نشریه علمی علوم و فنون نقشه‌برداری، دوره ۱۳، شماره ۴، صفحه ۶۷ تا ۸۱، خرداد ماه ۱۴۰۳ - مقاله پژوهشی

ارائه چارچوب مفهومی و معماری زیرساخت دانش مکانی (SKI) با کاربست وب سرویس‌های مکانی

مرتضی امیدی‌پور[۱*]، آرا تومانیان[۲]، نجمه نیسانی سامانی[۳]، علی منصوریان[۴]

[۱] دانشجوی دکتری سنجش از دور و GIS، دانشکده جغرافیا، دانشگاه تهران، ایران
omidipoor@ut.ac.ir

[۲] دانشیار گروه سنجش از دور و GIS، دانشکده جغرافیا، دانشگاه تهران، ایران
a.toomanian@ut.ac.ir

[۳] دانشیار گروه سنجش از دور و GIS، دانشکده جغرافیا، دانشگاه تهران، ایران
nneysani@ut.ac.ir

[۴] استاد گروه گروه جغرافیای طبیعی و علوم اکوسیستم، دانشگاه لوند، سوئد
ali.mansourian@nateko.lu.se

(دریافت: بهمن ۱۴۰۱، تصویب: اردیبهشت ۱۴۰۳)

چکیده

با مطرح شدن پارادایم‌های کاربردپذیری و تعامل‌پذیری ماهیت و نحوه بکارگیری داده‌های جغرافیایی به سرعت در حال تغییر است. توسعه استانداردهای باز، مفاهیم سرویس‌گرایی، توسعه سامانه‌های اطلاعات جغرافیایی تحت وب و بکارگیری فرمت داده‌های جدید در قالب وب سرویس‌ها حاکی از آن است که پارادایم جدیدی در علم اطلاعات جغرافیایی شکل گرفته است. توسعه فناوری‌های مرتبط با علوم کامپیوتر و فناوری اطلاعات از یک سو و استفاده از سرویس‌های مکانی در کنار مفاهیم زیرساخت داده مکانی (SDI) از دیگر سو موجب شده که چالش دسترسی و انتشار داده مکانی تا حد زیادی رفع گردد. متاسفانه علی رغم این پیشرفت‌ها، استخراج دانش از مجموعه داده جغرافیایی به شکل موثر، کارا و در زمان مناسب همواره به عنوان چالشی بزرگ در جوامع اطلاعات جغرافیایی به قوت خود باقی مانده است. برای رفع این چالش استفاده از روش‌های داده‌کاوی، یادگیری ماشین، هوش مصنوعی و به به طور کلی اکتشاف دانش از داده جغرافیایی ضرورتی اجتناب ناپذیر است؛ با این وجود در بسیاری از سناریوها داده جغرافیایی ماهیتی توزیع یافته دارد. برای استخراج دانش از این مجموعه داده نیاز به روش‌های استخراج دانش توزیع یافته و زیرساختی فناورانه است که بتوان دانش مکانی را از منابع توزیع یافته استخراج کرد. به طور کلی در زیرساخت دانش مکانی (SKI) هدف ایجاد بستری است که بتوان دانش مکانی تولید کرد، به آن دسترسی داشت و دیگران را در آن سهیم ساخت. چنین زیرساختی دربرگیرنده مولفه‌هایی سازگار شامل نرم‌افزار، سخت افزار، داده، سرویس‌ها، روال‌ها و استانداردهایی است که در راستای کشف دانش در زیرساخت‌های داده مکانی و در حالتی توزیع یافته بکار گرفته می‌شود. در این نوشتار چارچوب مفهومی و معماری چنین زیرساختی با بکارگیری وب سرویس‌های مکانی ارائه شده است.

واژگان کلیدی: وب سرویس مکانی، داده‌کاوی، استخراج دانش، زیرساخت داده مکانی (SDI)، زیرساخت دانش مکانی(SKI).

* نویسنده رابط

۱- مقدمه

طی سالیان گذشته مشکل دسترسی و فقدان منابع داده همواره از مهمترین چالش‌های جامعه اطلاعات جغرافیایی بوده است. با فراگیر شدن سامانه‌های اطلاعات جغرافیایی (GIS) و توسعه نیازها و کاربردهای آن تلاش‌هایی در زمینه تولید داده جغرافیایی از سوی دولت‌ها، سازمان‌ها، موسسات و افراد مختلف صورت گرفت به گونه‌ای که با گذشت چند دهه حجم داده تولیدی به شدت افزایش یافته است. تولید داده از منابع مختلف خود سرآغاز و موجب شکل‌گیری چالش‌های جدید از جمله دسترسی، افزونگی، ناهمگونی، توزیع یافتگی، دوباره کاری، مشکلات سازمانی و بروکراتیک و نیز اشتراک‌گذاری داده مکانی شد [۱]. پس از شکل‌گیری مفاهیم زیرساخت داده مکانی (SDI) در پایان قرن بیستم و نهادینه شدن آن در سالیان اخیر مشکل دسترسی و انتشار داده مکانی تاحدی بر طرف شد؛ از سوی دیگر توسعه فناوری‌های مرتبط با علوم کامپیوتر و فناوری اطلاعات (IT) و استفاده از سرویس‌های مکانی، چالش سازگاری، افزونگی و ترکیب داده از منابع مختلف نیز تا حد زیادی رفع شده است.

افزایش بی‌سابقه در جمع‌آوری داده جغرافیایی خود چالش‌های جدید و اجتناب‌ناپذیری را موجب شده که از جمله آنها میتوان به استخراج اطلاعات سطح بالا و دانش در این اقیانوس داده به شکل موثر، کارا و در زمان مناسب اشاره کرد. در این راستا داده‌کاوی مکانی و استخراج دانش از داده جغرافیایی به عنوان یک حوزه پژوهشی و زمینه تحقیقاتی مهم ظهور پیدا کرد که تمرکز آن بر توسعه نظری، روش‌شناسی و ایجاد روش‌های عملیاتی برای استخراج اطلاعات و دانش مفید از مجموعه داده‌های بزرگ جغرافیایی است؛ اگرچه توسعه این روش‌ها بسیار ارزشمند بوده است با این حال در رابطه با استخراج دانش از داده مکانی مشکل کاربردپذیری و تعامل پذیری از جمله مشکلات اساسی است.

با مطرح شدن پارادایم‌های کاربردپذیری و تعامل‌پذیری ماهیت و نحوه بکارگیری داده‌های جغرافیایی به سرعت در حال تغییر است [۲]. توسعه استانداردهای باز، مفاهیم سرویس‌گرایی، توسعه سیستم‌های اطلاعات جغرافیایی تحت وب و بکارگیری فرمت داده‌های جدید از قبیل GML[۱] و دیگر فرمت‌های مبتنی بر XML[۲] در قالب وب سرویس‌ها همه حاکی از آن است که پارادایم‌های جدیدی در علوم اطلاعات جغرافیایی شکل گرفته است. برای استخراج دانش از داده‌های جغرافیایی در یک معماری توزیع شده و با استفاده از معماری سرویس-گرا مهمترین چالش ترکیب مولفه‌های توزیع شده و یا داده‌های ناهمگون در یک جریان کاوش داده به شکلی موثر، تعامل پذیر، با قابلیت و عملکرد مناسب و در نهایت به شکلی ساده است.

در بسیاری از سناریوها و به طور خاص در رابطه با SDI[۳] داده جغرافیایی جمع آوری شده از منابع مختلف ماهیتی توزیع شده دارد. بنابراین برای استخراج دانش از این مجموعه داده‌ها نیاز به روش‌های استخراج دانش توزیع یافته و زیرساختی فناورانه است که بتوان دانش مکانی را از منابع مختلف استخراج کرد. نیاز فوری به فناوری‌های جدید و ابزارهای خودکار می‌تواند نقش موثری در تبدیل این داده‌های عظیم به اطلاعات و دانش مفید داشته باشد [۳]. یکی از این فناوریها "وب سرویس‌ها" هستند. در حال حاضر روش‌های سنتی تبدیل داده جغرافیایی به دانش و تصمیم متکی بر تجزیه، تحلیل و تفسیر دستی و در بیشتر موارد در یک معماری متمرکز صورت می‌گیرد. چنین روش‌هایی بسیار کند، هزینه‌بر و وابسته به زمینه هستند و در آنها به دو رهیافت قابلیت استفاده مجدد [۴] و تعامل‌پذیری[۵] توجه نمی‌شود. در واقع با افزایش حجم داده جغرافیایی، این نوع تجزیه و تحلیلها در بسیاری از حوزه‌ها غیر عملی و یا اینکه محدود به داده و موضوعات خاصی می‌گردد. این در حالی است که ارزش واقعی چنین داده‌هایی در گزارش‌های مفید و اطلاعات سطح بالا مستتر است. علی‌رغم اینکه در یک چارچوب و معماری واحد نیازمندی‌های قابل توجهی برای استفاده از وب سرویس‌ها در راستای استخراج دانش از داده‌های جغرافیایی وجود دارد، با این حال وب سرویس‌های مبتنی بر دانش نسل جدیدی از قابلیت‌ها را به گونه‌ای فراهم می‌کنند که میتوان هوشمندانه و به طور خودکار از حجم عظیمی از داده جغرافیایی دانش مفید استخراج کرد. استفاده از وب سرویس‌ها به منظور استخراج دانش از داده جغرافیایی موضوعی است که هم از جنبه نظری و هم از جنبه کاربردی

۱ Geography Markup Language - GML
۲ Extensible Markup Language - XML
۳ Spatial Data Infrustructure - SDI
۴ Reusability
۵ Interoperability

کمتر مورد توجه قرار گرفته است. البته باید به این موضوع اشاره کرد که ترکیب وب سرویس‌های مکانی موجود برای تحلیل داده جغرافیایی در موضوعات مختلفی مورد توجه قرار گرفته است؛ ولی ارائه وب سرویسی اختصاصی در یک معماری توزیع یافته به منظور استخراج دانش موضوعی است که تاکنون کمتر مورد توجه قرار گرفته است و میتواند در حوزه‌های مختلف مورد استفاده قرار گیرد.

