۹۰
آدرس پست الکترونیک : m.eshghi@mail.kntu.ac.ir, m.eshghi70@gmial.com

۱- مقدمه

پیشرفت‌های انجام شده در زمینه مکانی[1] همچون توسعه نگاشتِ وب۲[2] فرصت‌های ارزشمندی برای استفاده از مزایای حجم عظیمی از کاربرانی که به صورت داوطلبانه اقدام به اخذ و غنی‌سازی داده‌های مکانی می‌کنند را در اختیار قرار می‌دهد [۱]. همچنین سخت‌افزارهای مدرنِ موجود و در دسترس کاربران همچون تلفن‌های هوشمند، منجر به ایجاد فرصتی برای مشارکت شهروندان در جمع‌آوری داده‌های مکانی شده و نقش آن‌ها را در استراتژی تولید داده به شدت تغییر داده، پر اهمیت و تاثیر گذار کرده است [۲، ۳]. نتیجه این امر داده‌هایی می‌باشد که توسط عموم مردم تولید شده است و در سال ۲۰۰۷ توسط Goodchild با نام داده‌های داوطلبانه به دنیا معرفی شدند [۳]. پروژه‌های شناخته شده‌ای مانند OSM [3]، Wikimapia، و Google Mapmaker (که براساس این ایده شکل گرفته‌اند) کاربران را قادر می‌سازند تا نقشی در تولید داده‌های مکانی داشته باشند. نرخ بالای افزایش کاربران چنین محیط‌هایی نشانی از سطح محبوبیت این پروژه‌ها می‌باشد [۱]. برای مثال، تعداد عضوهای ثبت شده در پایگاه داده OSM در سال ۲۰۰۹ میلادی، صدهزار کاربر بوده است، در حالیکه تا اواخر سال ۲۰۱۵ میلادی، این تعداد به بیش از سه میلیون کاربر افزایش یافته است [۴]. همچنین Maps-Google، که از منابع مختلف اطلاعاتی برای ایجاد یک پایگاه داده مناسب بمنظور به کار گرفتن آن برای کاربردهای مختلف اطلاعات مکانی، استفاده می‌کند. منابع اطلاعاتی نظیر سیل عظیم اطلاعات تولید شده توسط مردم در شبکه‌های اجتماعی مختلف همچون فلیکرِ[4]، بعلاوه منابع اطلاعاتی مختصِ گوگل ازجمله اطلاعات هواشناسی و سنجش از دور[5] بدست آمده از طریق پایش زمین توسط ماهواره‌ها [۵]. به عنوان نتیجه، جریان داده عظیمی در حوزه اطلاعات مکانی شکل گرفته است که در آن الزام به بهترین استفاده در مناسب‌ترین زمان ممکن از داده‌های جغرافیایی و به‌ویژه مکانی، وجود دارد. برای این منظور، نیاز به بررسی این داده‌ها، پردازش و سازمان‌دهی آن‌ها بمنظور رسیدن به بهترین نتایج، احساس می‌شود. نتایجی که از مهم‌ترین آن‌ها می‌توان به شناسایی اطلاعات دارای دقت و صحت مناسب اشاره کرد [۶].

با توجه به ذات داده‌های داوطلبانه، این داده‌ها دارای مزایا و معایبی هستند. با در نظر گرفتن مزایای این داده‌ها، ولی معایب آن‌ها سدِ جدی بر سر راه استفاده از آن‌ها می‌باشد که پراهمیت‌ترین آنها به بحث کیفیت این داده‌ها برمی‌گردد [۷و۸]. چون کیفیت داده‌های داوطلبانه تولید شده معلوم نیست بررسی کیفیت این نوع از داده‌ها همیشه محل بحث و تحقیق بسیاری از محققان بوده است [۹و۱۰]. بنابراین، در این مقاله نیز تلاش به بررسی بخشی از کیفیت این داده‌ها از جنبه‌ی دقت معنایی‌شان شده است.

