Archive of SID

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

# ارائه ژئوپرتالی مبتنی بر سیستم‌های توصیه گر معنایی

**علیرضا رحمتی*[۱]، آرا تومانیان[۲]، نجمه نیسانی سامانی[۲]**

[۱] کارشناسی ارشد سنجش از دور و سیستم‌های اطلاعات مکانی - دانشکده جغرافیا - دانشگاه تهران
rahmatya685@ut.ac.ir

[۲] استادیار گروه سنجش از دور و سیستم‌های اطلاعات مکانی - دانشکده جغرافیا - دانشگاه تهران
{a.toomanian, nneysani}@ut.ac.ir

(تاریخ دریافت بهمن ۱۳۹۴، تاریخ تصویب مرداد ۱۳۹۵)

## چکیده

حجم بالای داده یک مشکل عمده در وب کنونی به شمار می‌رود و از طرف دیگر هشتاد درصد داده‌های موجود نیز دارای بعد مکانی هستند .بنابراین مسئله گرانباری[۱] اطلاعات در علوم اطلاعات مکانی نیز چالش برانگیز است. حجم گسترده داده و سرویس‌های موجود نیز هنگامی مفید است که برای کاربر قابل‌کشف شدن باشد. در ژئوپرتالها[۲] کاربران با مشکلاتی از قبیل انتخاب کلیدواژهای مناسب برای جستجو، چگونگی پرکردن اطلاعات فرم جستجو، انتخاب فیلترهای جستجوی مناسب، روالهای متفاوت جستجو در ژئوپرتالها و غیره که سبب سردرگمی آنها می‌شود مواجه می‌شوند. راه حل‌های بسیار برای حل این مسائل ارائه‌شده است که اغلب سعی در افزایش توانایی جستجوگری بر اساس کلیدواژه‌های ارائه‌شده از سوی کاربر تأکید داشته‌اند؛ اما همچنان این مشکلات و عدم همسانی‌ها در ژئوپرتالها وجود دارد. هدف این پژوهش ارتقای عملکرد جستجوگری ژئوپرتالها با افزودن مکانیسم بازیابی اطلاعات جدید در ژئوپرتال است. سیستم‌های توصیه‌گر روشی جدید در بازیابی و ارائه اطلاعات به کاربران هستند که روالی مجزا از مکانیسم‌های جستجوگری موجود در ژئوپرتالها ارائه می دهند. پژوهش حاضر در پی ترکیب سیستم‌های توصیه گر با ژئوپرتال است. در این صورت می‌توان نارسایی‌های موجود در ژئوپرتالها را از طریق ارائه آیتم‌های پیشنهادی بر اساس پروفایل کاربر و فراداده و استخراج ارتباطات بین آن‌ها کاهش دهد. در این راستا به منظور استخراج مشابهت‌های بین پروفایل کاربر و فراداده از هستی‌شناسی[۳] استفاده‌شده است. سیستم توصیه گر معناگرا[۴] بر پایه الگوریتم تجزیه مقادیر تکین (SVD[۵]) و تحلیل مشابهت معنایی[۶] جهت استخراج آیتم‌های اطلاعاتی مرتبط و حل مشکل شروع سرد طراحی‌شده است. برای پیاده‌سازی ژئوپرتال از پایگاه داده مکانی سازمان راهداری استفاده‌شده است و طراحی هستی‌شناسی نیز بر پایه مفاهیم مرتبط در این حوزه بوده است.

**واژگان کلیدی:** سیستم‌های توصیه گر، ژئوپرتال، وب معنایی، مشابهت هستی‌شناسی

* نویسنده رابط

۱ Overload
۲ Geoportal
۳ Ontology
٤ Semantic Recommender
٥ Singular Value Decomposition
٦ Ontological Similarity



www.SID.ir

Archive of SID



## ۱- مقدمه

کاربرد گسترده داده‌های مکانی در عرصه‌های مختلف تصمیم سازی، مدیریت منابع و غیره؛ و در سطوح مختلف ملی، محلی، منطقه‌ای بر کسی پوشیده نیست (Buehler,McKee et al. 1996). علاوه بر داده‌های موجود، نرخ رشد داده‌های در حال تولید نیز روزبه‌روز در حال افزایش است. این نرخ رشد بالا حاصل پیشرفت‌های تکنولوژیکی مانند افزایش روزافزون قدرت تفکیک تصاویر ماهواره‌ای، افزایش حجم و توان نرم‌افزاری‌های مدیریت پایگاه داده (DBMS) و افزایش توان پردازشی نرم‌افزارها و همچنین پیشرفت‌های عظیم سخت‌افزاری و ظهور تکنولوژی رایانش ابری و... می‌باشد. علاوه بر این هرروزه داده‌های موجود از هر طیفی (تصویر، ویدئو، صدا، داده‌های تاریخی، اسناد علمی و هر نوع اقلام اطلاعاتی قابل‌تصور) توسط سازمان‌های دولتی، خصوصی و مردم عادی زمین مرجع می‌شوند(Nogueras-Iso, Muro-Medrano et al. 2005). این حجم عظیم داده سبب می‌شود که کاربران مختلف نتوانند به‌راحتی داده موردنیاز را کشف نمایند. بنابراین می‌بایست راهکارهایی جهت سهولت جستجو، ارزیابی و دسترسی به داده‌ها ارائه شود. زیرساخت داده مکانی[۱] نظامی است که یکی از اهداف آن برآورده ساختن همین نیاز است. (Infrastructures 2004). ژئوپرتال به‌عنوان درگاه این زیرساخت، محلی برای انتشار و جستجوی داده‌ها و سرویس‌های مکانی است و می‌توان گفت ۵۰ درصد از کاربرانی که از ژئوپرتال استفاده می‌کنند به دنبال کشف داده و دسترسی به آن هستند(Resch and Zimmer 2013). بنابراین کارکرد اصلی ژئوپرتالها جستجو و کشف داده‌ها و سرویس‌های مکانی است.

حجم گسترده داده و سرویس‌های موجود، هنگامی مفید است که برای کاربر قابل‌کشف شدن باشد. در ژئوپرتالهای موجود جستجوی منابع برای کاربران همواره با مشکلاتی روبروست. ممکن است کاربران ندانند از چه کلیدواژه‌هایی برای جستجوی منابع استفاده کنند، چگونه اطلاعات فرم جستجو را پر نمایند و یا اینکه از کدام فیلتر جستجو استفاده نمایند(Larson, Hochmair 2005 , Siliceo et al. 2006). واژه‌ها و توصیفاتی که در فراداده‌ها وجود دارد، حتی در شرایطی که عبارت جستجو بسیار واضح و گویا باشد، می‌تواند عامل محدودکننده‌ی جستجوی کاربران باشد. نتایج جستجو کاملاً بستگی به عناصر و توصیفات به‌کاررفته در فراداده دارد( Athanasis, Kalabokidis et al. 2009). علاوه بر این به سبب زمینه‌های علمی متفاوت کاربران و ارائه‌دهندگان داده، کلیدواژه‌های متفاوتی برای جستجو و انتشار یک داده به‌کاربرده می‌شود که سبب بروز ناهمگونی معنایی[۲] در جستجوی منابع می‌شود. علاوه بر این ممکن است زبان کاربر و ارائه‌دهنده متفاوت باشد و یا درک متفاوتی از یک مفهوم حتی در بین کاربران مختلف وجود داشته باشد. موارد مذکور سبب می‌شود که نتایج جستجو بسیار کم باشد و اقلام اطلاعاتی کمی بازگردانده شود و همچنین ممکن است نتایج جستجو از دقت کمتری برخوردار بوده و اقلام اطلاعاتی مدنظر کاربر همراه با گستره وسیعی از اقلام اطلاعاتی دیگر بازگردانده شود(Smits and Friis-Christensen 2007 , Athanasis, Kalabokidis et al. 2009). راه حل‌های بسیاری برای حل مسائل مرتبط با ناهمگونی‌های معنایی و زمینه‌های علمی متفاوت کاربران و تهیه‌کنندگان داده ارائه‌شده است که اغلب سعی در افزایش توانایی جستجوگری بر اساس کلیدواژه‌های ارائه‌شده از سوی کاربر تأکید داشته‌اند. درواقع رویکرد اصلی در اغلب پژوهش‌ها توسعه روش‌های جستجوی داده در ژئوپرتالهای بوده است. بسیاری از مشکلات مطرح‌شده همانند وجود روش‌های متفاوت جستجو، انتخاب فیلتر مناسب، انتخاب کلیدواژه‌های جستجو، چگونگی پر کردن فرم جستجو و غیره همچنان در ژئوپرتالها وجود دارد.