زیرساخت دانش مکانی (SKI) سازوکاری است گسترده و توزیع یافته که از طریق آن دانش مکانی ایجاد، ساماندهی، ذخیره، مدیریت و در نهایت به اشتراک گذاشته می‌شود. SKI، مکانیزمی را ایجاد می‌کند تا فرایندهای ضروری دانش جغرافیایی با بیشترین کارایی انجام و مورد استفاده قرار گیرد. هدف SKI ایجاد بستری است که بتوان دانش مکانی تولید کرد، به آن دسترسی داشت، آنرا به اشتراک گذاشت و در نهایت جامعه را از مزایا آن بهره‌مند کرد. چنین زیرساختی دربرگیرنده مولفه‌هایی سازگار شامل نرم‌افزار، سخت افزار، داده، سرویس‌ها، روال‌ها و استانداردهایی است که در راستای کشف دانش در مجموعه‌های بزرگ داده از قبیل زیرساخت‌های داده مکانی و در حالتی توزیع یافته بکار گرفته می‌شود.

در این نوشتار چارچوب مفهومی و معماری زیرساخت دانش مکانی شامل مولفه‌ها و ارتباطات میان آنها ارائه شده است.

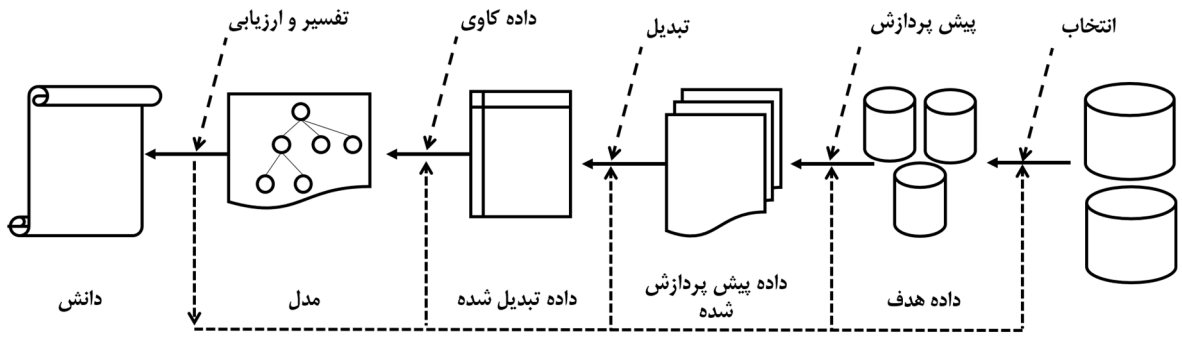
۲- پیشینه پژوهش

منصوریان (۲۰۰۶) و تومانیان (۲۰۱۲) مهمترین چالش‌های جوامع اطلاعات جغرافیایی در رابطه با داده را در ۴ مورد خلاصه کرده‌اند: ۱) موجود بودن[۶] داده ۲) قابلیت دسترسی[۷] به داده، ۳) کاربست‌پذیری[۸] و ۴) کاربردپذیری[۹] داده. در این حالت اهداف SDI در سطوح مختلف محلی، منطقه‌ای، ملی و بین‌المللی همسو با محدودیت‌های ذکر شده دنبال شد و در طول دو دهه گذشته، با هدف بهبود دسترسی، استفاده و به اشتراک گذاری داده جغرافیایی در سطوح مختلف مفهوم SDI توسعه یافته است.

هرچند به نظر می‌رسد کاربردپذیری داده در توسعه SDI کم‌تر مورد توجه قرار گرفته است، با این وجود به منظور حل چالش‌های مرتبط با داده مکانی در SDI چارچوبی بر مبنای پنج مولفه اساسی ارائه شده است. این مولفه‌ها عبارتند از: شبکه‌های دسترسی[۱۰]، سیاست‌ها[۱۱]، استانداردها[۱۲]، داده و مردم[۱۳]. یکی از مهمترین اهداف SDI اشتراک‌گذاری داده مکانی است. اشتراک گذاری داده مکانی به این معناست که داده برای اهداف گوناگون به طور مکرر مورد استفاده قرار گیرد و افزایش ارزش آن بدون افزایش هزینه باشد. ژنگ و لی[۱۴] (۲۰۰۵) دو شکل مختلف اشتراک گذاری داده مکانی را مورد بحث قرار داده‌اند. این دو عبارتند از ۱) اشتراک گذاری داده به صورت فایل [۱۵] و ۲) اشتراک گذاری بیوقفه داده[۱۶]. مسائل و مشکلات اشتراک گذاری داده مکانی به صورت فایل موجب استفاده از وب سرویس‌های مکانی در زمینه اشتراک گذاری داده مکانی شد. بنابراین با توسعه سیستم های تحت شبکه و توسعه Web2 اهمیت سیستم‌های توزیع شده و تعامل‌پذیر به منظور اشتراک‌گذاری داده مکانی رو به روز بیشتر می‌شود.

تمرکز صرف بر "داده" در چارچوب SDI موجب شد تا استخراج دانش از داده جغرافیایی همواره به عنوان یک چالش و شکاف باقی بماند. موضوع استخراج دانش از داده جغرافیایی در حوزه داده‌کاوی مکانی (SDM)[۱۷] قرار دارد. در GIS مبتنی بر دانش هدف کسب دانش جغرافیایی سطح بالا به واسطه استنتاج و نتیجه‌گیری از حقایق و روابط جغرافیایی ذخیره شده است [۶]. در این رابطه SDM رهیافتی ارائه می‌دهد که با بکارگیری آن میتوان دانشی از قبل ناشناخته، مفید و قابل تفسیر در حجم عظیمی از داده جغرافیایی استخراج کرد [۱۳]. در روش‌های فعلی (کلاسیک) به منظور استخراج دانش از داده‌های جغرافیایی مراحل مختلفی مورد انتظار است.

۶ Availability
۷ Accessibility
۸ Applicability
۹ Usability
۱۰ Access networks
۱۱ Policy
۱۲ Standards
۱۳ People
۱۴ Zhang, & Li
۱۵ File-level spatial data sharing
۱۶ Real-time spatial data sharing
۱۷ Spatial Data Mining (SDM)



شکل ۱- فرایند استخراج دانش به روش داده کاوی سنتی

انتخاب داده، پیش‌پردازش، تبدیل داده، اجرای الگوریتم‌های داده‌کاوری و در نهایت تفسیر و ارزیابی نتایج از جمله مهمترین مراحلی است که در داده‌کاوی کلاسیک مورد تاکید قرار می‌گیرد. مرحله نهایی در فرایند داده‌کاوی اشتراک دانش است؛ در این مرحله به واسطه روش‌های مختلف (از قبیل بصری سازی، درخت تصمیم، ساختارهای شرطی، زبان طبیعی و ...) دانش کشف شده از حجم عظیم داده به ذینفعان ارائه می‌گردد. لازم به ذکر است که در رهیافت سنتی فرایند داده‌کاوی در مرحله تولید دانش به اتمام می‌رسد و بحث اشتراک گذاری دانش به ندرت در رابطه با داده‌کاوی مورد تاکید قرار می‌گیرد. با این حال در حوزه مدیریت دانش بیشتر محققان عناصر اصلی چرخه مدیریت دانش را شامل تولید، ذخیره وساماندهی، اشتراک و استفاد می‌دانند [۱۶، ۱۵، ۱۴].

اگر اشتراک گذاری داده را فعالیتی بدانیم که داده توسط دیگران قابل دسترس باشد، بر این اساس میتوان اشتراک گذاری دانش مکانی را فعالیتی دانست که دانش استخراج شده از داده جغرافیایی را در اختیار دیگران قرار داد. در این رابطه دو سناریو مطرح است. ۱) دانش ممکن است در اختیار کسانی قرار گیرد که به داده مربوطه نیز دستری داشته باشند و ۲) دانش ممکن است در اختیار کسانی قرار گیرد که به داده مربوطه دسترسی ندارند. در SKI رویکرد دوم مورد تاکید است. بر این اساس در این پژوهش اشتراک‌گذاری داده و دانش مکانی به صورت تبادل داده و یا دانش بین دو یا چند کاربر (شخص یا سازمان) در قالب وب سرویس‌های مکانی و با بهره‌گیری از پروتکل‌های استاندارد تبادل داده در یک شبکه تعریف می‌گردد.

مطالعات اخیر توسط [۱۸، ۱۹، ۲۰ و ۲۱] نشان داده‌اند که چارچوب‌های پردازش موازی و توزیع‌شده، استخراج دانش از داده مکانی در مقیاس بزرگ را ممکن می‌سازد. در این زمینه، فناوری‌های مختلفی که اغلب آنها به صورت منبع باز ارائه شده است، برای محاسبات قابل اعتماد، مقیاس پذیر و توزیع شده توسعه یافته است. با استفاده از این فناوریها میتوان داده و محاسبات را در میان خوشه‌های کامپیوتری ذخیره و مدیریت کرد. هدوپ[۱۸] و سیستم فایل توزیع شده آن[۱۹] به همراه آپاچی اسپارک[۲۰]، از جمله فناوریهای پیشرو در این زمینه هستند که محاسباتی با کارایی بالا برای بازیابی الگوها و دانش از حجم عظیمی از داده‌های مکانی را تسهیل کرده است.