۲- پیشنه تحقیق

در این تحقیق، در ابتدا به مدل‌ها و مفاهیم ارائه شده در زمینه تشابه معنایی پرداخته و سپس کاربرد خاص آن را در زمینه اطلاعات مکانی داوطلبانه بررسی کرده و تحقیقات انجام شده در این زمینه بررسی می‌شود.

۲-۱- مدل‌ها و مفاهیم تشابه معنایی

در حالیکه انسان‌ها توانایی تشخیص و تمییز دادن مفاهیم مشابه را دارند، انجام چنین کاری برای ماشین و سیستم‌های کامپیوتری بسیار سخت و دشوار می‌باشد. برای مثال، انسان‌ها هنگامی که به "راه" فکر می‌کنند بسیار طبیعی و محتمل‌تر است که به جای فکر به "رود" و یا "دره‌ای ممتد"، به "خیابان" فکر کنند. ایجاد چنین دانشی در سیستم-های کامپیوتر نیازمند استفاده از مدل‌های محاسباتی ویژه‌ای می‌باشد. یک راه برای ایجاد تمایز بین مفاهیم و اندازه-

[1] Geospatial
[2] Web 2.0 mapping
[3] OpenStreetMap
[4] Flicker
[5] Remote Sensing

گیری روابط بین مفاهیم، استفاده از تشابه معنایی است. تشابه معنایی براساس معانیِ مفاهیم تعیین می‌کند که چه مقدار این مفاهیم مشابه هستند. چندین روش برای اندازه‌گیری و تعیین تشابه معنایی بین اصطلاحات ارائه شده است. روش‌های ارائه شده تاکنون را براساس دانش و مفهوم شباهت استفاده شده در آنها، می‌توان به ۵ دسته کلی تقسیم-بندی کرد [۱۱، ۱۲]: مدل‌های هندسی، خصوصیت، هم‌ترازی، شبکه، و تبدیل.

۲-۲- تشابه معنایی استفاده شده در داده‌های مکانی داوطلبانه

در OpenStreetMap، پدیده‌ها دنیای واقعی بوسیله اختصاص مجموعه‌ای از خصوصیات به شکل‌های هندسی اولیه جغرافیایی توصیف می‌شوند. در حال حاضر، OSM از چهار نوع از شکل‌های هندسی اولیه جغرافیایی (نقطه، خط، پلیگون، و روابط) بهره می‌برد. خصوصیات توصیفی با استفاده از ساختار دوگانه کلید/ارزش[۶] که تگ[۷](برچسب) نامیده می‌شوند، توصیف می‌شوند. در این ساختار، "Key" مشابه خصوصیت توصیفی و "Value" معادل ارزش و مقداری مربوط به خصوصیت توصیفی برای شیء هندسی مربوطه می‌باشد. یک شیء در پایگاه داده OSM می‌تواند توسط هر تعداد از برچسب‌ها توصیف شود، با این شرایط که Key برای هر شیء می‌بایست منحصر بفرد باشد. این برچسب‌ها نتیجه بحث‌های پیوسته و غیررسمی در اجتماع[۸] می‌باشد و محبوب‌ترین برچسب‌ها در ویکیِ[۹] OSM توصیف شده‌اند، با این حال، محدودیتی در استفاده از برچسب‌ها وجود ندارد و مشارکت کنندگان می‌توانند آزادانه و در مواقع ضروری برچسب‌های مورد نظر خود را ایجاد کنند [۱۳]. درحالیکه چنین روش‌هایی که اجتماع-مبنا[۱۰] هستند اجازه ایجاد مفاهیم غنی را می‌دهند، و در عین حال می‌توانند عدم تجانس معنایی قابل توجه‌ای را بواسطه مشارکت کنندگان و وسوسه آنها به استفاده از برچسب‌های نادر (و یا ایجاد برچسب دلخواهی خودشان زمانیکه از وجود برچسبی که به صورت معمول توسط مشارکت کنندگان برای توصیف یک عارضه به کار می‌رود، خبر ندارند) ایجاد کنند. چنین انعطاف‌پذیری همچنین می‌تواند منجر به عدم اجماع در رابطه با برچسب‌های مورد استفاده شود. برای مثال، برخی از کاربران سایت OSM ممکن است یک جنگل را به صورت "landuse=forest" برچسب بزنند، در صورتی که سایرین ممکن است به صورت "natural=wood" این کار را انجام دهند [۱۴]. دو پروژه مهم در زمینه مشکل معنایی داده‌های VGI در ادامه بررسی شده‌اند.