ترکیب روش‌های جدید جستجو در ژئوپرتالها هم‌زمان با توسعه مفهوم زیرساخت داده مکانی صورت می پذیرد. با توجه به اینکه اکنون، دوره انتقال از مفهوم زیرساخت داده مکانی داده محور و یا محصول مبنا به سمت مفهوم زیرساخت داده مکانی کاربر محور و یا فرایند مبنا می باشد(Rajabifard, Feeney et al. 2002)، نسل جدید ژئوپرتالها نیز نشان‌دهنده افزایش روزافزون علایق به‌سوی تأمین نیازهای کاربران و افزایش تجربه تعاملی آن‌ها با اطلاعات مکانی را خاطر نشان می سازد( van Oort, Kuyper et al. 2009). از اینرو راهکارهایی به منظور

---

۱ SDI(Spatial Information Infrastructure)

۲ Semantic Heterogeneity



www.SID.ir

Archive of SID



جمع‌آوری داده‌های مرتبط با رفتار، ترجیحات و علایق کاربران در ژئوپرتال ها موردنیاز است. در صورتی که چنین داده‌هایی در دسترس باشد می‌توان از طریق شناسایی اطلاعات و سرویس‌های مکانی مطلوب، نتایج موتورهای جستجو را ارتقا بخشید ( Fernández and Castellanos 2006). بسیاری از فروشگاه‌های اینترنتی همانند آمازون[۱]، هنگام خرید و یا دیدن یک مورد در وب‌سایت به کاربران خود آیتم‌هایی را پیشنهاد می‌کنند که ممکن است حتی کاربر از وجود چنین آیتم‌هایی مطلع نباشد. حتی ممکن است آیتم پیشنهادی، آیتم موردعلاقه‌ی کاربر باشد و کاربر نتوانسته کلیدواژه‌ای متناسب برای جستجوی این آیتم به کار ببرد. حتی ممکن است آیتم پیشنهاد آیتمی مرتبط و مکمل آیتم مدنظر کاربر باشد. ون اوورت[۲] و همکاران (۲۰۰۹) به تشریح چگونگی استفاده از مفاهیم نظریه بازاریابی اینترنتی در ژئوپرتالهای می‌پردازد و سیستم‌های توصیه گر به‌عنوان یک رویکرد فروش متقاطع[۳] که درواقع متقاعد کردن مشتری به خرید کالای دوم مکمل کالای اول و یا کالای دوم مرتبط با خرید اول معرفی می‌کند . پیشنهاد داده‌ها و سرویس‌های مکانی در ژئوپرتالهای، می‌تواند تلنگری به کاربران در مورد داده‌های و سرویس‌های موردعلاقه‌ی آن‌ها باشد و به‌نوعی بازاریابی برای فراهم آورندگان داده‌ها و سرویس‌های مکانی است.

در این پژوهش سعی بر ارتقای عملکرد جستجوگری در ژئوپرتالها از طریق افزودن یک مکانیسم جدید بازیابی اطلاعات در ژئوپرتال است. سیستم‌های توصیه گر روش نوینی در کشف و ارائه اطلاعات به کاربران هستند که دارای روالی مجزا از مکانیسم جستجوگری موجود در ژئوپرتالها می باشند، از این لحاظ که علاوه بر نتایج جستجوی ارائه‌شده توسط مؤلفه پرسشگر سیستم، پانلی جداگانه از اقلام اطلاعاتی به کاربر ارائه می‌شود. این پانل درواقع می‌تواند نقش تبلیغاتی در مورد داده‌ها سرویس‌های مکانی داشته باشد. در این پژوهش که به پیاده‌سازی یک ژئوپرتال توصیه گر در حوزه راهداری و حمل‌ونقل جاده‌ای خواهد. در سیستم توصیه گر مد نظر پژوهش حاضر از یکی از ابزارهای وب معنایی به نام هستی‌شناسی[۴] استفاده خواهد شد. سیستم توصیه گر مدنظر با استفاده از ثبت تعاملات کاربر با سیستم و استخراج مشابهت معنایی آیتم‌های اطلاعاتی و کاربران، بر مبنای الگوریتم تجزیه مقادیر تکین (SVD) به پیشنهاد آیتم‌های اطلاعاتی به کاربر می‌پردازند. هدف اصلی پژوهش حاضر طراحی و پیاده سازی ژئوپرتالی در حوزه راهداری مبتنی بر سیستم توصیه گر با استفاده از مشابهت هستی شناسی است. در سیستم‌های توصیه گر همانند سیستم مدنظر پژوهش حاضر، که به‌تازگی در یک حوزه خاص همانند راهداری پیاده‌سازی می‌شوند مشکلی به نام مسئله شروع سرد به وجود می‌آید. در کنار هدف اصلی پژوهش، یکی از اهداف جزئی پژوهش به کارگیری ابزار هستی شناسی برای حل مشکل شروع سرد در سیستم های توصیه گر است.

## ۲- مبانی نظری

در این قسمت به بررسی مفاهیم نظری مورداستفاده در پژوهش حاضر پرداخته می‌شود:

### ۲-۱- سیستم های توصیه گر

سیستم‌های توصیه‌گر سیستم‌هایی هستند که به جمع‌آوری اطلاعات در مورد علائق، ترجیحات، زمینه‌های کاری و غیره‌ی کاربران نسبت به اقلام اطلاعاتی[۵] به دو شکل آشکار (از طریق امتیازدهی کاربران) و یا نهان (از طریق پایش رفتارهای کاربران از قبیل تعداد دفعات گوش دادن به یک موسیقی، مرور صفحات وب، دانلود یک نرم‌افزار، مطالعه کتاب و ...)، می‌پردازند( Choi, Yoo et al. 2012). علاوه براین این سیستم‌ها به جمع‌آوری اطلاعات دموگرافیک کاربران همانند سن، جنسیت، ملیت، و یا اطلاعات مربوط به شبکه‌های اجتماعی که کاربر در آن‌ها عضویت دارد، نیز می‌پردازد. این سیستم‌ها قادرند به کاربران در یافتن اقلام اطلاعاتی مدنظرشان، با ارائه اقلام اطلاعاتی پیشنهادی کمک نماید.(Park, Kim et al. 2012). سیستم‌های توصیه‌گر، فناوری شخصی‌سازی فیلترینگ اطلاعات هستند که به پیش‌بینی احتمال پسندیدن یک شئ اطلاعاتی خاص توسط کاربری خاص و شناسایی N شئ

۱ Amamzon.com
۲ Van Oort
۳ Cross-Selling Approach

٤ Ontology
٥ Items



www.SID.ir

Archive of SID



موردعلاقه کاربر می‌پردازد(Bigdeli and Bahmani 2008). سیستم‌های توصیه گر به دودسته کلی سیستم‌های فیلترینگ مشارکتی[۱] و سیستم‌های محتوی محور[۲] تقسیم‌بندی می‌شوند(Burke 2002Pazzani 1999 ;).

### ۲-۱-۱- سیستم‌های توصیه گر مشارکتی

این سیستم‌ها بر اساس رفتار کاربران در گذشته به توصیه‌گری می‌پردازند و بر اساس این فرض زندگی عادی بنانهاده شده‌اند که هر فرد در زندگی عادی خود از گفته‌ها و تجربیات نزدیکان خود استفاده می‌کند. این دسته از سیستم‌ها به دودسته حافظه مبنا[۳] و مدل مبنا[۴] تقسیم می‌شوند. در روش‌های حافظه مبنا صرفاً از ماتریس امتیازات کاربران و اقلام آیتم ها استفاده می‌شود و از تمامی امتیازدهی‌های کاربران در هرزمانی استفاده می‌شود. جنس و نوع اقلام پیشنهادی با توجه با تغییرات علائق و زمینه‌های کاری کاربران در بستر زمان دائماً در حال تغییر است. به‌عبارتی‌دیگر این سیستم‌ها زمانمند[۵] بوده و اغلب از سنجه‌های مشابهت برای سنجش میزان نزدیکی کاربران و اقلام اطلاعاتی استفاده می‌کنند(Adomavicius and Tuzhilin 2005) در سیستم‌های مدل مبنا از پایگاه داده امتیازدهی کاربران برای ساخت مدل استفاده می‌شود و این داده‌های خام به‌صورت آفلاین پیش‌پردازش شده و سپس مدل کاربر ساخته می‌شود. پس از اتمام ساخت مدل، امتیازدهی‌های اخیر کاربران در نظر گرفته نمی‌شود. بر مبنای مدل ساخته‌شده سیستم به توصیه گری آیتم‌های اطلاعاتی می‌پردازد. اگر از هر کاربری اطلاعات جدید به دست آید مدل منقضی[۶] خواهد شد(Su and Khoshgoftaar 2009). دو مسئله مهم در سیستم‌های مشارکتی مسئله شروع سرد[۷] و تفرق داده‌ای[۸] است. شروع سرد در هنگام راه‌اندازی یک سیستم جدید توصیه گر اتفاق می‌افتد. در این حالت اطلاعاتی از امتیازدهی کاربران در مورد آیتم‌های اطلاعاتی وجود ندارد و به بیان‌دیگر اطلاعاتی در مورد نحوه تعامل کاربران و آیتم‌های اطلاعاتی در دست نیست. این مشکل معمولاً در سه سطح اتفاق می‌افتد: در سطح جامعه کاربران، در سطح کاربر جدید و در سطح آیتم جدید (Schafer, Frankowski et al. 2007). مسئله تفرق داده‌ای به این دلیل ایجاد می‌شود که در اکثر مواقع کاربران تنها به آیتم‌های خاصی علاقه دارند و صرفاً همان آیتم‌ها رتبه‌بندی می‌کنند بنابراین در هر بردار ماتریس کاربران و آیتم‌های اطلاعاتی تعداد کثیری از درایه‌ها دارای مقدار null می‌باشد(Huang, Chen et al. 2004) . بسیاری از محققان برای حل این مسئله از تکنیک‌های کاهش بعد[۹] که عمدتاً بر پایه فاکتورگیری ماتریس[۱۰] است، استفاده کرده اند (Sarwar, Karypis et al. 2000) .