۳- چارچوب نظری

۳-۱- سرویس‌های مکانی

با پیشرفت زیرساخت‌های اطلاعاتی اهمیت استفاده از سرویس‌های وب به حدی رسیده که عناوینی چون "علم سرویس‌گرایی" در رابطه با محاسبات توزیع یافته و قابل تعامل مورد استفاده قرار می‌گیرد. پیشرفت‌های صورت گرفته در معماری سرویس‌گرا، GIS را نیز تحت تاثیر خود قرار داده و چشم‌اندازی روشن به منظور انجام عملیات توزیع یافته بر روی داده جغرافیایی فراهم شده است. در این میان توسعه GIS در بستر وب، تاکید بر تعامل‌پذیری داده جغرافیایی و همچنین استاندارد کردن این فرایندها موجب شد که در سال ۱۳۷۳ نهاد OGC [۲۱] به عنوان مرجعی بین‌المللی تحقیق، توسعه و تدوین استانداردهای سرویس‌های مکانی وب را مورد

۱۸ Hadoop
۱۹ HDFS
۲۰ Apache Spark
۲۱ Open Geospatial Consortium - OGC

توجه قرار دهد (http://www.opengeospatial.org). تدوین فرمت‌ها و کدگذاری‌های استاندارد و همچنین تدوین مشخصات پیاده‌سازی[۲۲] از جمله مهمترین فعالیت‌های OGC در رابطه با توسعه معماری سرویس‌گرا در GIS می‌باشد. چنین استانداردهایی در ادبیات سرویس‌گرایی تحت عنوان سرویس‌های مکانی OGC یا به اختصار OWS[۲۳] شناخته می‌شوند که جنبه‌های مختلفی از قبیل جست و جو، دریافت، ذخیره‌سازی، پردازش، نمایش و انتشار داده مکانی را شامل می‌شود.

در سال‌های اخیر سرویس‌های مکانی متنوعی با اهداف مختلف توسعه یافته که میتوان آنها را در چند دسته کلی مورد بررسی قرار داد. سرویس‌های شناسایی، سرویس‌های نمایشی، سرویس‌های مدیریت داده و سرویس‌های پردازشی از جمله مهمترین این موارد هستند. همچنین با توجه به اینکه در دهه اخیر مباحث مربوط به حوزه IOT در GIS متاثر از IT مورد تاکید فراوان قرار گرفته است، سرویس‌های مختلفی برای مدیریت، برنامه‌ریزی و دسترسی به داده‌های سنسورهای مکانی نیز توسعه یافته است. علاوه بر موارد ذکر شده در چند چند سال اخیر سرویس‌های متنوعی توسعه یافته که انتشار عمومی آنها در کارگروه‌های OGC در حال بررسی است. بنابراین میتوان پیشبینی کرد که در سال‌های اخیر سرویس‌های مکانی جنبه‌ها مختلفی از مدیریت داده‌های جغرافیایی را مورد توجه قرار خواهند داد و پارادیم سرویس‌گرایی در علم اطلاعات جغرافیایی بیش از پیش مورد استفاده قرار خواهد گرفت. چنین امری حتی تا جایی ممکن است پیش رود که پارادایم علمی GIS نیز دستخوش تغییر شود و سرویس‌های اطلاعات جغرافیایی بیش از سامانه‌های اطلاعات جغرافیایی مورد توجه جوامع دانشگاهی و تجاری قرار گیرد.

۳-۲- زیرساخت دانش مکانی(SKI[۲۴])

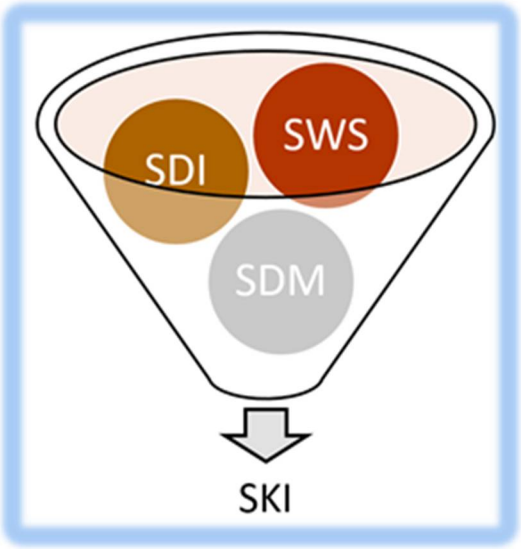
دانش مکانی اصطلاحی مبهم است که در رشته‌های مختلف مورد استفاده قرار می‌گیرد و تعاریف مختلفی نیز از آن ارائه شده است. به طور کلی در تعاریف مختلف اطلاعات سطح بالا یا حقایقی که به حل مسائل جغرافیایی در زمینه‌های مختلف کمک می‌کند تحت عنوان دانش مکانی شناخته می‌شود [۴]. به عبارت دیگر درک چهار پرسش اصلی (چه چیزی، چرا، چگونه و کجا) در رابطه با پدیده‌های جغرافیایی منجر به کشف و درک ماهیت پدیده‌های جغرافیایی و شناسایی قوانین و دانش مکانی خواهد شد [۵]. در این رابطه هندسه و توپولوژی دو ویژگی مهم برای تعیین رابطه بین اشیاء و پدیده‌های جغرافیایی می‌باشد [۶].

توسعه فناوری‌های اطلاعاتی از قبیل وب‌سرویس‌ها، چارچوب‌های توزیع یافته، اینترنت اشیاء (IOT) و GIS (از قبیل GIS فراگستر) منجر به آن شده است که ماهیت دانش مکانی در طول دهه‌های گذشته تکامل یابد. امروزه با ابداع مفاهیم جامعه و اقتصاد مبتنی بر دانش[۲۵] دسترسی بی‌وقفه به دانش مکانی در زمان مناسب، در شکل مناسب (تعامل‌پذیر)، در بستری مناسب (توزیع یافته) و با هزینه مناسب بیش از پیش حیاتی است. این بدان معناست که دانش مکانی را می‌توان به عنوان خروجی مجموعه‌ای از سرویس‌ها تصور کرد که روش کسب و اشتراک‌گذاری آن با حالت سنتی معمول تفاوت دارد و در قالب سرویس‌هایی مکانی مبتنی بر دانش معنا و هویت می‌یابد.

اگرچه SDI برای فرایندهای کسب، ذخیره‌سازی، اشتراک‌گذاری و پردازش داده مکانی سازوکار مناسبی را ارائه می‌دهد [۱، ۷، ۸،۹]، برای استخراج دانش مکانی و اشتراک‌گذاری آن سازوکاری تعامل پذیر وجود ندارد. از اینرو با بسط مفهوم SDI متناسب با چشم انداز تعریف شده از دانش مکانی چارچوب آن را میتوان به سمت زیرساختی مبتنی بر دانش سوق داد. چنین مکانیزمی را میتوان با ترکیب مفاهیم وب سرویس‌های مکانی (SWS)، داده‌کاوی مکانی (SDM) در بستر یک زیرساخت داده ایجاد نمود. چنین زیرساختی را تحت عنوان زیر ساخت دانش مکانی (SKI) خواهیم شناخت.

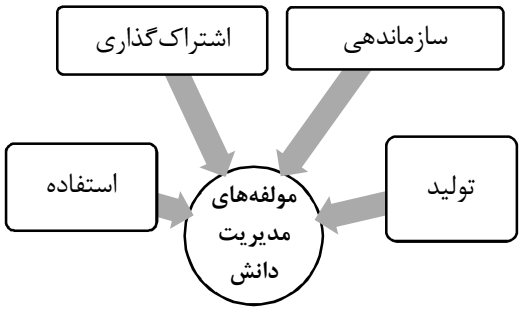
با در نظر داشتن مفاهیم اولیه، زیرساخت دانش مکانی دربرگیرنده مجموعه‌ای از مولفه‌های تعامل‌پذیر و سازگار شامل نرم افزار، داده، سخت افزار، فرایندها، استانداردها و سرویس‌هایی است که به منظور ایجاد، کاوش و اشتراک‌گذاری دانش مفید از داده‌های جغرافیایی توزیع شده با هم در ارتباطاند [۵].

۲۲ Implementation specification
۲۳ OGC Web Service
۲۴ Spatial Knowledge Infrustructure - SKI
۲۵ Knowledge society & Knowledge economy



شکل۲- ترکیب مفاهیم وب سرویس‌های مکانی (SWS)، داده‌کاوی مکانی (SDM) و قابلیت‌های زیرساخت داده مکانی (SDI) به منظور پشتیبانی از عملیات استخراج و اشتراک دانش مکانی

با توجه با اینکه GIS در حال حاضر به سمت رهیافت شبکه محور، فراگستر، محاسبه-مبنا و دانش-مبنا گام برمی‌دارد [۱۰، ۱۱، ۱۲] . بنابراین به زودی لزوم توسعه زیرساخت مبتنی بر دانش در رابطه با اطلاعات جغرافیایی شکل خواهد گرفت. تنها به واسطه چنین سرویس‌هایی میتوان انتظار داشت که GIS از پارادایمی داده محور، متمرکز و ایستا به پارادایمی شبکه مبنا، سرویس مبنا، فراگستر و دانش مبنا تغییر رویه دهد.