پروژه انجام شده توسط Mülligann و همکاران (۲۰۱۱) تشابهات معنایی را با آنالیز الگوهای نقطه‌ای در یک روش دو مرحله‌ای برای مطالعه جغرافیای معناهای OSM، ترکیب کرد. در ابتدا، تشابه معنایی از مجموعه داده OSM بوسیله هم‌رخدادی مکانی عوارض محاسبه شده و در مرحله بعد، الگوی مکانی-معنایی بوسیله نمایش ارتباط بالقوه بین دو نوع مختلف از عوارض جغرافیایی (یا بین دو نوع یکسان از عوارض جغرافیایی) شناسایی شده است. برای مثال، اضافه کردن یک ایستگاه آتش‌نشانی نزدیک به یک ایستگاه موجود ، می‌تواند بعنوان خطا شناخته شود، به این معنا که میانگین فاصله بین این دو عارضه زیرمجموعه‌ای از الگوی مکانی-معنایی این عارضه می‌باشد [۱۵].

پروژه انجام شده توسط Ballatore و همکاران (۲۰۱۲) از یک شبکه معنایی برای استخراج روابط مفاهیم OSM در ویکی استفاده کرده است. این استخراج با استفاده از ویکی کِراولِر[۱۱] که در OSM توسط نویسندگان توسعه داده شده، انجام شده است. در این شبکه معنایی، مفاهیم OSM بعنوان ورتکس[۱۲] (Vertex)، و لبه‌های بین این مفاهیم (ورتکس‌ها) بوسیله اندازه‌گیری ارتباط بین صفحات ویکی، نمایش داده شده‌اند. با استفاده از این شبکه معنایی،

[6] Key/Value
[7] Tag
[8] Community
[9] Wiki
[10] Community-based
[11] Wiki Crawler
[12] Vertex

اندازه‌گیری‌های متفاوتی از تشابه (برای مثال، SimRank و P-Rank) براساس الگوریتم استنادی آزمایش شده‌اند و یک امتیاز تشابه معنایی بین دومفهوم محاسبه شده است [۱۴].