### ۲-۱-۲- تکنیک‌های کاهش بعد

هدف این تکنیک‌ها کاهش مقدار داده‌های مرتبط هم‌زمان با حفظ محتوای اطلاعاتی عمده موجود در آن‌ها است. این تکنیک‌ها اغلب درزمینه هایی مثل داده‌کاوی، یادگیری ماشین و تحلیل خوشه‌ای کاربرد دارند. اکثر روش‌های کاهش بعد شامل استخراج خصیصه‌ها با استفاده از متغیرهای پنهان در داده‌ها و یا اصطلاحاً latent variables در جهت توصیف علل پنهان هم رویدادی دادها می‌شوند. تکنیک‌های کاهش بعد به‌طور خاص در فیلترینگ مشارکتی (گاهی اوقات فیلترینگ مشارکتی مدل مبنا نامیده می‌شود) کاربرد دارند.

SVD یکی از روشهای پرکاربرد در زمینه‌ی فاکتورگیری ماتریس است که از ماتریس $A_{n,m}$ به سه ماتریس $A = USV^T$ فاکتورگیری می‌شود. ماتریس‌های $U$ و$V$ دو ماتریس متعامد به ترتیب با ابعاد $n \times y$ و $m \times y$ هستند و $y$ مرتبه ماتریس A می باشد. S ماتریس قطری با ابعاد $y \times y$ است که شامل تمام مقادیر تکین ماتریس A است. در این حالت امکان کاهش ابعاد ماتریس $S_y$ ، که تنها شامل k بزرگترین مقادیر قطری ماتریس $S_y$ باشد و از آن ماتریس $S_k$ به دست آید، وجود دارد و k<y خواهد بود (Polat and Du 2005). سرور و همکاران (۲۰۰۰) الگوریتم مشارکتی مبتنی بر SVD ارائه دادند. مقادیر null ماتریس متفرق کاربر-آیتم (A) با استفاده از

۱ Collaborative Filtering
۲ Content Based
۳ Memory-Based
٤ Model-Based
٥ Time Aware
٦ Outdated
۷ Cold Start Problem
۸ Data Sparsity
۹ Dimensionality Reduction
۱۰ Matrix Factorization



www.SID.ir

Archive of SID



متوسط امتیازات هر کاربر تکمیل می‌شود تا بتواند روابط معنادار پنهان را استخراج نمایند. ماتریس تکمیل شده با استفاده از تبدیل مقادیر به z-score استانداردسازی و نرمال می شود. ماتریس نرمال شده $A_{norm}$ توسط SVD به ماتریسهای U، S و V فاکتورگیری می‌شود. سپس ماتریس $S_k$ با حفظ تنها k مقدار بزرگ یکه حاصل می شود. همزمان ابعاد ماتریسهای U و V نیز کاهش می یابد. سپس ماتریسهای $U_k\sqrt{S_k}$ و $\sqrt{S_k}V_k^T$ محاسبه می شوند. این دو ماتریس را می توان برای پیش بینی مقدار ترجیح کاربر u نسبت به آیتم k را به دست آورد. برای پیش بینی مقدار محبوبیت یک آیتم برای کاربری خاص، ضرب اسکالر ردیف(بردار) $u^{th}$ ماتریس $U_k\sqrt{S_k}$ و ستون(بردار) $q^{th}$ ماتریس $\sqrt{S_k}V_k^T$ محاسبه می شود و مقدار حاصله دنورمالایز[1] می شود:

$$p_{uq} = \overline{v}_u + \sigma_u \left[U_k\sqrt{S_k}(u).\sqrt{S_k}V_k^T(q)\right] \quad (1)$$

مقدار $\overline{\mathbf{v}}_\mathbf{u}$ برابر میانگین و $\boldsymbol{\sigma}_\mathbf{u}$ برابر انحراف معیار امتیازدهی های کاربر u می باشد.

روش‌های کاهش بعد مبتنی بر SVD به‌وسیله ی محققانی از قبیل Billsus و Pazzani و همچنین Sarwar و همکاران و بسیاری از محققان دیگر درزمینه فیلترینگ مشارکتی مورداستفاده قرار گرفت (Ekstrand, Riedl et al. 2011) .

## ۲-۲- سیستم‌های توصیه گر محتوا محور یا شناختی

در این نوع سیستم‌ها برای کاربر یک پروفایل به‌صورت پنهانی، بر اساس محتویات و خصیصه‌های مشترک آیتم‌هایی که توسط کاربر امتیازدهی شده‌اند (Like/Dislike و یا ۵+ تا ۵-)، ساخته می‌شود. سپس بر اساس مشابهت محتوی و خصیصه‌های آیتم‌ها و پروفایل کاربر به وی پیشنهادهایی ارائه می‌گردد.(Borges and Lorena 2010) اخیراً این نوع توصیه گرها به دلیل توسعه فناوری وب۲[2] و امکان ترکیب داده‌های بیشتر به شکل برچسب[3]، پست، ایده و نظرات و محتویات چندرسانه‌ای توسط کاربران در آیتم‌ها، بسیار موردتوجه قرارگرفته‌اند (Bobadilla, Ortega et al. 2013). فرا خصوصی‌سازی[4] از مسائل تحقیقاتی موجود در این زمینه است. به سبب ماهیت این رویکرد که صرفاً برحسب مشابهت خصیصه‌های آیتم‌ها به توصیه‌گری می‌پردازند، سیستم به جستجوی آیتم‌هایی می‌پردازد که شباهت کامل با آیتم مدنظر کاربر دارد، بدین سان نتیجه، پیشنهاد آیتم‌هایی است که کاربر بیشتر اوقات به آن‌ها آشنایی کامل دارد (Blanco-Fernández, Pazos-Arias et al. 2008). در این حالت آیتم‌هایی که موردعلاقه کاربرند ولی کاربر از وجود آن‌ها باخبر نیست به وی پیشنهاد نمی‌شود. مسئله دیگر تحلیل محتوای محدود[5] است که مربوط به استخراج خودکار اطلاعات قابل‌اطمینان از محتوای آیتم‌ها (موزیک، تصویر، ویدئو و ...) است (Adomavicius and Tuzhilin 2005). در حیطه ژئوپرتالها، سیستم‌های توصیه گر محتوا محور و سیستم‌های اشتراکی هر دو دارای مسائل خاصی هستند .زیرا که در فیلترینگ اشتراکی به سبب استفاده از تمایلات اکثر کاربران در مورد آیتم‌ها، توصیه‌های عمومی به کاربران متخصص ژئوپرتالهای ارائه می‌گردد که این متناسب با نیاز کاربران نیست و همچنین به دلیل مسئله فرا خصوصی‌سازی که در سیستم‌های محتوا محور وجود دارد، نمی‌توان اتکای کاملی به این سیستم‌ها داشت و می‌بایست به نحوی به تنوع آیتم‌های پیشنهادی سیستم افزود. این مسئله را می‌توان با استفاده از ابزارهای وب معنایی[6] همانند هستی‌شناسی تا حد زیادی رفع نمود. هنگامی پروفایل آیتم‌های اطلاعاتی و پروفایل کاربران با مجموعه‌ای از خصیصه‌های ثابت و یکسان (همانند استانداردهای فراداده برای داده‌ها و سرویس‌های مکانی) که دارای مقادیر مشخص و پیش فرض باشد توصیف شوند، پروفایل‌ها را می‌توان با یک ساختار نظام‌مند نمایش داد (Pazzani and Billsus 2007). هستی‌شناسی یکی از ابزارهای اصلی جهت نمایش این ساختار است.

## ۲-۳- وب معنایی و هستی‌شناسی

ایده توسعه صفحات وب با اضافه کردن مقدار اطلاعات تکمیلی، به نحوی‌که اطلاعات موجود قابل‌فهم برای ماشین‌ها باشد، برای اولین بار توسط تیم برنرزلی مطرح

---

۱ Denormalize
۲ Web 2
۳ Tag
٤ Overspecialization
٥ Limited Content Analysis
٦ Semantic Web



www.SID.ir

Archive of SID



شد (Berners-Lee, Hendler et al. 2001). وب معناگرا سعی دارد تا اطلاعات موجود در وب را برای ماشین‌ها قابل‌درک کند. در وب‌های معناگرا از روش‌ها و تکنیک‌های مطرح‌شده در هوش مصنوعی مانند پردازش زبان طبیعی، پردازش تصویر، هستی‌شناسی و... استفاده می‌کنند (Maedche 2002).

گرابر[1] در سال ۱۹۹۵ هستی‌شناسی را ذکر خصوصیات واضح و روشن یک مفهوم (ادراک یا تصور) تعریف می‌کند. در واقع هستی‌شناسی نمایش یک فهم مشترک از یک دامنه (حوزه) و واژگان مرتبط آن است. ازلحاظ فنی، هستی‌شناسی‌ها نمایش‌دهنده‌ی مفاهیم و ارتباطات طبقه‌بندی‌شده و طبقه‌بندی نشده آن‌ها است. در وب معنایی، هستی‌شناسی به ما کمک می‌کند تا واژگان مورداستفاده در نمایش و به اشتراک‌گذاری داده‌ها را تعریف کرد (Gruber 1995). اغلب هستی‌شناسی‌ها به‌صورت شبکه معنایی از گره‌های مفهومی مرتبط، متصور و بصری سازی می‌شوند. در حال حاضر OWL زبان غالب برای توصیف و ساخت و نمایش هستی‌شناسی‌های رسمی است (Sugumaran and Gulla 2011).