شکل ۳- عناصر اصلی چرخه مدیریت دانش

۳-۳- نقش وب سرویس‌های مکانی در استخراج و اشتراک‌گذاری دانش

گرچه فناوری‌های وب سرویس به طور گسترده در حوزه علوم اطلاعات جغرافیایی مورد استفاده قرار می‌گیرد، همواره چالش‌های قابل توجهی در توسعه سرویس‌های سطح بالاتر و اختصاصی در بخش‌ها و زمینه‌های مختلف علمی وجود دارد [۱۲]. در حال حاضر حجم زیادی داده جغرافیایی در سراسر جهان ذخیره شده است که هر روزه حجم این داده ها با سرعت فزاینده ای در حال افزایش است. چنین داده‌هایی خواه در بستر یک زیرساخت ساختار یافته همانند SDI و پایگاه داده‌ها و خواه به صورت غیرساختار یافته باشند، بدیهی است که چنین داده‌هایی بالقوه حاوی دانش جغرافیایی‌اند که در صورت استخراج میتوان از آنها در برنامه‌ریزی، تصمیم گیری، مدیریت و در نهایت حل مسائل جغرافیایی استفاده کرد.

علی رغم انجام تحقیقات قابل توجه در زمینه داده کاوی و ارائه الگوریتم های مختلف دو چالش اصلی وجود دارد [۱۷] به واسطه افزایش بیش از حد حجم داده‌ها استخراج دانش به صورت متمرکز میسر نیست و ۲) داده‌ها ماهیتی کاملا توزیع‌شده دارند و در زمینه استخراج دانش در حالت توزیع یافته تحقیقات چندانی صورت نگرفته است. این دو چالش در رابطه با داده‌های جغرافیایی نیز همچنان صدق می‌کند، مضاف بر آنکه داده جغرافیایی دارای ویژگی‌هایی هستند که استخراج دانش از آنها نسبت به سایر داده‌ها با مشکلات بیشتری همراه است؛ از جمله این ویژگی‌ها میتوان به وجود وابستگی‌های فضایی، چارچوب‌های اندازه‌گیری مختلف جغرافیایی، ناهمگونی فضایی، ابعاد زیاد، پیچیدگی فضایی-زمانی و در نهایت تنوع زیاد در نوع‌های داده اشاره کرد. برای استخراج دانش از چنین داده‌هایی نیازمندی‌های خاصی باید فراهم گردد. همچنین دانش استخراج شده از داده‌های جغرافیایی تنها محدود به ویژگی‌ها و حقایق جغرافیایی نیستند و دانش جغرافیایی در سطح روابط هندسی و توپولوژیکی نیز بسیار اهمیت دارد. در این زمینه تاکنون الگوریتم‌هایی نیز ارائه شده است. الگوریتم‌های طبقه‌بندی فضایی، وابستگی فضایی، خوشه‌بندی، قطعه‌بندی فضایی، قوانین انجمنی فضایی و روندیابی فضایی از جمله مهمترین الگوریتم‌هایی هستند که در این زمینه تاکنون ارائه شده است [۶].

معمولا در شیوه سنتی داده کاوی داده ها به صورت متمرکز در یک پایگاه داده (یا در حالتی غیر ساختاری) ذخیره وسپس دانش مفید از آن استخراج می‌شود[۱۷]. با توجه به معایبی که در رابطه با این معماری ارائه شده است به منظور استخراج دانش از داده‌های جغرافیایی در یک معماری توزیع یافته چارچوبی بر مبنای داده کاوی مکانی توزیع یافته را

میتوان متصور شد. داده کاوی توزیع یافته[۲۶] راه‌حلی مفید برای حل چالش حجم و توزیع یافتگی داده‌هاست. ایده کلی آن است که در الگوریتم‌های داده‌کاوی برای یادگیری مدل از مجموعه داده‌های مجزا (توزیع شده از نظر جغرافیایی) بهره گرفته شود و سپس به واسطه مدل آموزش دیده دانش کلی از تمام مجموعه داده‌ها استخراج شود. استفاده از چنین رهیافتی دو مزیت کلیدی را فراهم می‌سازد:

۱) صحت آموزش: به واسطه بکارگیری الگوریتم‌های داده کاوی بر روی مجموعه گسترده‌ای از داده‌های توزیع یافته صحت یادگیری بهتر از روش های متمرکز است.

۲) زمان محاسبه و محدودیت حافظه: محاسبات را میتوان بر روی یک شبکه توزیع کرد و از قابلیت‌های تمام سخت‌افزارها بهره جست.

حرکت اجتناب ناپذیر و سریع به سمت توزیع‌شدگی داده‌ها و محیط های محاسباتی پیچیده و ناهمگون چالش‌های تحقیقاتی جدیدی را نیز برای داده کاوی ایجاد کرده است. در برخی از منابع نیز زمینه‌ای تحت عنوان کشف دانش در محیط فراگستر[۲۷] (KDubiq) به تازگی در حال شکل گیری است. KDubiq زمینه تحقیقاتی نوظهور است که فصل مشترک دو چالش استخراج دانش از داده‌های توزیع شده و سامانه‌های متحرک[۲۸] است. هدف کشف دانش فراگستر [۲۹] ایجاد چارچوبی واحد برای بررسی سیستماتیک وابستگی‌های متقابل و دانش به واسطه فناوری‌های نسل جدید هوشمند از قبیل: یادگیری ماشین، داده کاوی، شبکه‌های حسگر، گریدها، سیستم‌های (P2P[۳۰])، سیستم‌های کاوش جریان داده، شناسایی فعالیت[۳۱] ، وب۲، حریم خصوصی، مدلسازی کاربر و غیره است [۱۰].

در یک جمع‌بندی از ادبیات مرتبط پژوهش میتوان چنین نتیجه گرفت که به منظور فراهم ساختن زیرساخت دانش مکانی(SKI) حداقل شرایط زیر باید فراهم باشد:

۱) دسترسی به داده های جغرافیایی به واسطه وب سرویس‌های مکانی مبتنی بر داده

۲) امکان استفاده از الگوریتم های داده کاوی مکانی در یک معماری توزیع شده

۳) استخراج دانش از داده‌های جغرافیایی توزیع یافته به واسطه وب سرویس مکانی مبتنی بر دانش

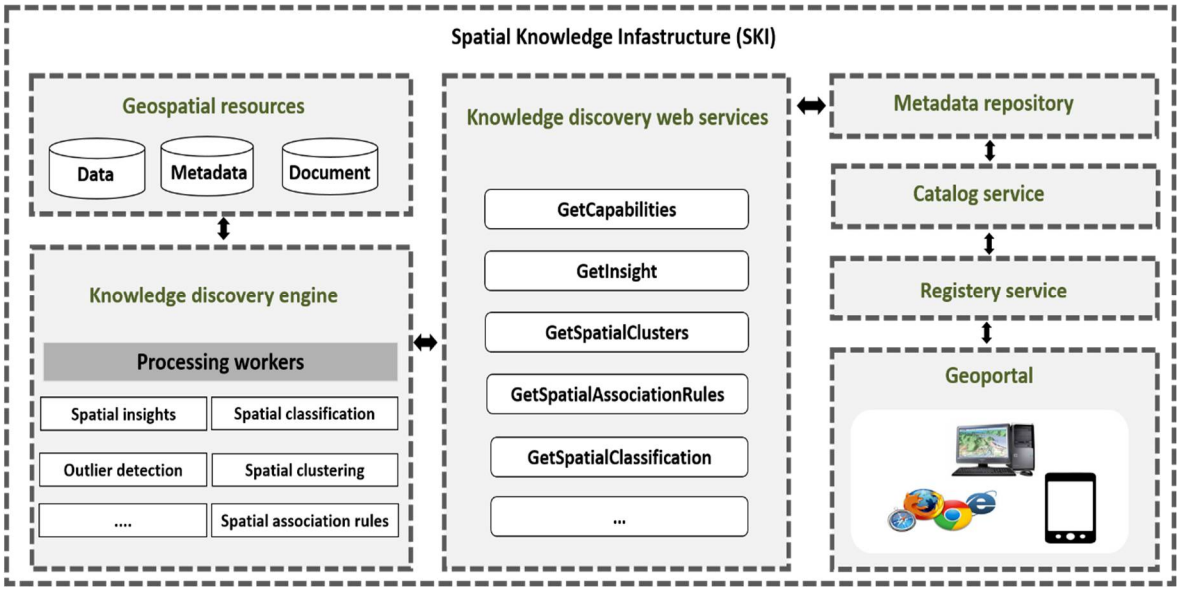
۴- معماری زیرساخت دانش مکانی

معماری زیرساخت دانش مکانی در شکل (۴) نمایش داده شده است. به منظور استخراج دانش از داده جغرافیایی در چارچوب پیشنهادی قابلیت‌های فناوریهای SOA و الگوریتم‌های داده‌کاوی مکانی در یک موتور محاسباتی توزیع شده و موازی ترکیب شده است. همانطور که در شکل زیر نشان داده شده است، برای ارائه عملکردها و قابلیت های مطلوب، این معماری شامل چهار لایه اصلی است. با استفاده از فناوری های مدرن ذخیره سازی [۳۲]، در لایه اول یکپارچه سازی و مدیریت داده‌های توزیع شده صورت می‌گیرد. در لایه دوم که با عنوان "موتور استخراج دانش" نام گذاری شده، پردازش و پشتیبانی از تکنیک های داده کاوی مکانی با کارایی بالا توسط کلاسترهایی از رایانه‌ها [۳۳] فراهم می‌شود. تعامل یکپارچه و متقابل بین کاربران و لایه‌های پایین (داده) توسط لایه سرویس میسر می‌شود. وب سرویس مکانی مبتنی بر دانش که به اختصار KDWS نامیده می‌شود در این لایه ایفای نقش می‌کند. در نهایت در بالاترین سطح کاربران و استفاده کنندگان قادر هستند که دانش موجود در داده مکانی را دریافت و به اشتراک گذارند. در ادامه اجزای معماری زیرساخت دانش مکانی شرح داده شده است.