۳- محاسبه دقت معنایی داده‌های مکانی داوطلبانه

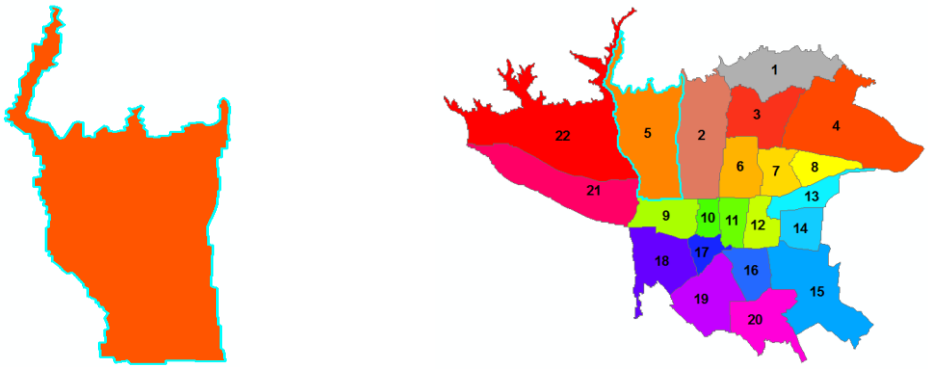
همانطور که Siebritz (۲۰۱۴) در پایان نامه خود بیان کرده است [۱۶]، در این مقاله برای تعریف دقت معنایی یک مجموعه داده در مقایسه با طبقه بندی داده‌های مرجع این مسئله بررسی شده است که آیا عوارض دسته بندی شده در داده‌های داوطلبانه (OSM) به دسته‌های درستی طبقه بندی شده اند یا خیر؟ در مورد کلاس های مربوط به عارضه راه در دو مجموعه داده مرجع و داوطلبانه (OSM)، نتیجه بررسی دقت معنایی داده تعیین می‌کند که آیا یک عارضه مشخص بصورت یکسان در دو مجموعه داده طبقه بندی شده است یا نه (حتی اگر نام های متفاوتی برای هر طبقه در دو دسته داده متفاوت باشد). بنابراین برای انجام این کار، در اولین گام می‌بایست مجموعه داده‌های داوطلبانه‌ای که دارای متناظر در مجموعه داده مرجع می‌باشند را شناسایی کرد. در مرحله بعد، می‌بایست این بررسی انجام شود که کدام یک از کلاس های راه، عارضه راهِ مشابه‌ای بین دو مجموعه داده مرجع و داده‌های داوطلبانه (که دارای متناظر در مجموعه داده مرجع می‌باشند) تعیین می‌کنند. با انجام این کار، در واقع یک فرهنگ لغت ایجاد می‌شود که برچسب‌های استفاده شده در پایگاه داده OSM را قابل ترجمه کرده و امکان تلفیق این پایگاه داده را با پایگاه داده‌های مرجع موجود برای افزایش و غنی‌سازی داده‌ها می‌کند. برای بررسی دقت مجموعه داده های داوطلبانه OSM مقایسه کلاس‌های از قبل تعریف شده برای هر دو دسته داده مرجع و داوطلبانه (OSM) برای ایجاد تطبیق کلاس‌های راه‌ها استفاده شده است.

۴- پیاده‌سازی و ارزیابی

منطقه مورد مطالعه در این تحقیق، منطقه ۵ شهرداریِ شهر تهران می‌باشد. داده‌های شبکه راه‌ها و معابر تهران، تولیدی شهرداری تهران، با مقیاس ۱:۲۰۰۰ به عنوان مجموعه داده مرجع، و داده‌های سایت OSM به عنوان مجموعه داده داوطلبانه مورد استفاده این تحقیق هستند. بدلیل رایگان بودن داده‌های سایت OSM و همچنین امکان دریافت آسان این داده‌ها در قالب شیپ-فایل (Shape-file) و در حالت آماده برای استفاده در سیستم‌های اطلاعات مکانی[۱۳]، از داده‌های این وب‌سایت استفاده شده است. شکل (۱) نشان دهنده منطقه مطالعاتی می‌باشد. با توجه به شکل (۲)، مرحله تناظریابی میان داده‌های مرجع و داوطلبانه بوسیله یک الگوریتم عارضه-مبنا (روشی بهبود یافته براساس روش ارائه شده توسط Koukoletsos، ۲۰۱۲، [۱۷]) انجام شده و داده‌های داوطلبانه متناظر و نامتناظر استخراج شده‌اند. با توجه به جدول (۱)، همانطور که در مقایسه کلاس‌های مجموعه داده مرجع و داوطلبانه دارای متناظر برای منطقه ۵ شهر تهران می‌توان مشاهده کرد: ۱- کلاس‌ها و دسته‌هایی وجود دارد که به مقدار لازم عارضه در آنها وجود ندارد در نتیجه کلاس‌های انتخابی در داده‌های داوطلبانه برای تعیین دقت معنایی عبارتند از: Primary، Secondary، و Trunk. در نتیجه، همانطور که در جدول (۲) نشان داده شده است، ۳ کلاس راه از پایگاه داده داوطلبانهِ OSM به صورت کامل با مجموعه داده مرجع انطباق پیدا کردند؛ که این انطباق با توجه به درصد مشابهتی که میان داده‌های مرجع و داوطلبانه متناظر از نظر مکانی و توصیفی دارند و براساس برچسب‌های ارائه شده برای نشان دادن نوع داده برای هرکدام از آن‌ها، استخراج شده است. دو ستون به ترتیب تعداد تطابقات و درصد مطابقت را نشان می‌دهند و به ترتیب بیان می‌کنند که چند بار و به چه درصدی این انطباق میان این دو مجموعه داده اتفاق افتاده است. برای مثال کلاس Trunk در مجموعه داده داوطلبانهِ OSM دارای متناظر تندراه در مجموعه داده مرجع می‌باشد. این درصد مشابهت برچسب‌ها بیانگر دقت معنایی داده‌های داوطلبانه سایت OSM نیز می‌باشد. لازم به ذکر است که با توجه به