## ۳- پیشینه پژوهش

سیستم های توصیه گر در حوزه های محتلف علوم اطلاعات مکانی کاربرد دارند اما درزمینه کاربرد سیستم‌های توصیه گر در حوزه ژئوپرتالهای پژوهش‌های بسیار اندکی انجام‌گرفته است. بشه[2](۲۰۱۱) در پژوهشی با عنوان تبدیل موتور جستجوی مکانی به سیستم توصیه گر مکانی با استفاده از مفهوم برازش برای کاربرد[3] یک سیستم توصیه گر جهت تعیین میزان برازش یک داده مکانی برای یک کاربرد خاص طراحی کرد. رویکرد وی عمدتاً بر دیدگاه کاربران نسبت به ابعاد کیفی داده مکانی و توصیفی منابع داده مکانی تأکید داشت (besheh 2011). عدم به‌کارگیری الگوریتم‌های توصیه گر، عدم استفاده از پیشینه رفتاری کاربر برای توصیه اقلام اطلاعاتی، تکیه صرف بر یک شاخص کیفی بر اساس برازش برای کاربرد.

همچنین ووکر[4] و همکاران (۲۰۱۳) در دو مقاله به طراحی سیستم توصیه گری برای ژئوپرتال در حوزه انرژی پرداختند. آن‌ها از الگوریتم های قواعد رابطه‌ای[5] و Slope One که در بسته نرم‌افزاری ایزی رک[6] وجود داشت و جزو توصیه گرهای مشارکتی آیتم مبنا به‌حساب می‌آیند استفاده کردند. تکیه آن‌ها بر استخراج میزان مشابهت فرادادهها بر اساس تکنیک‌های مشابهت معنایی دو فراداده به‌عنوان ورودی اضافی برای بسته نرم‌افزاری ایزی رک است. سپس آن‌ها از سرویس توصیه گری این نرم‌افزار باواسط سرویس REST استفاده کردند. آن‌ها رفتار کاربر را تنها بر اساس سه عمل دیدن، مرور و خرید که در خریدهای برخط در فروشگاه‌های اینترنتی وجود دارد پایش کردند و بر اساس این سه عمل و مشابهت معنایی دو فراداده به توصیه گری می پرداختند. در تحلیل معنایی متون هنگامی‌که طول (تعداد کلمات) متن‌ها (فراداده‌ها) متفاوت باشد ارتباط و مدل‌سازی معنایی متون به‌درستی انجام نخواهد شد. در این سیستم به دلیل عدم وجود پروفایل کاربر امکان شخصی‌سازی توصیه گریها وجود ندارد. همچنین برای آیتم‌های اطلاعاتی نیز پروفایلی وجود ندارد. توصیه گری صرف بر مبنای مقادیر امتیازدهی‌ها انجام می‌شود. ابعاد دیگر داده‌های مکانی در نظر گرفته نشده است.

## ۴- روش تحقیق

پژوهش حاضر طراحی و پیاده‌سازی ژئوپرتالی را مبتنی بر یک سیستم توصیه گر ارائه می‌دهد. سیستم توصیه گر در ژئوپرتال با ارائه آیتم‌های بازیابی شده در پانل جداگانه‌ای، به کاربران در دسترسی به داده‌ها و سرویس‌های مکانی کمک بسزایی خواهد کرد. رویکرد مورداستفاده در طراحی سیستم توصیه گر مدنظر پژوهش، رویکردی ترکیبی مبتنی بر سیستم‌های مشارکتی مدل مبنا و سیستم‌های محتوامحور معناگرا است. در سیستم‌های توصیه گر موجود الگوریتم‌های کاهش بعد همانند SVD بسیار رایج هستند و کارایی آن‌ها به اثبات رسیده است. به دلیل مدل مبنا بودن و تعداد بسیار زیاد کاربران، این نوع الگوریتم‌ها نیاز به منابع سخت‌افزاری بالایی دارند. منابع سخت‌افزاری برای اغلب کاربردها در اختیار نیست بنابراین

[1] Gruber
[2] Besheh
[3] Fitness for Use
[4] Vocker
[5] Association Rules
[6] EasyRec http://easyrec.org/



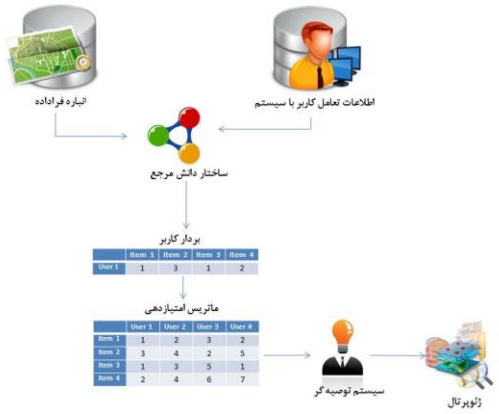
www.SID.ir

Archive of SID



الگوریتم‌های آیتم مبنا باوجود دقت پایین اغلب مورداستفاده هستند؛ اما در مورد ژئوپرتالهای این روال صادق نمی باشد. اغلب به دلیل تخصصی بودن، تعداد کاربران ژئوپرتالهای نسبت به سایر پرتال‌هایی که در حوزه‌های عمومی کاربرد دارند (همانند سایت آمازون)، کمتر است بنابراین نیاز به منابع سخت‌افزاری نیز کمتر و درنتیجه دقت پیش‌بینی‌ها نیز بالاتر خواهد بود؛ بنابراین در پژوهش حاضر نیز از این نوع الگوریتم استفاده‌شده است. در سیستم‌های توصیه گر مشارکتی که به‌تازگی در حوزه کاربردی جدیدی پیاده‌سازی می‌شوند دو مسئله تفرق ماتریس امتیازدهی‌ها و شروع سرد کارایی این سیستم‌های را به‌شدت تحت تأثیر قرار می‌دهند. در این پژوهش برای حل این دو مسئله از یکی از ابزارهای تحلیل معنایی به نام مشابهت معنایی استفاده‌شده است. مشابهت معنایی بر مبنای ساختار هستی‌شناسی مرجع موجود در حوزه کاربردی پژوهش (راهداری و حمل‌ونقل جاده‌ای) استخراج می‌شود. ساختار هستی‌شناسی مرجع در پروفایل کاربر قرار می‌گیرد و مختص آن کاربر خواهد بود. ساختار هستی‌شناسی مختص کاربر تمامی اطلاعات مربوط به تعامل کاربر با سیستم را در خود ذخیره خواهد کرد. هر آیتم جدید که وارد سیستم می‌شود در هستی‌شناسی مرجع وارد می‌شود و هستی‌شناسیهای مختص کاربران نیز برحسب این تغییرات به‌روزرسانی می‌شود. کلیدواژه‌ها و محدوده‌های مکانی مورد جستجوی کاربر، نام و محدوده‌ی مکانی و کلیدواژه‌های مرتبط با آیتم‌های مشاهده‌شده توسط کاربر در پانل نتایج جستجو و پانل آیتم‌های پیشنهادی، اطلاعاتی هستند که در ساختار هستی‌شناسی مختص کاربر ذخیره می‌شوند. اطلاعاتی که در ساختار هستی‌شناسی پس از ثبت‌نام کاربر در سیستم، همگی به‌صورت پنهانی جمع‌آوری می‌شوند. بر مبنای تابع مشابهت معنایی میزان مشابهت آیتم‌های اطلاعاتی با کاربر مدنظر محاسبه می‌شود و در برداری که مختص کاربر است ذخیره می‌شود. این بردار نیز در پروفایل کاربر قرار دارد.

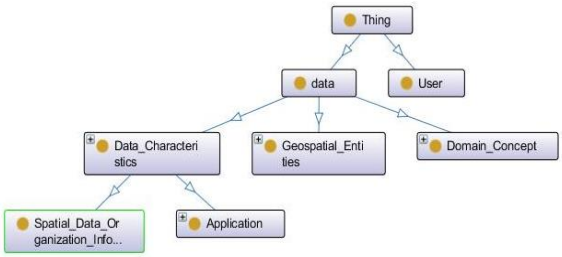
ماتریس امتیازدهی مستقیماً توسط کاربران سیستم مقداردهی می‌شود. با ترکیب بردارهای کاربران در ماتریس امتیازدهی، ماتریس امتیازدهی از حالت متفرق خارج می‌شود و نهایتاً بر مبنای این ماتریس سیستم توصیه گر مشارکتی مبتنی بر الگوریتم SVD، آیتم‌های اطلاعاتی را در ژئوپرتال به کاربر ارائه می‌نماید.




شکل۱- فلوچارت روش تحقیق در پژوهش حاضر (مأخذ: نگارنده)

## ۴-۱- ساختار هستی‌شناسی مرجع

این ساختار هستی‌شناسی مرجع، توسط نرم‌افزار Porotege بر اساس دو هستی‌شناسی پایه ( GDN(Geographic Data Files و OTN(Ontology of Transportation Network) و همچنین حذف و اضافه نمودن برخی مفاهیم مرتبط و غیر مرتبط در این هستی‌شناسی ، طراحی گردید. رده‌های کلی این هستی‌شناسی به شرح ذیل است:




شکل۲- کلاس‌های اصلی هستی‌شناسی ایجادشده (مأخذ: نگارنده)

الف) کلاس User: این کلاس شامل تمامی کاربران سیستم است.

ب) کلاس data: این کلاس شامل دو زیر کلاس اقلام اطلاعاتی و خصیصه‌های آن‌ها و مفاهیم دامنه مدنظر است.