۴-۱- لایه اول: داده

اولین لایه در معماری SKI داده می‌باشد که وظیفه آن یکپارچه سازی و ادغام منابع توزیع یافته است. باید توجه داشت که علاوه با توزیع‌یافتگی در اینجا با چالش ناهمگونی داده نیز مواجه هستیم. بنابراین هر چند که در SKI هدف آن نیست که داده‌ها را در یک مخزن واحد جمع آوری کرد ولی برای استخراج دانش با حداکثر کارایی لازم منابع داده به شکل مناسب در یک انباره داده بر خط ذخیره شود.

۲۶ Distributed data mining
۲۷ Ubiquitous
۲۸ Mobile systems
۲۹ Ubiquitous knowledge discovery
۳۰ Peer to peer
۳۱ Activity recognition
۳۲ Big data storage technologies
۳۳ Cluster of workers



شکل ۴- معماری زیرساخت دانش مکانی (SKI)

برای استخراج، تبدیل و بارگذاری داده در یک سیستم ذخیره سازی کلان داده مدرن، لازم است از ابزارهای ETL[۳۴] استفاده شود. از نقطه نظر عملیاتی این مولفه مسئول بارگذاری داده از منابع مختلف داده در یک خوشه HDFS است. HDFS برای پشتیبانی از فایل‌های بسیار بزرگ (ترابایت) طراحی شده و برای کاربردهایی که از الگوی «نوشتن-یک بار-خواندن-خیلی» پیروی می‌کنند به خوبی قابل استفاده است. بنابراین این فناوری کاملا با اهداف داده کاوی مکانی مطابقت دارد. یک خوشه HDFS[۳۵] از یک NameNode (وظیفه ذخیره متادیتا) و یک DataNodes (وظیفه ذخیره سازی داده واقعی) تشکیل شده است. در این حالت یک فایل به یک یا چند بلوک تقسیم می شود و این بلوک ها در مجموعه ای از DataNodes ذخیره می‌شوند. برای ارائه قابلیت اطمینان، هر فایل به صورت دنباله‌ای از بلوک‌ها نگهداری می‌شود. NameNode و DataNodeها دارای سرویس‌هایی هستند که نظارت بر وضعیت خوشه را فراهم می سازد. شرح جامع معماری HDFS را می توان در معماری Hadoop HDFS موجود در مستندات رسمی آن یافت.

در معماری پیشنهادی، طیف وسیعی از فرمت داده‌های مکانی توسط ابزارهای ETL قابل مدیریت است. خوشبختانه، در حال حاضر، می توان با استفاده از ابزارهای ETL مبتنی بر حافظه، حجم عظیمی از داده را به طور موثر ذخیره کرد. لازم به ذکر است که ETLهای مکانی مبتنی بر فراداده این امکان را فراهم می‌کنند که با استفاده از سرویس‌های مکانی فرایند ادغام و ذخیره سازی در HDFS صورت گیرد.

۴-۲- لایه دوم: موتور استخراج دانش

موتور استخراج دانش مولفه پردازشی برای استخراج دانش در یک پلتفرم کلان داده مکانی است. برای جلوگیری از تاخیر زمانی کم، از راه حل‌های پردازش موازی یا خوشه‌ای استفاده می‌شود. در این راستا از چارچوب های پردازش موازی متن باز مانند GeoSpark میتوان استفاده کرد. به طور کلی در این لایه مهندسان داده امکانی را فراهم میکنند تا فرایند استخراج دانش از داده مکانی توسزیع یافته با حداکثر کارایی صورت گیرد.

۴-۳- لایه سوم: سرویس‌های مبتنی بر دانش

امکان فراهم سازی زیر ساخت دانش مکانی بدون وجود این مولفه امکان پذیر نمی‌باشد. استخراج دانش در حالتی توزیع یافته و ترکیب دانش تنها به واسطه این مولفه قابل انجام است. این سرویس‌ها هم قادر به انجام الگوریتم‌های داده کاوی مکانی هستند و هم امکاناتی را برای خلاصه‌سازی داده‌ها فراهم می‌سازند. مهمترین ویژگی‌های این وب سرویس‌ها عبارتند از:

۳۴ Extract, Transform, Load (ETL)
۳۵ Hadoop Distributed File System (HDFS)

داینامیک بودن: کشف داینامیک و استفاده از داده‌ها و الگوریتم‌های داده‌کاوی در بستری مبتنی بر سرویس به صورت داینامیک صورت می‌گیرد.

مقیاس پذیر بودن: وب سرویس‌های مورد استفاده در چارچوب توسعه داده شده امکان ترکیب منابع داده جدید، الگوریتم های جدید و ابزارهای جدید را فراهم می سازد. از سوی دیگر کاربران با هر پلتفرمی قابل به دریافت سرویس هستند.

کاهش پیچیدگی: در مقایسه با روشهای سنتی استخراج دانش از داده های مکانی کاربر تمرکز خود را بر آنچه که از داده میخواهد می گذارد نه بر چگونگی انجام و جزئیات آن. در نتیجه چارچوب پیشنهادی پیچیدگی عملیات کاوش دانش از داده جغرافیایی ناهمگون و توزیع یافته را بسیار کاهش می‌دهد.

قابلیت توسعه: سرویس‌های مورد استفاده را میتوان برای الگوریتم‌ها و روش‌های مختلف استخراج دانش از داده‌های جغرافیایی تعریف کرد.

۴-۴- لایه چهارم: کاربرد

در زیرساخت دانش مکانی هدف اصلی کاربران کاوش، استخراج و اشتراک‌گذاری دانش مکانی است. در اینجا منظور از کاربر صرفا عامل انسانی نیست و سیستم‌ها و یا نرم‌افزارهای مختلف می‌توانند جزو کاربران محسوب گردند.

در حال حاضر هدف اصلی در ژئوپورتال‌ها اشتراک گذاری داده مکانی است. در این رابطه میتوان ژئوپورتالی را متصور شد که از طریق آن بتوان دانش نهفته در داده مکانی را به اشتراک گذاشت. اگر چه هدف اولیه از توسعه ژئوپورتال ها مباحث مرتبط با تحلیل داد و یا استخراج دانش نبوده است، در سال‌های اخیر ضرورت تسهیل قابلیت های تحلیلی، آماری و یا روش های استخراج دانش در ژئوپورتال‌ها در برخی از تحقیقات مطرح شده است [۱۹].. ژئوپورتال مبتنی بر دانش قابلیتی را در اختیار قرار می دهد تا به کمک آن بتوان الگوها و مدل های قابل فهمی از داده مکانی را کسب کرد. به منظور فراهم سازی چنین بستری در این لایه لازم است افراد یا سازمان ها متادیتا مربوط به سرویس های خود را ثبت نمایند. در این لایه نیز از سرویس‌های کاتالوگ و سرویس‌های ثبت متادیتا همانند روش‌های مرسوم در SDI استفاده خواهد شد.

۵- فناوری‌های مورد استفاده

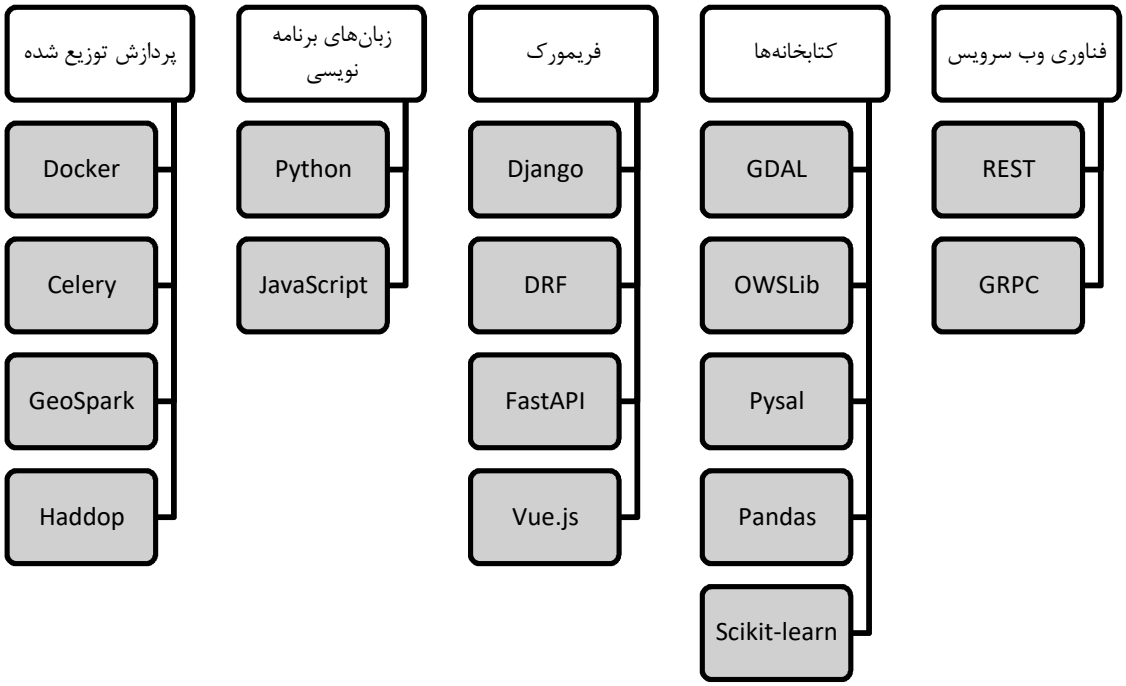
فناوری‌ها و ابزارهای مورد استفاده در پژوهش حاضر شامل فناوری‌های وب سرویس، کتابخانه‌ها، زبان‌های برنامه‌نویسی، فریمورک‌ها، و فناوری‌های مرتبط با پردازش‌های توزیع شده می‌باشد. تلاش بر آن بوده که در تمامی بخش‌ها ابزارهای متن باز و استانداردهای شناخته شده و مورد تایید نهادهای متولی امر از قبیل OGC مورد تاکید قرار گیرد. مهمترین فناوری‌ها و ابزارهای مورد استفاده در پژوهش حاضر در شکل زیر ارائه شده است.