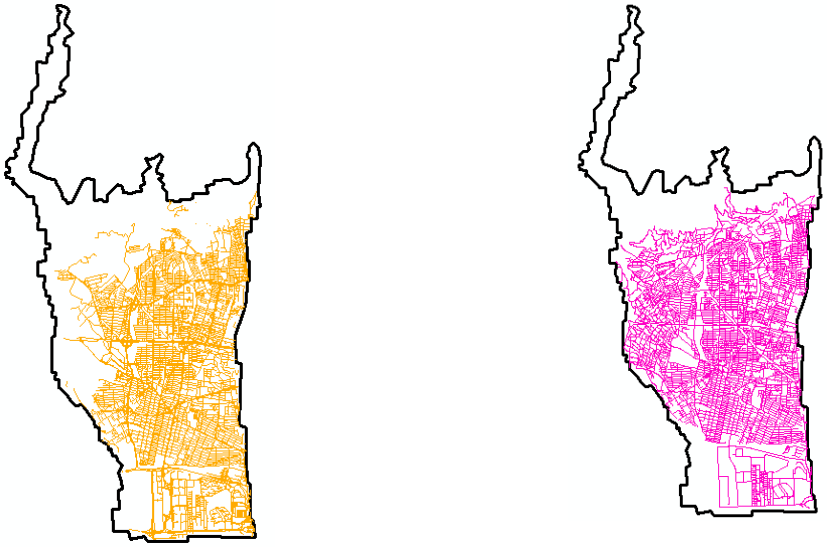
[13] GIS-Ready

نام گذاری‌های متفاوتی که در کشورهای مختلف انجام می‌شود، این دسته بندی و میـزان مشـابهت را نمـی‌تـوان بـرای سایر مکان‌ها در نظر گرفت و کاملا محلی می‌باشد.

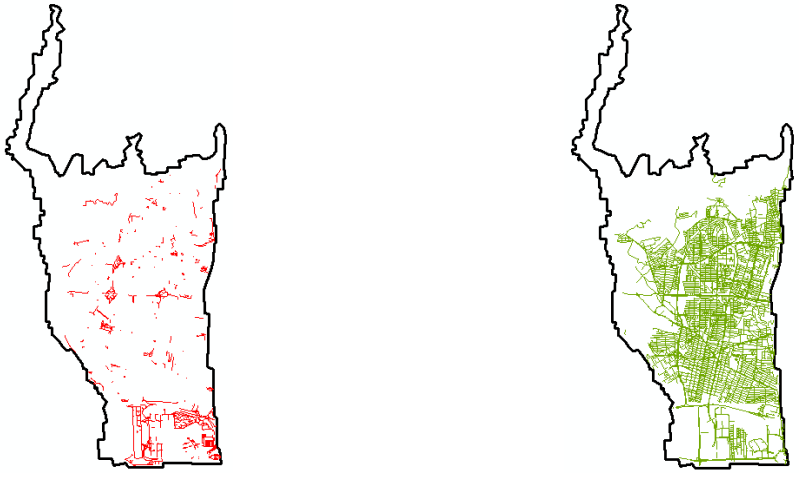


الف ب

شکل ۱: منطقه مطالعاتی: الف) کل مناطق شهردایِ شهر تهران ب) منطقه ۵ شهردایِ شهر تهران



الف ب



ج د

شکل ۲: الف) داده‌های راه برای مجموعه داده مرجع ب) داده‌های راه برای مجموعه داده داوطلبانه سایت OSM ج) داده‌های راه برای مجموعه داوطلبانه سایت OSM که دارای متناظر در مجموعه داده مرجع می‌باشند د) داده‌های راه برای مجموعه داوطلبانه سایت OSM که بدون متناظر در مجموعه داده مرجع می‌باشند