ج) کلاس Geospatial_Entities: این کلاس شامل تمامی آیتم‌های اطلاعاتی است. هر آیتم جدیدی که وارد سیستم می‌شود به این کلاس نگاشت می‌شود. زیر کلاس‌ها و نمونه‌های این کلاس درواقع مدل انتزاعی از پایگاه داده مکانی سازمان راهداری هستند.

د) کلاس Data_Characterisitcs: این کلاس نشان‌گر خصیصه‌های آیتم‌های اطلاعاتی است.

ه) کلاس Domain_Concepts: این کلاس بیان‌کننده کلیه مفاهیمی مورداستفاده برای داده‌های حوزه مورد مطالعه است. از طرف دیگِ این کلاس خود نوعی داده



www.SID.ir

Archive of SID



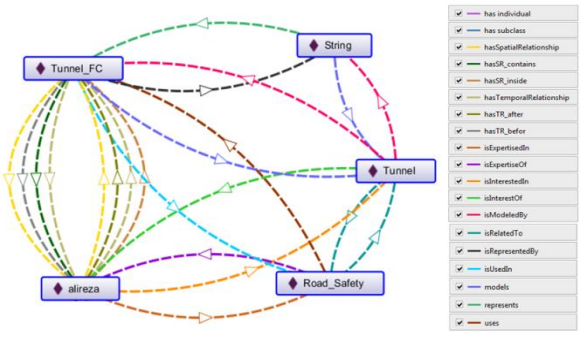
است که داده‌های دیگر را توصیف می‌کند بنابراین زیرکلاسی از کلاس data است. بیشتر مفاهیم این کلاس از هستی‌شناسی های پایه استخراج شد.

ی) کلاس Application: این کلاس بیان‌کننده کلیه کاربردهایی است که از یک داده مکانی در حوزه مدنظر مورد انتظار است. همچنین این کلاس به‌عنوان تخصص‌های کاری کاربران نیز مطرح است. برای تولید زیر کلاس‌ها و نمونه‌های این کلاس از منابع مختلفی استفاده شد. مصاحبه‌های انجام‌گرفته با کارشناسان سازمان راهداری و اسناد مختلف موجود در سازمان ازجمله این منابع بودند.

ر)کلاس Spatial_Data_Organization_Information: کلاسی است که بیان‌کننده نوع نمایش داده‌های مکانی است. نوع نمایش یک داده مکانی می‌تواند کاربرد این داده در حوزه خاصی را بیان نماید. برای ایجاد زیرکلاسها و نمونه‌های این کلاس از استاندارد فراداده FGDC استفاده شد است.

مرحله دوم طراحی هستی‌شناسی تعریف روابط بین کلاس‌ها و نمونه‌هاست. روابط تعریف‌شده بین کلاس‌ها روابط مکانی و زمانی و توصیفی بین آیتم‌های اطلاعاتی و کاربران را بیان می‌نماید. روابط مکانی مورداستفاده همان روابط RCC8 در علوم اطلاعات مکانی است.

به دلیل این که تعریف روابط بین کلاسها و اختصاص نمونه ها به آنها، سبب بازتعریف روابط بین کلاسی در بین نمونه ها نمی شود می بایست پس از تعریف روابط اصلی بین کلاس‌ها، این روابط بین نمونه های کلاسها نیز برقرار شود. با تکیه‌بر مطالعات انجام‌شده در سازمان راهداری، این روابط بین نمونه‌های کلاس‌های اصلی هستی‌شناسی نیز اختصاص داده شد (شکل ۳).




شکل۳- نمونه‌ای از روابط تعریف‌شده بین نمونه‌های کلاس‌های هستی‌شناسی (مأخذ: نگارنده)

## ۴-۲- استخراج محدوده مکانی و زمانی کاربر

در سیستم طراحی شده تعاملات کاربر با سیستم ثبت می گردد. دو نوع از اطلاعاتی که در حین تعامل کاربر با سیستم ثبت می شوند محدوده های مکانی و بازه های زمانی آیتم های اطلاعاتی است. در هر جلسه، کاربر ممکن است بر روی چندین آیتم های اطلاعاتی توصیه شده و یا جستجو شده کلیک کند و به بررسی و واکاوی آیتم اطلاعاتی بپردازد. بنابراین می بایست با توجه به خصیصه های زمانی و مکانی آیتم های که کاربر با آنها در تعامل بوده است، برای هر جلسه کاربر یک محدوده مکانی و بازه زمانی موثر استخراج می شود. این محدوده مکانی و بازه زمانی موثر در پروفایل کاربر ثبت می شود. محدوده ی مکانی و بازه زمانی که برای کاربر استخراج می شود بر مبنای بازه های زمانی و محدوده های مکانی مستخرج از ده جلسه پیشین کاربر است. البته پیش فرض سیستم ۱۰ جلسه است که مطابق با نیاز کاربر می تواند کاهش یا افزایش یابد. هنگامی که این کابر از سیستم لاگ اوت می شود سیستم به پردازش خصیصه های زمانی و مکانی ثبت شده می پردازد. این پردازش در دو مرحله انجام می گیرد. مرحله اول شامل استخراج محدوده مکانی و بازه زمانی مختص جلسه اتمام یافته می شود. مرحله دوم نیز شامل استخراج روابط مکانی و زمانی کاربر با آیتم های اطلاعاتی با توجه به ده بازه زمانی و مکانی موثر است.

مرحله اول با توجه به تابع ارائه شده توسط گومز[۱] و همکاران (۲۰۱۴) انجام می گیرد. در واقع این تابع، تابع تطبیق یافته ی تابع ارائه شده توسط تراورسکی[۲](۱۹۷۷) برای ارزیابی مشابهت بین دو ماهیت است. درجه اشتراک محدوده های مکانی A و B به صورت رابطه ۳-۷ محاسبه می شود.

$$Overlap(A,B) = \frac{area(A \cap B)}{area(A \cap B) + \propto * area(A-B) + (1-\propto) * area(B-A)}$$

به دلیل اینکه درجه اشتراک نامتقارن است و سطح مکمل سطح جستجو نشان دهنده ی مشابهت کمتر محدوده های مکانی است بنابراین مقدار $\propto = ۰.۸۴$ است. مقدار این ضریب به وسیله ی روش وزندهی استخراج شده است(Fox & Shaw, 1994).

همچنین گومز و همکاران برای محاسبه درجه اشتراک بازه های زمانی t و T رابطه ۳-۸ را پیشنهاد کردند.

$$Overlap(T,t) = \frac{duration(t \cap T)}{duration(t \cap T) + \propto * duration(t-T) + (1-\propto) * duration(T-t)}$$

۱ Gomes
۲ Tversky



www.SID.ir

Archive of SID



این دو تابع به ازای تمامی حالتهای غیرتکراری در دو آرایه محدوده های مکانی و بازه های زمانی اجرا می شوند تا از اجتماع تمامی محدوده هایی که بیش از ۸۰ درصد مشابهت مکانی یا زمانی دارند، محدوده مکانی و بازه زمانی کاربر استخراج شود.

## ۴-۳- مشابهت معنایی

مشابهت معنایی در پژوهش حاضر نوعی از مشابهت سنجی در سطح هستی‌شناسی است. این مشابهت سنجی درواقع در محدوده نمونه‌های هر کلاس انجام می‌گیرد. تابع مشابهت هستی‌شناسی مدنظر بر پایه تابعی است که والتر و همکاران (۲۰۱۲) ارائه دادند. در این پژوهش که تابع مشابهت معنایی به‌صورت رابطه ۲ تعریف‌شده است.

$$S_{ab} = Similarity(a,b)\sum_{i=1}^{\#p}\left(\frac{common(a,b,P[i])}{\max(deg\ (a,P[i]),deg\ (b,P[i]))}\right) * Weight(P[i]) \quad (۲)$$

در این تابع $S_{ab}$ میزان مشابهت دو نمونه a و b است.

P (Propertices): برداری است که شامل تمامی خصیصه‌ها (خصیصه‌های داده‌های) و روابطی (خصیصه‌های کلاسی) است که برای کلاس مدنظر تعریف‌شده است. برای مثال:

P=[isInterestedIn, isExpertisedIn], P#=2

deg (a, p): نشان‌دهنده‌ی تعداد نمونه‌هایی است که از طریق رابطه p با نمونه a در ارتباط است.

UserNew.isInterestedIn=[Tunnel,Public_Transport]
deg(isInterestedIn, UserNew)=2

common (a,b,p): نشان‌دهنده‌ی تعداد نمونه‌هایی است که از طریق رابطه p هم‌زمان با a و b در ارتباط هستند.

MainRoads_FC.isDescribedBy=[Route,Road, Network ,Traffic,Traffic_Restriction]
Liniation_FC.isDescribedBy=[Traffic_Restriction, Road_Sign,Road,Road_furniture]
Common(MainRoads_FC,Liniation_FC, isDescribedBy)=2

Weight(P): نشان‌دهنده‌ی درجه اهمیت هر رابطه است.