از نقطه نظر فنی در زیر ساخت دانش مکانی مهمترین چالش در رابطه با فرایند درخواست، دریافت و اشتراک دانش بحث ناهمزمانی [۳۶] است. به عبارت ساده‌تر، در فرایند اشتراک‌گذاری دانش درخواست، دریافت و اشتراک دانش توسط افراد / سازمان‌ها باید بدون وقفه صورت گیرد. چنین چالشی بیشتر وابسته به فناوری است که راه‌حل آن استفاده از یک چارچوب پیام‌رسانی بی‌وقفه و ناهمزمان است. در این پژوهش به منظور تبادل پیام، دریافت و ارسال پاسخ به صورت بی‌وقفه از Celery استفاده شده است. این فناوری یک صف وظیفه ناهمگام [۳۷]منبع باز است که بر پایه تبادل پیام توزیع شده [۳۸] بنا شده است. این فناوری علاوه بر پیام‌رسانی از پردازش‌های زمان‌بندی[۳۹] شده نیز پشتیبانی می‌کند.

۶- پیاده‌سازی نمونه

با هدف کشف و استخراج دانش از داده‌های مکانی سرویسی تحت عنوان KDWS توسعه یافته است. این سرویس با هدف پیاده‌سازی مجموعه‌ای از رابط‌ها برای انجام فرایند استخراج دانش از مجموعه داده مکانی بر مبنای اصول معماری سرویس‌گرا توسعه یافته است. در این سرویس بر اساس پارامتر های مشخص شده در درخواست کاربر خروجی مورد انتظار (در فرمت های استاندارد مکانی) به صورت تعامل پذیر ارائه می‌گردد.

۳۶ Asynchronous
۳۷ Asynchronous task queue
۳۸ Distributed message passing
۳۹ Scheduling

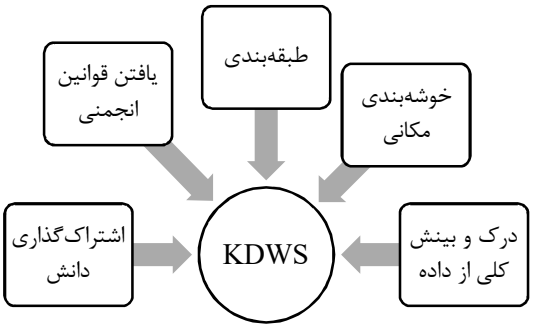


شکل ۵- فناوری‌های و ابزارهای مورد استفاده

اگرچه داده‌کاوی تنها یکی از روش های استخراج دانش است، با این حال در KDWS تاکید خاصی بر الگوریتم‌های داده کاوی مکانی و اشتراک گذاری دانش شده است.

در سرویس KDWS رابط‌هایی پیاده‌سازی می‌شود که قابلیت‌های را فراهم خواهد ساخت که به واسطه آن میتوان درک و بینش کلی از یک مجموعه داده مکانی بدست آورد، داده‌ها را به شکل مناسب خلاصه کرد، الگوریتم‌های داده کاوی مختلف را بر روی داده‌ها اعمال کرد و در نهایت دانش بدست آمده را به اشتراک گذاشت.

از نقطه نظر عملیاتی، کاربر درخواست خود را در قالب مجموعه‌ای از پارامترها به کارگزار سرویس ارئه می‌کند. با دریافت درخواست از سوی کارگزار یک وظیفه/ پردازش داده کاوی مکانی به موتور استخراج دانش ارسال میشود. در نهایت موتور استخراج، دانش بدست آمده را بر اساس پارامترهای دریافت شده تحویل می‌دهد. لازم به ذکر است که از طریق فناوری مجازی سازی و به طور خاص فناوری مبتنی بر کانتینر با استفاده از Docker شبیه سازی توزیع یافتگی داده سازماندهی شده است. در این رابطه سه ماشین مجازی شبیه سازی شد که در هر یک داده‌های مکانی در قالب سرویس‌های WFS، WMS و CSW منتشر شده است.

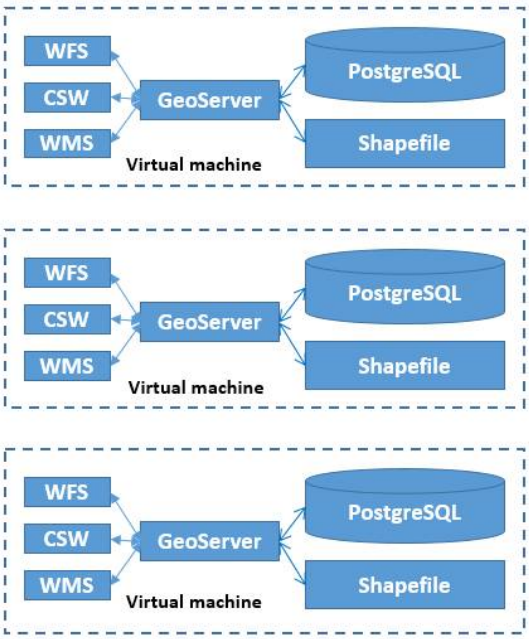


شکل ۶- قابلیت‌های سرویس KDWS

در شکل ۸ رابط کاربری نمونه سیستم پیاده‌سازی شده برای وب سرویس مکانی توسعه یافته ارائه شده است.

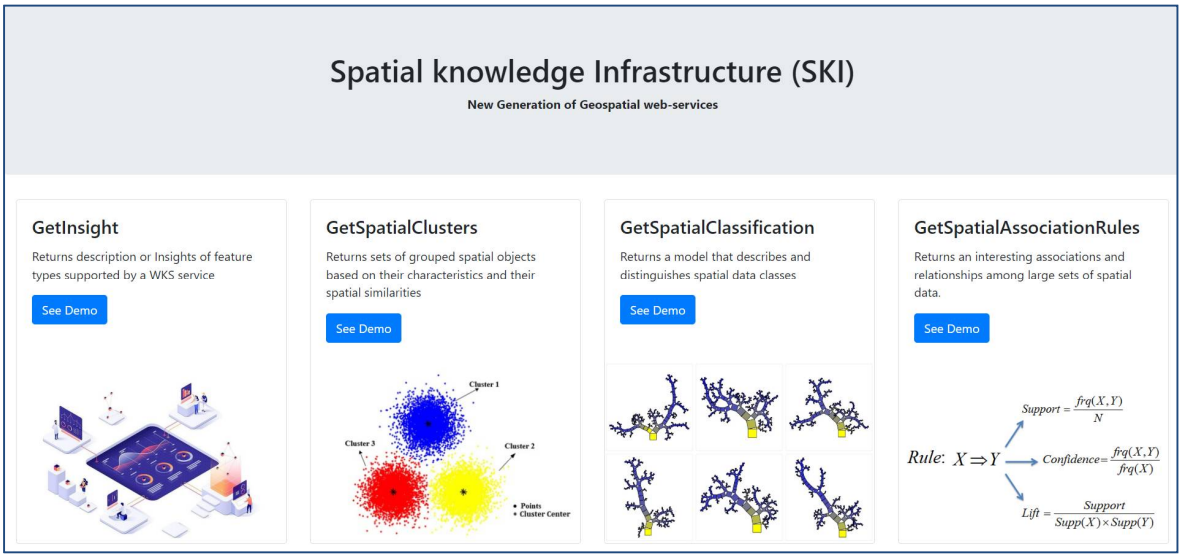
مهمترین مفاهیم مورد استفاده در KDWS عبارتند از سرویس، رابط و عملیات. سرویس مجموعه ای از رابط‌های ارائه شده توسط یک موجودیت است. سرویس های مختلف عملکردهای متفاوتی دارند که مستقل و قابل تفکیک هستند. به عنوان مثال، خدمات مختلف مانند بازیابی، نقشه برداری و پردازش به عنوان خدمات جداگانه در نظر گرفته می شوند. برای اجرای یک فرآیند (در اینجا استخراج دانش از داده مکانی) روش ارجاع درخواست‌ها به یک شی (در اینجا به معنای سرور است) توسط یک رابط مشخص می‌شود. در ساده ترین حالت، رابط KDWS نام عملیات، لیست پارامترها و مقادیر مجاز برای استخراج دانش از پایگاه داده را توصیف

می‌کند. در ارتباط با KDWS، عملیات مختلفی انتظار می رود که توسط رابط سرویس تعریف می‌شوند.