جدول ۱: تعداد کلاس‌های تطابق یافته مجموعه داده مرجع و داوطلبانه برای منطقه ۵ شهرداریِ شهر تهران

طبقه بندی کلاس‌های راه براساس استاندارد ۵۰۴ شهرداری تهران برای داده‌های مرجع																					تعداد عوارض OSM	
آزادراه	تندراه	شریانی درجه ۱ (خیابان)	شریانی درجه ۲ (خیابان)	معابر جمع کننده	معابر محلی دسترسی	معابر اختصاصی	با دسترسی	با محدودیت	معبر ویژه علو پیاده													
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	Footway	طبقه بندی کلاس‌های راهِ OSM برای داده‌های داوطلبانه در منطقه مورد مطالعه	۳۹									
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	living_street		۸۶									
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	Motorway		۲									
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	Motorway-Link		۱۳									
۰	۰	۱۲	۲	۰	۰	۰	۰	۰	۰	Primary		۱۴									
۰	۰	۰	۰	۵	۰	۰	۰	۰	۰	Primary-Link		۲۸									
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	Residential		۱۹۱۷									
۰	۰	۰	۴۸	۰	۰	۰	۰	۰	۰	Secondary		۷۵									
۰	۰	۰	۰	۱۵	۰	۰	۰	۰	۰	Secondary-Link		۳۹									
۰	۰	۰	۰	۱۰	۲۰	۰	۱۱	۳۳	۰	Service		۱۱۰									
۰	۰	۰	۰	۰	۰	۰	۰	۸	۰	Steps		۱۷									
۰	۰	۰	۰	۰	۰	۴۰	۶۳	۰	۰	Tertiary		۱۳۰									
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	Tertiary-Link		۱۵									
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	Track		۴									
۰	۶۰	۹	۶	۰	۰	۰	۰	۰	۰	Trunk		۷۵									
۰	۰	۰	۰	۴۰	۰	۰	۰	۰	۰	Trunk-Link		۱۲۶									
۰	۰	۰	۰	۰	۰	۰	۰	۱۰	۱۷	Unclassified		۷۱									

جدول ۲: اندازه‌گیری تشابه کلاس‌های مجموعه داده مرجع و داوطلبانه برای منطقه ۵ شهرداریِ شهر تهران

مطابقت %	تعداد	مجموعه داده داوطلبانه OSM	مجموعه داده مرجع
٪ ۸۵/۷	۶۰	Trunk	تندراه
٪ ۸۵/۷	۱۲	Primary	شریانی درجه ۱ (خیابان)
٪ ۶۴	۴۸	Secondary	شریانی درجه ۲ (خیابان)

جدول ۳: کلاس‌های متناظر شناخته شده در دو مجموعه داده مرجع و داوطلبانه برای منطقه ۵ شهرداریِ شهر تهران

مجموعه داده داوطلبانه OSM	مجموعه داده مرجع
Trunk	تندراه
Primary	شریانی درجه ۱ (خیابان)
Secondary	شریانی درجه ۲ (خیابان)