Weight (isDescribedBy) =0.2

جدول۱- روابط تعریف‌شده بین کلاس‌ها (مأخذ: نگارنده)

| | نام رابطه | نوع رابطه | کلاس مبدأ | کلاس مقصد | معکوس |
|---|---|---|---|---|---|
| روابط توصیفی | isDescribedBy | | Geospatial_Entities | Domain_Concepts | describes |
| | Describes | | Domain_Concepts | Geospatial_Entities | isDescribedBy |
| | isExpertisedIn | | User | Application | isExpertisedIn |
| | isExpertiseOf | | Application | User | isExpertiseOf |
| | isInterestedIn | | User | Domain_Concepts | isInterestOf |
| | isInterestOf | | Domain_Concepts | User | isInterestedIn |
| | isModeledBy | | Domain_Concepts | Spatial_Data_Organization_Information | models |
| | models | | Spatial_Data_Organization_Information | Domain_Concepts | isModeledBy |
| | isRelatedTo | Symetric(متقارن) | Application | Domain_Concepts | isRelatedTo |
| | isRepresentedBy | | Geospatial_Entities | Spatial_Data_Organization_Information | represents |
| | represents | | Spatial_Data_Organization_Information | Geospatial_Entities | isRepresentedBy |
| | isUsedIn | | Geospatial_Entities | Application | uses |
| | uses | | Application | Geospatial_Entities | isUsedIn |
| روابط مکانی | hasSR_contains | | Geospatial_Entities | User | hasSR_inside |
| | hasSR_inside | | Geospatial_Entities | User | hasSR_contains |
| | hasSR_coveredBy | | Geospatial_Entities | User | hasSR_covers |
| | hasSR_covers | | Geospatial_Entities | User | hasSR_coveredBy |
| | hasSR_disjoint | Symetric(متقارن) | Geospatial_Entities | User | hasSR_disjoint |
| | hasSR_equlas | Symetric(متقارن) | Geospatial_Entities | User | hasSR_equlas |
| | hasSR_meets | Symetric(متقارن) | Geospatial_Entities | User | hasSR_meets |
| | hasSR_overlaps | Symetric(متقارن) | Geospatial_Entities | User | hasSR_overlaps |
| روابط زمانی | hasTR_after | | Geospatial_Entities | User | hasTR_befor |
| | hasTR_befor | | Geospatial_Entities | User | hasTR_after |
| | hasTR_covers | | Geospatial_Entities | User | hasTR_during |
| | hasTR_during | | Geospatial_Entities | User | hasTR_covers |
| | hasTR_equals | | Geospatial_Entities | User | hasTR_equals |
| | hasTR_finishes | | Geospatial_Entities | User | hasTR_starts |
| | hasTR_starts | | Geospatial_Entities | User | hasTR_finishes |
| | hasTR_meets | | Geospatial_Entities | User | hasTR_met |
| | hasTR_met | | Geospatial_Entities | User | hasTR_meets |
| | hasTR_overlapped | | Geospatial_Entities | User | hasTR_overlaps |
| | hasTR_overlaps | | Geospatial_Entities | User | hasTR_overlapped |



www.SID.ir

Archive of SID



تابع مذکور توسط والتر برای بررسی مشابهت دو ماهیت یکسان (همانند یک فیلم با فیلمی دیگر و یا یک داده مکانی با داده‌ای مکانی دیگر) کاربرد دارد ولی در پژوهش حاضر میزان مشابهت و یا به‌عبارت‌دیگر مجاورت معنایی دو ماهیت غیر یکسان کاربر با آیتم اطلاعاتی مدنظر است. از طرف دیگر در این تابع تنها نمونه‌هایی که به‌صورت مستقیم با یک رابطه P با نمونه a مرتبط‌اند مدنظر هستند اما در ساختار هستی‌شناسی RMTO_Onto ارتباط بین ماهیت‌های کاربر و آیتم اطلاعاتی ممکن است از طریق یک تا چندین رابطه متوالی برقرار شود؛ بنابراین تنها با بررسی یک رابطه نمی‌توان به بررسی مشابهت معنایی پرداخت.

اگر $\mathcal{C}$ و $\mathcal{P}$ برابر تمامی کلاس‌ها و روابط تعریف‌شده در هستی‌شناسی باشند، یک توالی روابط $PS$ نشان‌دهنده تعداد متناهی از روابط $[P_1, \dots, P_n]$ است که تعدادی از کلاس‌های یک هستی‌شناسی را به یکدیگر مرتبط می‌کند(رابطه ۳).

$$\boldsymbol{PS} = \left\{ \frac{[\boldsymbol{P_1}, \dots, \boldsymbol{P_n}]}{\boldsymbol{P_j \epsilon \mathcal{P}}} \forall\, \mathbf{1} \le \mathbf{J} \le \mathbf{N}, \mathbf{range}(\boldsymbol{P_i}) = \boldsymbol{domain}(\boldsymbol{P_{i+1}}) \forall\, \mathbf{1} \le \boldsymbol{i} < \boldsymbol{N} \right\} \quad (۳)$$

برای مثال توالی روابط $PS = [isExpertisedIn, isRelatedTo, describes]$ کلاسهای $[User, Application, Domain_Concepts, Gospatial_Entites]$ را به هم دیگر مرتبط می‌کند.

یک نمونه از توالی روابط $PS$ که با $ps$ مشخص می‌شود مجموعه‌ای از روابط است که نمونه‌های موجود در توالی روابط $PS$ را به یکدیگر مرتبط می‌کند. از عبارت $y \xrightarrow{owl:typeOf} Y$ برای بیان اینکه $y$ نمونه‌ای از نوع $Y$ است و $Y$ نیز کلاس یا رابطه‌ای تعریف‌شده در هستی‌شناسی است استفاده می‌شود.

$$\boldsymbol{ps} = \{[\boldsymbol{p_1}, \dots, \boldsymbol{p_n}]/\boldsymbol{p_j} \xrightarrow{\boldsymbol{owl:typeOf}} \forall\, \mathbf{1} \le \boldsymbol{j} \le \boldsymbol{N} \quad (۴)$$

نمونه‌ای از مثال توالی روابط ذکرشده می‌تواند به شکل زیر باشد:

$$\boldsymbol{ps} = [\boldsymbol{alireza}, \boldsymbol{Road\ Saftey}, \boldsymbol{Tunnel}, \boldsymbol{Tunnel_FC}]$$

در ساختار شبکه‌ای هستی‌شناسی با استفاده از نمونه‌ی توالی رابطه‌ای مذکور، می‌توان از موجودیت کاربر به یک آیتم اطلاعاتی مسیریابی کرد. این توالی‌ها را می‌توان توالی‌های از پیش تعریف‌شده در نظر گرفت. هرکدام از این توالی‌ها درواقع یک مسیر (پلی لاین) از پیش تعیین‌شده از یک گره ابتدایی کاربر به یک گره انتهایی آیتم اطلاعاتی هستند. لبه‌های[۱] ( همانند isExpertisedIn، isRelatedTo، و غیره ) این مسیر ثابت است اما گره های این مسیر می‌تواند متغیر باشد. لبه ها در واقع همان روابط بین کلاسهاست. گره های مسیر نیز در واقع نمونه های کلاسها هستند. بنابراین نمونه‌های متفاوتی از هر توالی رابطه‌ای یا مسیرکاربر-آیتم می‌تواند وجود داشته باشد.

مسیرها یا توالی‌های از پیش تعریف‌شده برای مسیریابی از کاربر به یک آیتم در ساختار گراف هستی‌شناسی به شرح ذیل هستند. این مسیرها برای تمای کاربران و آیتم های می تواند موجود باشد. گرههای ابتدایی و انتهایی کاربران و آیتم ها و گرههای وسطی نیز نمونه های کلاسها هستند. در واقع در ساختار هستان شناسی این مسیرها، توالی های از پیش تعریف شده ای هستند و در سیستم مدنظر به ازای هر کاربر یک کپی از ساختار اصلی در پروفایل کاربر ایجاد می شود.

{isExpertisedIn, uses},
{isExpertisedIn, isRelatedTo,describes},
{isExpertisedIn, isRelatedTo,isModeledBy, represents},
{isInterestedIn, describes },
{isInterestedIn, isRelatedTo, uses },
{isInterestedIn, isModeledBy,represents },
{hasSR_contains},{hasSR_inside},{hasSR_coveredBy},{hasSR_covers},{hasSR_disjoint}{,{hasSR_equlas},{hasSR_meets},{hasSR_overlaps}
{hasTR_after},{hasTR_befor},{ hasTR_covers},{ hasTR_during},{ hasTR_equals},{ hasTR_finishes},{ hasTR_meets},{ hasTR_met},{hasTR_overlapped},{ hasTR_overlaps},{hasTR_starts}

با استفاده از این مسیرهای پیش‌فرض می‌توان میزان مشابهت معنایی یک کاربر u با یک آیتم اطلاعاتی d را به شرح رابطه ۵ تعریف نمود.

$$S_{ab} = \text{Similarity}(u, d) = \sum_{i=1}^{\#ps} \left( \frac{\text{common}(u, d, PS[i])}{\max(\deg(u, PS[i]), \deg(d, PS[i]))} \right) * \text{Weight}(PS[i]) \quad (۵)$$

[۱] Edge



www.SID.ir

Archive of SID



با فرض اینکه برای هر کاربر یک ساختار هستی‌شناسی ایجاد می‌گردد بنابراین در هر هستی‌شناسی تنها یک نمونه کاربر وجود خواهد داشت در این حالت:< deg (u, PS[i]) deg (d, PS[i]) خواهد بود بنابراین تابع بالا را به شرح رابطه ۶ می‌توان خلاصه‌سازی نمود:

$$S_{ab} = \text{Similarity}(u, d) = \sum_{i=1}^{\#ps} \left(\frac{\text{common}(u, d, PS[i])}{\deg(u, PS[i])}\right) * \text{Weight}(PS[i]) \quad (۶)$$

همچنین تمامی روابط تعریف‌شده بین کلاس‌ها دارای رابطه‌ی معکوس هستند و یا از نوع متقارن هستند.