شکل ۷- نحوه شبیه‌سازی توزیع یافتگی داده به منظور پیاده‌سازی نمونه

مهمترین مفاهیم پایه‌ای در این رابطه عبارتند از: سرویس، رابط [۴۰] و عملیات[۴۱] . سرویس مجموعه‌ای از رابط‌های ارائه شده توسط یک موجودیت است. یک سرویس قابلیت هایی را فراهم می کند که یک موجودیت را از دیگر موجودیت‌ها متمایز می‌سازد (ISO19119: 2015) در این رابطه، منظور از موجودیت یک سرور در بستر وب است که در اغلب موارد از طریق پروتکل تبادل داده HTTP پاسخگوی نیازهای کاربران است. در واقع در تعریف سرویس به این نکته باید توجه داشت که سرویس‌های مختلف وظایف مختلفی دارند که از هم مستقل و قابل تفکیک هستند. برای مثال سرویس‌های مختلفی از قبیل نمایش، پردازش و کسب داده مکانی از یکدیگر تفکیک می‌شوند و در قالب سرویس‌های مجزا پیاده‌سازی می‌شوند. به منظور اجرای یک فرایند (در اینجا استخراج دانش از داده مکانی) نحوه ارجاع درخواست‌ها به یک شی (در اینجا منظور سرور) توسط یک رابط مشخص می‌گردد. در ساده‌ترین حالت رابط سرویس KDWS نام عملیات، لیست پارامترها و مقادیر مجاز برای استخراج دانش از پایگاه داده را توصیف می‌کند.



شکل ۸- رابط کاربری سیستم پیاده‌سازی شده نمونه

در رابطه با سرویس KDWS، عملیات‌های مختلفی مورد انتظار است که این عملیات‌ها توسط رابط سرویس تعریف می‌شود. شکل زیر رابط سرویس KDWS را در قالب یک نمودار کلاس UML [۴۲] نشان می‌دهد.

اگرچه عملیات‌های سرویس پیشنهادی تفاوت‌هایی با سایر سرویس‌های OGC ، از قبیل WPS، WFS، WCS و غیره دارد، شباهت‌های بسیاری نیز با رابط‌ها رایج وجود دارد. بنابراین با توجه به اینکه OWSها در پیاده‌سازی‌های مشترک وب سرویس‌ها از استانداردی تحت عنوان "مشخصات

۴۰ Interface
۴۱ Operation
۴۲ Unified Modeling Language - UML

پیاده‌سازی مشترک[۴۳] " بهره می‌گیرند، در این پژوهش نیز همین چارچوب مورد استفاده قرار گرفته است (رجوع شود به OGC 06-121r3). نحوه کدگذاری عملیات درخواست در سرویس KDWS بر اساس استاندارد (OGC 05-008) صورت گرفته است. بنابراین درخواست میتواند بر اساس HTTP GET بر اساس کدگذاری KVP و یا بر اساس HTTP POST و با استفاده از XML صورت گیرد. در ادامه توضیحات عملیات‌های سرویس ارائه شده است.

۶-۱- عملیات GetCapabilities

در این عملیات فراداده مربوط به سرویس KDWS فراخوانی می‌شود. به محض دریافت یک درخواست از سرویس گیرنده، مواردی از قبیل عملیات‌ها، داده‌های در دسترس (در سرویس KDWS جداول رابطه‌ای مد نظر است) و همچنین الگوریتم‌های استخراج دانش پشتیبانی شده به ارسال کننده درخواست ارسال می‌شود. این فراداده در قالب XML ارسال می‌گردد.

برای درخواست فایل متادیتا سرویس KDWS نمونه درخواست بر اساس کد گذاری KVP به شکل زیر است.

```
http://demoserver.ut.edu/KDWS?service=KDWS&request
=getCapabilities
```

۶-۲- عملیات GetInsight

در این عملیات تلاش بر آن است که بینشی کلی از داده جغرافیایی کسب شود. بینش می‌تواند از طریق خلاصه‌سازی داده، توزیع داده، رابطه بین متغیرها و غیره باشد. در ساده ترین حالت کاربر مایل است که خلاصه‌ای از داده مکانی مد نظر خود در قالب فرمت‌های مختلف (از قبیل نقشه، Json، XML و غیره) داشته باشد. معیارهایی همانند میانگین، میانه، چارک‌ها و غیره برای هر فیلد و یا به طور کلی برای جدول درخواستی از جمله مواردی است که در خلاصه‌سازی داده مورد تاکید است. از سوی دیگر توزیع داده‌ها و ارتباط بین متغیرها نیز از جمله مواردی‌اند که در سرویس KDWS مورد انتظار است. دو نوع اصلی کسب بینش که در عملیات GetInsight پیاده‌سازی شده است، عبارتند از Summarization، Relations.

برای درخواست عملیات GetInsight در سرویس KDWS نمونه درخواست بر اساس کد گذاری KVP به شکل زیر است.

```
http://demoserver.ut.edu/KDWS?service=KDWS&version=
1.0.0&request=getinsight&insighttype=relations
&layers=demo_table&variables=a_1, a_2, a_3, a_4, a_5,
a_6, a_7, a_8, a_9 &outputformat=image/png
```

خروجی سرویس KDWS در مثال فوق به صورت شکل زیر به کاربر ارائه خواهد شد.

۶-۳- عملیات GetClusters

خوشه بندی یکی از پایه‌ای ترین الگوریتم‌های داده‌کاوی است، از این رو یکی از عملیات‌های سرویس پیشنهادی را به خود اختصاص داده است. در این عملیات کاربر نوع الگوریتم، مجموعه داده و متغیرهای مربوطه را به سرور ارسال میکند و پاسخی متناسب با پارامترهای ارسالی دریافت خواهد کرد. برای درخواست عملیات GetClusters در سرویس KDWS نمونه درخواست بر اساس کد گذاری KVP به شکل زیر است.

```
http://demoserver.ut.edu/KDWS?service=KDWS&version=
1.0.0&request= getclusters &algorithm=k-means
&layers=demo_table&variables= a_1, a_2, a_3, a_4, a_5,
a_6, a_7, a_8, a_9 &outputformat=application/json
```

جدول ۱- پارامترهای مهم عملیات GetClusters

عملیات	پارامتر	نوع
خوشه بندی	Dataset	اجباری
	Algorithem	اجباری
	Options	وابسته

در جدول ۱ نیز مهمترین پارامترهای عملیات GetClusters ارائه شده‌است.

۶-۴- عملیات GetClassification

پیاده‌سازی الگوریتم‌های طبقه‌بندی در قالب وب سرویس پیچیدگی بیشتری نسبت به دیگر الگوریتم‌ها دارد. دلیل این پیچیدگی را باید در وجود داده آزمایش و آموزش

۴۳ Common implementation specification

مدل برای اجرای طبقه‌بندی دانست. برای حل این پیچیدگی در رابط سرویس KDWS قابلیت در نظر گرفتن بخشی از مجموعه داده برای آزمایش و یادگیری مدل فراهم شده است. برای مثال میتوان از مجموعه داده ۲۰ درصد از داده را به صورت رندم به عنوان داده آزمایش در نظر گرفت و بعد از اجرای الگوریتم صحت مدل را محاسبه کرد. برای درخواست عملیات GetClassification در سرویس KDWS نمونه درخواست بر اساس کد گذاری KVP به شکل زیر است.

```
http://demoserver.ut.edu/KDWS?service=KDWS&version=1.0.0&request= getclassification
&algorithm=knn&layers=demo_table&testset=20&label=a10&variables=a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9
&outputformat=application/json
```

همچنین باید به این نکته اشاره کرد که تعداد پارامترهای اجباری و اختیاری بسته به الگوریتم مورد نظر متفاوت است.

۶-۵- عملیات GetAssociationRules

یکی از جذاب‌ترین الگوریتم‌های داده‌کاوی الگوریتم کشف قوانین انجمنی است. قوانین انجمنی روابط متقابل بین مجموعه داده را به صورت یک سری قوانین ارائه می‌دهد. از اینرو چنین دانشی در حوزه‌های مختلف مورد توجه بوده و کاربردهای متفاوتی دارد. در رابط سرویس پیشنهادی یکی از عملیات‌های مهم GetAssociationRules است. بسیاری از قوانینی که در رابطه با داده‌های مکانی تولید می‌شوند ممکن است اهمیت زیادی نداشته باشند بنابراین در سرویس پیشنهادی قوانین استخراج شده به همراه دو معیار ارزیابی مهم یعنی Support و Confidence ارائه می‌شوند. بنابراین در تمامی قوانین استخراج شده در رابطه با داده‌های مکانی ارزش و معیار مقبولیت قوانین انجمنی نیز ارائه خواهد شد. برای درخواست عملیات GetAssociationRules در سرویس KDWS نمونه درخواست بر اساس کد گذاری KVP به شکل زیر است.

```
http://demoserver.ut.edu/KDWS?service=KDWS&version=1.0.0&request= GetAssociationRules &algorithm= appriori
&layers=demo_table &label=v10&variables=v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9
&outputformat=application/json
```

با اعمال عملیات GetAssociationRules روابط موجود بین متغیرهای مختلف به صورت قوانین انجمنی در قالب Json به کاربر ارائه می شود.