۵- نتایج و پیشنهادها

در این مقاله به ارزیابی دقت معنایی داده‌های داوطلبانه با استفاده از مقایسه برچسب‌های استفاده شده برای تعیین نوع این داده‌ها و مقایسه آنها با داده‌های متناظر در مجموعه داده مرجع پرداخته شده است. برچسب‌های ارائه شده برای "نوع داده" در مجموعه داده مرجع و داوطلبانه بررسی شده و برچسب‌هایی که دارای بالاترین شباهت بودند به عنوان متناظر شناخته شده و از میزان این شباهت‌ها دقت معنایی عوارض خطیِ راه داده‌های داوطلبانه ارائه و ارزیابی شد. نتایج نشان داد که از میان طیف گسترده‌ای از برچسب‌های تعریف شده از قبل برای هرکدام از پایگاه داده‌های مرجع و داوطلبانه، تنها تعدادی از برچسب‌ها در منطقه مطالعاتی متناظر یکدیگر هستند، برای مثال برچسب‌های تندراه و Trunk با ۶۰ عارضه متناظر (درصد تشابه ٪ ۸۵/۷) به عنوان برچسب‌های مشابه استخراج شده‌اند. بررسی های انجام شده در رابطه با دقت معنایی داده‌های داوطلبانه با استفاده از این روش، بیان کننده نیاز به ارائه یک مدل مفهومی برای ذخیره‌سازی و بازیابی داده‌های داوطلبانه است، بطوریکه برچسب‌های بکار رفته برای ذخیره‌سازی داده‌های داوطلبانه به راحتی قابلیت تشخیص متناظر در مجموعه داده مرجع را داشته باشند. با انجام این کار مفاهیم استفاده شده در پایگاه داده‌های مختلف قابل ترجمه به یکدیگر بوده و می‌توان از منابع داده بیشتری به صورت همزمان استفاده کرده و قابلیت تعامل‌پذیری پایگاه داده‌های مختلف افزایش می‌یابد. در نتیجه علاوه بر به حداقل رساندن خطا در زمینه معنا و تعیین نوع داده در هنگام ذخیره‌سازی، استفاده از تمامی داده‌های ممکن و مناسب برای یک پرسش مطرح شده از جانب کاربر را ممکن می سازد، و موجب غنی‌سازی داده‌های داوطلبانه می‌شود.

مراجع

[۱] Hashemi, P., & Abbaspour, R. A. (2015). Assessment of Logical Consistency in OpenStreetMap Based on the Spatial Similarity Concept. In OpenStreetMap in GIScience (pp. 19-36). Springer International Publishing.

[۲] Haklay, M., A. Singleton, and C. Parker, (2008), "Web Mapping 2.0: The Neogeography of the GeoWeb", Geography Compass, 2, 6. p. 2011-2039.

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

[۴] OpenStreetMap, full history dump. http://wiki.openstreetmap.org/wiki/Planet.osm/full. Accessed 26 Dec. 2015

[۵] عشقی، م. مروری بر چالش‌های موجود در عدم قطعیت داده‌های مکانی داوطلبانه. سمینار دوره کارشناسی طوسی خواجه نصیرالدین صنعتی ارشد رشته مهندسی نقشه‌برداری - سیستم اطلاعات جغرافیایی. دانشگاه.

[۶] Wikipedia, 2015. http://wikipedia.org/wiki/Main_Page

[۷] و مرجع. پایان- VGI نقشه‌های بین هندسی اختلاف توصیف برای مکانی قواعد ر. (۱۳۹۲) استخراج عابدی،

.طوسی خواجه نصیرالدین صنعتی جغرافیایی. دانشگاه اطلاعات سیستم‌های رشته ارشد کارشناسی نامه

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

[۹] 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.

[۱۰] 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.

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

[۱۲] Li, B., Fonseca, F., 2006. Tdd: A comprehensive model for qualitative spatial similarity assessment. Spatial Cognition and Computation, 6, pp. 31–62.

[۱۳] Mooney, P., Corcoran, P., 2012a. The Annotation Process in OpenStreetMap. Transactions in GIS, 16(4), pp. 561–79.

[۱۴] Ballatore, A., Bertolotto, M., Wilson, D.C., 2012. Geographic knowledge extraction and semantic similarity in OpenStreetMap. Knowledge and Information Systems (KAIS), pp. 1–21.

[۱۵] Mülligann, C., Janowicz, K., Ye, M., Lee, W.-C., 2011. Analyzing the spatial-semantic interaction of points of interest in volunteered geographic information. Proceedings of the 10th international conference on Spatial information theory. Springer-Verlag, Berlin, Heidelberg, pp. 350–70.

[۱۶] Siebritz, L. A. (2014). Assessing the accuracy of openstreetmap data in south africa for the purpose of integrating it with authoritative data.

[۱۷] Koukoletsos, T. (2012). A Framework for Quality Evaluation of VGI linear datasets. PhD thesis, University College London, UK.