این امر سبب ایجاد مسیرهای جدید کاربر-آیتم می شود که نهایتا سبب افزایش دقت توصیه گری سیستم خواهد شد.

### ۴-۳- پیاده‌سازی سیستم

معماری مد نظر برای پیاده سازی سیستم معماری لایه ای است. معماری پیاده سازی شده معماری سه لایه است. لایه اول سیستم لایه داده است. در این لایه داده های مربوط به پرفایل کاربر در پایگاه داده کاربران و پروفایل آیتم های اطلاعاتی در پایگاه داده آیتم های اطلاعاتی نگهداری می شود. همچنین انباره هستان شناسی نیز در این لایه قرار دارد. لایه دوم، لایه کاربرد است که هسته اصلی سیستم پیاده سازی شده را تشکیل می دهد. این لایه بر روی سرور وب اجرا می شود. این لایه درخواستهای HTTP رسیده از لایه فوقانی را دریافت و پردازش لازمه را انجام داده و سپس پاسخ را به لایه فوقانی ارسال می نماید.

لایه سوم، لایه واسط ارتباطی و تعاملی کاربران با سیستم است و تقاضاهای کاربران را به سمت لایه کاربرد ارسال می نماید. این لایه پاسخ های ارسالی از لایه کاربرد که در فرمت JSON[۱] را به کاربر نمایش می دهد. تعامل این لایه با لایه زیرین با استفاده از فناوری AJAX[۲] انجام می گیرد. همچنین این لایه به نمایش داده های مکانی، نمایش لیست آیتم های پیشنهادی، نمایش لیست و محدوده ی مکانی آیتم های اطلاعاتی یافت شده در فرایند جستجو، نمایش فراداده می پردازد.

جهت پیاده‌سازی سیستم مبتنی بر وب از تکنولوژی Servlet 3 مبتنی بر زبان برنامه‌نویسی جاوا استفاده شد. این پیاده‌سازی در قالب سه لایه نمایش، کاربرد و داده انجام‌شده است. در لایه نمایش از کتابخانه‌های مختلف جاوا اسکریپتی استفاده‌شده است. همچنین ارتباط این لایه با لایه زیرین بر اساس تکنولوژی Ajax[۳] صورت می‌گیرد. در لایه کاربرد نیز از کتابخانه‌هایی ذیل مبتنی بر زبان جاوا جهت پیاده‌سازی سیستم استفاده‌شده است

Log4J[۴]: در این پژوهش این ابزار برای ثبت تعاملات کاربر با سیستم مورداستفاده قرار می‌گیرد. اطلاعات مستقیماً توسط این واسط وارد پایگاه داده می‌شوند.

GeoTools[۵]: در این پژوهش از این ابزار برای استخراج محدوده‌ی مکانی کاربر و روابط مکانی بین آیتم‌های اطلاعاتی و کاربران استفاده‌شده است.

Apache Jena[۶]: چهارچوبی متن‌باز و رایگان برای طراحی و ایجاد کاربردهای وب معنایی است. این ابزار دارای کتابخانه بسیاری برای انجام تحلیل‌های وب معنایی و کار با ساختارهای RDF و OWL است. در این پژوهش از این ابزار برای پیاده‌سازی تابع مشابهت معنایی کاربران و آیتم‌های اطلاعاتی و همچنین کار با هستی‌شناسی‌ها مورداستفاده قرارگرفته است.

Apache Mahout[۷]: هدف آن ایجاد محیطی برای توسعه سریع کاربردهای یادگیری ماشین و الگوریتم‌های مقیاس‌پذیر در سه حوزه‌ی خوشه‌بندی، طبقه‌بندی و توصیه گری است. از این ابزار برای پیاده‌سازی الگویتم توصیه گری استفاده‌شده است.

۱ JavaScript Object Notation
۲ Asynchronous JavaScript and XML

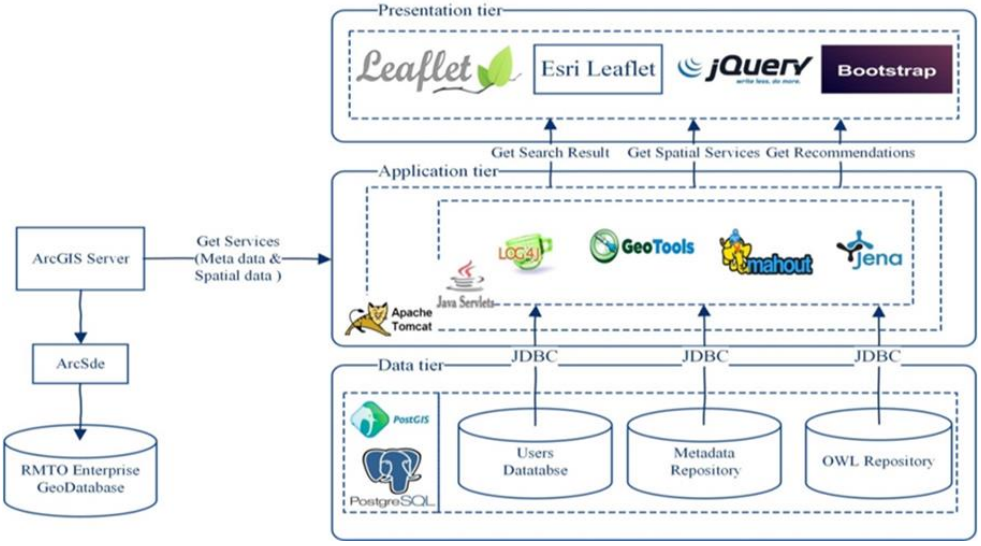
۳ Asynchronous JavaScript and XML
٤ http://logging.apache.org/log4j/2.x/
٥ http://www.geotools.org/
٦ https://jena.apache.org/
٧ http://mahout.apache.org/



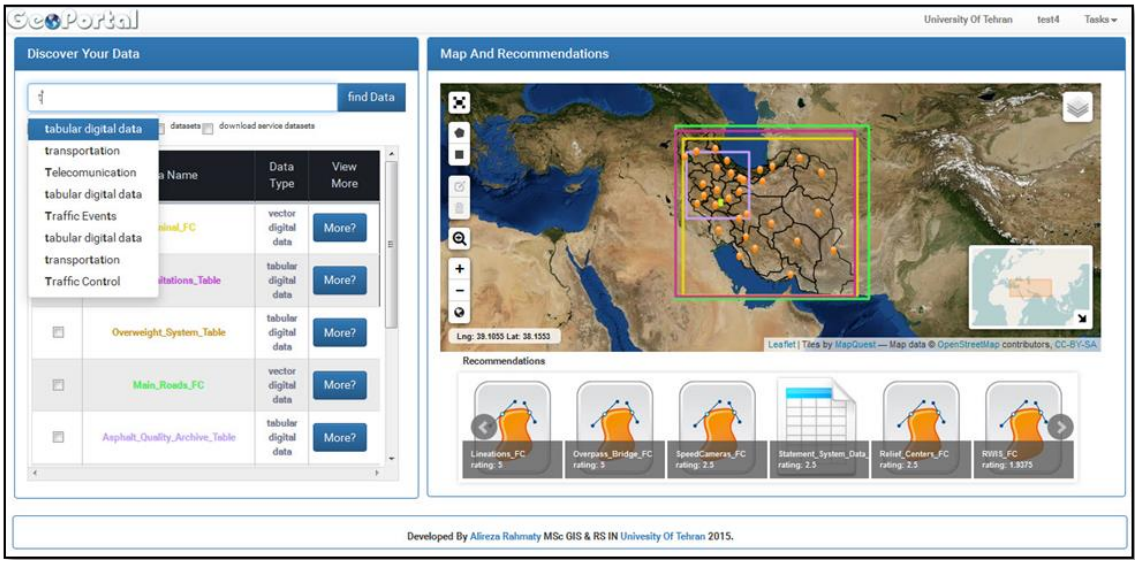
www.SID.ir

Archive of SID






شکل۴- معماری کاربرد سیستم پیاده‌سازی شده (مأخذ: نگارنده)




شکل۵- واسط گرافیکی کاربر سیستم پیاده‌سازی شده

## ۴- ارزیابی سیستم

ارزیابی کمی دقت و صحت توصیه گری سیستم پیاده سازی شده به دلیل اینکه داده‌های استاندارد امتیازدهی‌ها (پایگاه داده ماتریس امتیازات) در مورد آیتم‌های اطلاعاتی مشابه با آیتم‌های اطلاعاتی حوزه کاربردی پژوهش حاضر(راهداری و حمل‌ونقل جاده‌ای) وجود ندارد امکان پذیر نیست. تنها داده‌های استاندارد در این ارزیابی سیستم‌های توصیه گر داده‌های مربوط به فیلم و موزیک و یا سایر انواع آیتم‌های غیر مرتبط می‌باشد. حتی در دو پژوهش انجام‌یافته بسیار مشابه در مورد ژئوپرتالهای توصیه گر نیز به ارزیابی صحت و دقت آیتم های پیشنهادی پرداخته نشده است.