۷- بحث و نتیجه‌گیری

زیرساخت دانش مکانی سازوکاری است گسترده که از طریق آن دانش مکانی ایجاد، ساماندهی، ذخیره، مدیریت و ارائه می‌شود. SKI، مکانیزمی را ایجاد میکند تا فرایندهای ضروری دانش جغرافیایی با بیشترین کارایی انجام و مورد استفاده قرار گیرد. هدف SKI ایجاد بستری است که بتوان دانش مکانی تولید و به آن دسترسی داشت، آنرا به اشتراک گذاشت و در نهایت جامعه را از مزایا آن بهره‌مند کرد. چنین زیرساختی دربرگیرنده مولفه‌هایی سازگار شامل نرم‌افزار، سخت افزار، داده، سرویس‌ها، روال‌ها و استانداردهایی است که در راستای کشف دانش در مجموعه‌های بزرگ داده از قبیل زیرساخت‌های داده مکانی و در حالتی توزیع یافته بکار گرفته می‌شود. در این رابطه ابتدا چارچوب مفهومی زیرساخت دانش مکانی مورد بررسی و کنکاش قرار گرفت. علاوه بر تبیین مولفه‌ها و معماری، یک وب سرویس مکانی مبتنی بر دانش به منظور فرایند استخراج و اشتراک‌گذاری دانش توسعه یافته و قابلیت‌های آن مرور شد.

در یک جمع بندی کلی میتوان موارد زیر را به عنوان نقاط قوت و مزایای پژوهش برشمرد:

- استخراج دانش از داده‌های مکانی مبتنی بر سرویس‌های مکانی، فراتر از روش‌های سنتی و متمرکز
- معرفی و ارائه چارچوب مفهومی و معماری زیرساخت دانش مکانی
- وب سرویس مکانی توسعه داده شده در این پژوهش (KDWS) امکان بروزرسانی منابع داده جدید، و الگوریتم های جدید را فراهم می سازد. از سوی دیگر کاربران با هر پلتفرمی قابل به دریافت سرویس هستند.
- سرویس‌های مورد استفاده را میتوان برای الگوریتم‌ها و روش‌های مختلف استخراج دانش از داده‌های جغرافیایی تعریف کرد.

• با توجه به اینکه خروجی الگوریتم‌ها در قالب استاندارد و تعامل پذیر ارائه می‌گردد بنابراین فرایند اشتراک دانش تسهیل می‌گردد.

• تمامی فناوری‌های مورد استفاده در پژوهش حاضر Open Source بوده است.

علاوه بر موارد ذکر شده پژوهش حاضر دارای محدودیت‌هایی به شرح ذیل است:

سناریو پژوهش با فرض موجود بودن و قابل دسترس بودن داده مکانی از طریق سرویس‌های OGC صورت گرفته است؛ حال آنکه هنوز هم بسیاری از داده‌های مکانی در قالبی استاندارد و در یک شبکه قابل دسترس منتشر نشده و این مساله استخراج دانش به شکل توزیع یافته را با چالش‌هایی همراه می‌سازد. از سویی دیگر بحث مجوزها و قوانین دسترسی در ارتباط با داده‌ها در این پژوهش مورد توجه قرار نگرفته است.

۸- پیشنهادات

در این پژوهش تمرکز وب سرویس مکانی مبتنی بر دانش بر روی فرمت داده‌های وکتور بوده است که این داده‌ها غالبا در پایگاه داده‌های مکانی توزیع شده ذخیره و بازیابی شده است. بدیهی است بخش عظیمی از داده‌های مکانی خصوصاً تصاویر ماهواره‌ای در فرمت‌های رستری و غالبا به صورت File-base ساماندهی می‌شود؛ بنابراین پیشنهاد می‌شود در توسعه آتی سرویس‌های مکانی مبتنی بر دانش، روش‌های استخراج دانش از داده‌های توزیع یافته رستری نیز مورد تاکید قرار گیرد.

تکنولوژی مورد استفاده در اغب سرویس‌های مکانی مبتنی بر SOAP و اخیرا تمرکز بیشتر بر روی سرویس‌های مبتنی بر REST بوده است. به طور کلی همواره پروتکل HTTP/1 در این زمینه مورد تاکید بوده است. با توجه به اینکه انتخاب تکنولوژی وب سرویس‌های مکانی جهت استفاده در سیستم‌های نرم افزاری، یکی از مهمترین تصمیم‌ها می‌باشد، پیشنهاد می‌شود در یک مطالعه مستقل قابلیت‌های وب سرویس‌های مبتنی بر gRPC و به طور کلی HTTP/2 برای مدیریت داده‌ها و یا استخراج دانش از داده‌های مکانی مورد بررسی قرار گیرد. gRPC در مقایسه با راهکارهای کنونی مزایای مختلفی را ارائه می‌دهد. سرعت و کارایی بهتر، ارتباط دوسویه و statefull، ساختار و تعریف مشخص و قابلیت‌های streaming از جمله مهمترین این مزیت‌ها می‌باشد.

ترکیب وب‌سرویس‌های مکانی مبتنی بر دانش زمینه تحقیقاتی مهم دیگری است که پیشنهاد می‌شود در تحقیقی مستقل مورد بررسی قرار گیرد. به نظر می‌رسد که ترکیب چندین وب سرویس مکانی مبتنی بر دانش موجب تولید و یافتن دانشی جدیدتر خواهد شد که میتواند اهمیتی اساسی داشته باشد. در این رابطه روش‌های خودکار ترکیب وب‌سرویس‌ها، روش‌های مبتنی بر آنتولوژی، روش‌های مبتنی بر سیستم‌های خبره و غیره میتواند مورد استفاده قرار گیرد.

مراجع

[۱] Toomanian, A. (2012) Methods to Improve and Evaluate Spatial Data Infrastructures. Lund University, Sweden. Available at: http://www.lunduniversity.lu.se/lup/publication/40d93fc-98ca-4f12-becd-a0e8d25c9048.

[۲] Li, D., Wang, S. and Li, D. (2016) Spatial data mining : theory and application. Available at: http://www.springer.com/us/book/9783662485361 (Accessed: 28 August 2017).

[۳] Talia, D. and Trunfio, P. (2013) Service-oriented distributed knowledge discovery. CRC Press.

[۴] Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). Geographic information science and systems. John Wiley & Sons.

[۵] Omidipoor, M., Toomanian, A., & Samani, N. N. (2018, June). Towards Spatial Knowledge Infrastructure (SKI): Technological Understanding. In Proceedings of the 21st AGILE International Conference on Geographic Information Science, Lund, Sweden (pp. 12-15).

[۶] Miller, H. J., & Han, J. (2009). Geographic data mining and knowledge discovery. CRC press.

[۷] Rajabifard, A., Feeney, M.-E. F. and Williamson, I. P. (2002) 'Future directions for SDI', International Journal of Applied Earth Observation and Geoinformation. Elsevier, 4(1), pp. 11–22.

[۸] Williamson, I. P., Rajabifard, A. and Feeney, M.-E. F. (2003) Developing spatial data infrastructures : from concept to reality. Taylor & Francis.

[۹] Mansourian, A., Rajabifard, A., Valadan-Zoje, M. and Williamson, I. (2006) 'using SDI and web-based system to facitilate disaster management', Computers & Geosciences, 32(3), pp. 303–315. doi: http://dx.doi.org/10.1016/j.cageo.2005.06.017.

[۱۰] Kargupta, H. (2009) Next generation of data mining. CRC Press.

[۱۱] Li, S., Dragićević, S. and Veenendaal, B. (2011) Advances in web-based GIS, mapping services and applications. CRC Press/Balkema.

[۱۲] Yue, P. (2013) Semantic web-based intelligent geospatial web services, springer. Available at: http://link.springer.com/content/pdf/10.1007/978-1-4614-6809-7.pdf.

[۱۳] Li, S., Dragicevic, S., Castro, F. A., Sester, M., Winter, S., Coltekin, A., ... & Cheng, T. (2016). Geospatial big data handling theory and methods: A review and research challenges. ISPRS journal of Photogrammetry and Remote Sensing, 115, 119-133.

[۱۴] Dalkir, K. (2013). Knowledge management in theory and practice: Routledge.

[۱۵] Easterby-Smith, M., & Lyles, M. A. (Eds.). (2011). Handbook of organizational learning and knowledge management. John Wiley & Sons.

[۱۶] Rubenstein-Montano, B., Liebowitz, J., Buchwalter, J., McCaw, D., Newman, B., Rebeck, K., & Team, T. K. M. M. (2001). A systems thinking framework for knowledge management. Decision support systems, 31(1), 5-16.

[۱۷] Zhong, N. and Zhou, L. (1999) Methodologies for knowledge discovery and data mining : Third Pacific-Asia Conference, PAKDD-99, Beijing, China, April 26-28, 1999 : proceedings. Springer.

[۱۸] Omidipoor, M., Toomanian, A., Neysani Samany, N., & Mansourian, A. (2020). Knowledge discovery web service for spatial data infrastructures. ISPRS International Journal of Geo-Information, 10(1), 12.

[۱۹] Alkathiri, M., Jhummarwala, A., & Potdar, M. B. (2019). Multi-dimensional geospatial data mining in a distributed environment using MapReduce. Journal of Big Data, 6(1), 82.

[۲۰] Li, Z., Gui, Z., Hofer, B., Li, Y., Scheider, S., & Shekhar, S. (2020). Geospatial information processing technologies. Manual of digital earth, 191-227.

[۲۱] Jo, J., & Lee, K. W. (2018). High-performance geospatial big data processing system based on MapReduce. ISPRS International Journal of Geo-Information, 7(10), 399.