## ۵- نتیجه‌گیری

سیستم‌های توصیه گر در علوم اطلاعات مکانی در حوزه‌های مختلفی کاربرد دارند. تا به امروز مطالعات بسیار کمی درزمینه کاربرد سیستم‌های توصیه گر درزمینه ژئوپرتال ها انجام‌شده است. در این پژوهش برخلاف مطالعات دیگر درزمینه سیستم‌های توصیه گر معناگرا که به استخراج مشابهت‌ها کاربر با کاربر و یا آیتم با آیتم پرداخته‌اند در این پژوهش با تکیه ساختار هستی‌شناسی ایجادشده به محاسبه مشابهت آیتم اطلاعاتی با کاربر پرداخته‌شده است. در پژوهش‌های انجام‌شده درزمینه کاربرد سیستم‌های توصیه گر در ژئوپرتال ها، در پیاده‌سازی سیستم از بسته نرم‌افزاری جداگانه‌ای استفاده‌شده است و



www.SID.ir

Archive of SID



توصیه گری به‌صورت یک سرویس به ژئوپرتال ارائه می‌گردد؛ اما در پژوهش حاضر یک موتور توصیه گر جدای از معماری ژئوپرتال دیده نشده است و در معماری ژئوپرتال ترکیب‌شده است. در این ترکیب مؤلفه‌ی پروفایل آیتم یک سیستم توصیه گر در انباره داده قرارگرفته است. در پژوهش حاضر برخلاف پژوهش‌های انجام‌شده درزمینه ژئوپرتال توصیه گر که توصیه گری ها صرفاً بر مبنای سه عمل دیدن، مرور کردن و یا دانلود کردن بود در این پژوهش رفتارهای بیشتر و متنوع‌تری از کاربر ثبت می‌شود. این رفتارها شامل ثبت کلیدواژه‌های مورد جستجوی کاربر، ثبت آیتم‌های اطلاعاتی موردعلاقه‌ی کاربر، ثبت محدوده‌های مکانی آیتم‌های اطاعاتی برای کاربر و استخراج سطح مکانی مؤثر برای کاربر است. همچنین در پژوهش‌های انجام‌شده درزمینه ژئوپرتال توصیه گر مشابهت بین فراداده‌های آیتم‌های اطلاعاتی استخراج می‌شود. این مشابهت صرفاً مشابهت یک‌بعدی است و بر مبنای توصیفات به‌کاررفته در فراداده است درحالی‌که در پژوهش حاضر استخراج مشابهت بر مبنای سه دسته از روابط مکانی، زمانی و توصیفی است. نتایج پیاده‌سازی ایده پژوهش حاضر در قالب یک سامانه نرم‌افزاری تحت وب نشان می‌دهد که سیستم این قابلیت را دارد که با توجه به ساختار هستی‌شناسی موجود، امتیاز بیشتری به آیتم‌های اطلاعاتی مرتبط تخصیص دهد. دراین پژوهش از هستی شناسی نسبتا ساده ای استفاد شده است که پیشنهاد میشود در پژوهشهای آینده از ساختارهای هستی شناسی پیچیده تری استفاده شود تا نتایج توصیه گری دقیقتر باشد. در این پژوهش برای انجام توصیه گری از الگوریتم‌های موجود استفاده شد اما می‌توان در پژوهش‌های آینده به توسعه الگوریتم‌هایی پرداخت که در حوزه موردمطالعه بهینه‌سازی و مطابقت داده‌شده‌اند. همچنین در این پژوهش تنها به تحلیل معنایی مشابهت هستی شناسی در سطح نمونه‌های کلاس‌ها و روابط تعریف‌شده برای این کلاس‌ها پرداخته‌شده است؛ اما در پژوهش آینده می‌توان تحلیل‌های سلسله مراتبی کلاس‌ها و همچنین روابط داده‌ای را نیز به این تحلیل‌ها اضافه نمود. همچنین می‌توان این تحلیل معنایی را در توابع محاسبه مجاورت الگوریتم‌های توصیه گر ترکیب نمود تا روند توصیه گری در ژئوپرتال یک روند پیوسته‌ای باشد.

## مراجع


[1] Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions on, 17(6), 734-749 .

[2] Athanasis, N., Kalabokidis, K., Vaitis, M., & Soulakellis, N. (2009). Towards a semantics-based approach in the development of geographic portals. Computers & Geosciences, 35(2), 301-308 .

[3] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. Scientific american, 284(5), 28 .۳۷-

[4] besheh, S. (2011). turning spatial search engine to spatial recommednder system. (Master of Science), University of Twente .

[5] Bigdeli, E., & Bahmani, Z. (2008). Comparing accuracy of cosine-based similarity and correlation-based similarity algorithms in tourism recommender systems. Paper presented at the Management of Innovation and Technology, 2008. ICMIT 2008. 4th IEEE International Conference on.

[6] Blanco-Fernández, Y., Pazos-Arias, J. J., Gil-Solla, A., Ramos-Cabrer, M., López-Nores, M., García-Duque, J., . . . Bermejo-Muñoz, J. (2008). A flexible semantic inference methodology to reason about user preferences in knowledge-based recommender systems. Knowledge-Based Systems, 21(4), 305-320 .

[7] Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A .(2013) .Recommender systems survey. Knowledge-Based Systems, 46, 109-132 .

[8] Borges, H. L., & Lorena, A. C. (2010). A survey on recommender systems for news data Smart Information and Knowledge Management (pp. 129-151): Springer.

[9] Buehler, K., McKee, L & ,.Open GIS Consortium, I. O. T. C. (1996). The OpenGIS Guide: Introduction to Interoperable Geoprocessing ; Part 1of the Open Geodata Interoperability Specification (OGIS): Open GIS Consortium, Incorporated.

[10] Burke, R. (2002). Hybrid recommender systems: Survey and experiments. User modeling and user-adapted interaction, 12(4), 331-370 .




www.SID.ir

Archive of SID



[11] Choi, K., Yoo, D., Kim, G., & Suh, Y. (2012). A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis. Electronic Commerce Research and Applications, 11(4), 309-317 .

[12] de Andrade, F. G., de Souza Baptista, C., & Davis Jr, C. A. (2014). Improving geographic information retrieval in spatial data infrastructures. GeoInformatica, 18(4), 793-818.

[13] Ekstrand, M. D., Riedl, J. T., & Konstan, J. A. (2011). Collaborative filtering recommender systems. Foundations and Trends in Human-Computer Interaction, 4(2), 81-173 .

[14] Fernández, T .D., & Castellanos, E. (2006). Towards user-driven spatial data infrastructures. An approach oriented to sustainable development. Paper presented at the GSDI-9 World Conference, Santiago, Chile.

[15] Fox, E. A., & Shaw, J. A. (1994). Combination of multiple searches. NIST SPECIAL PUBLICATION SP, 243-243.

[16] Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? International journal of human-computer studies, 43(5), 907-928 .

[17] Hochmair, H. H. (2005). Ontology matching for spatial data retrieval from Internet portals GeoSpatial Semantics (pp. 166-182): Springer.

[18] Huang, Z., Chen, H., & Zeng, D. (2004). Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. ACM Transactions on Information Systems (TOIS), 22(1), 116-142 .

[19] Infrastructures, D. S. D. (2004). the SDI Cookbook. GSDI/Nebert .

[20] Larson, J., Siliceo, M. A. O., & dos Santos, M. P. (2006). Are geospatial catalogues reaching their goals? Paper presented at the 9th AGILE Conference on Geographic Information Science: Shaping the future of Geographic Information Science in Europe.

[21] Maedche, A. (2002). Ontology learning for the semantic web: Springer Science & Business Media.

[22] Nogueras-Iso, J., Muro-Medrano, P. R., & Zarazaga-Soria, F. F. (2005). Geographic information metadata for spatial data infrastructures : resources, interoperability and information retrieval. Berlin [u.a.]: Springer.

[23] Park, D. H., Kim, H. K., Choi, I. Y., & Kim, J. K. (2012). A literature review and classification of recommender systems research. Expert Systems with Applications, 39(11), 10059-10072 .

[24] Pazzani ,M. J. (1999). A framework for collaborative, content-based and demographic filtering. Artificial Intelligence Review, 13(5-6), 393-408 .

[25] Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems The adaptive web (pp. 325-341): Springer.

[26] Polat, H., & Du, W. (2005). SVD-based collaborative filtering with privacy. Paper presented at the Proceedings of the 2005 ACM symposium on Applied computing.

[27] Resch, B., & Zimmer, B. (2013). User experience design in professional map-based Geo-Portals. ISPRS international journal of geo-information, 2(4), 1015-1037 .

[28] Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000). Application of dimensionality reduction in recommender system-a case study: DTIC Document.

[29] Schafer, J. B., Frankowski, D., Herlocker ,J., & Sen, S. (2007). Collaborative filtering recommender systems The adaptive web (pp. 291-324): Springer.

[30] Smits, P. C., & Friis-Christensen, A. (2007). Resource discovery in a European spatial data infrastructure. Knowledge and Data Engineering, IEEE Transactions on, 19(1), 85-95 .

[31] Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. Advances in artificial intelligence, 2009, 4 .

[32] Sugumaran, V., & Gulla, J. A. (2011). Applied semantic web technologies: CRC Press.

[33] Tversky A (1977) Features of similarity. Psychol Rev 84(4):327–352. doi:10.1037/0033-295X.84.4.327

[34] van Oort ,P. A., Kuyper, M., Bregt, A., & Crompvoets, J. (2009). Geoportals: an internet marketing perspective. Data Science Journal, 8, 162-181 .

[35] Vockner, B., Richter, A., & Mittlböck, M. (2013). From geoportals to geographic knowledge portals. ISPRS International Journal of Geo-Information, 2(2), 256-275.



www.SID.ir