Archive of SID

NASTINFO

بازیابی اطلاعات جغرافیایی در محیط وب:
مروری بر نوشته‌ها و ترسیم مدل مفهومی

بهروز رسولی | سعید اسدی

چکیده

هدف: ارائه مدل مفهومی برای بازیابی اطلاعات جغرافیایی.

روش/ رویکرد پژوهش: از طریق مطالعه منابع مرتبط با حوزه بازیابی اطلاعات جغرافیایی، الگوی رایج شناسایی و سپس طراحی مدل مفهومی معماری نظام بازیابی اطلاعات مکانی انجام شد.

یافته‌ها: تاکنون سه رویکرد اصلی برای بازیابی اطلاعات جغرافیایی شامل مدل‌های الگو- محور، مبتنی بر هستی‌شناسی، و ماشین‌آموز پیشنهاد شده است. استفاده از یک پایگاه جاینامه‌ای یا یک پایگاه دانش اطلاعات جغرافیایی یکی از ملزوماتی است که بسیاری از روش‌های بازیابی اطلاعات جغرافیایی نیازمند آن هستند. رابط کاربری و رتبه‌بندی نیز از مسائل مهمی است که در طراحی نظام بازیابی اطلاعات جغرافیایی باید مورد توجه قرار گیرند.

نتیجه‌گیری: نوع‌شناسی روش‌های بازیابی اطلاعات جغرافیایی می‌تواند به مطالعه نظام‌مند این حوزه کمک کند. همچنین، مدل مفهومی ارائه‌شده می‌تواند برای طراحی موتور جستجوی بازیابی اطلاعات جغرافیایی استفاده شود.

کلیدواژه‌ها

بازیابی اطلاعات، اطلاعات مکانی، اطلاعات جغرافیایی، مدل مفهومی، اطلاعات جغرافیایی

www.SID.ir

Archive of SID

بازیابی اطلاعات جغرافیایی در محیط وب:
مروری بر نوشته‌ها و ترسیم مدل مفهومی

بهروز رسولی[1]
سعید اسدی[2]

تاریخ دریافت: ۹٤/۰۷/۲۰
تاریخ پذیرش: ۹٤/۰۹/۱۷

۱. دانشجوی دکترای علم اطلاعات و دانش‌شناسی، پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) (نویسنده‌مسئول)
rasouli@students.irandoc.ac.ir
۲. استادیار گروه علم اطلاعات و دانش‌شناسی، دانشگاه شاهد
s.asadi@shahed.ac.ir

3. Geographic Information Systems (GIS)

٤. Spatial: طبق تعریف دیکشنری آکسفورد این واژه از واژه لاتین spatium مشتق شده و به‌معنی "فضا" یا "مکان" است. بنابراین، spatial به‌معنی فضا یا مکانی بودن است. موقعیت، اندازه، و شکل مقیاس‌هایی هستند که می‌توانند به این واژه منتسب شوند. تفاوت این واژه با واژه جغرافیایی (مطالعه سطح زمین) دامنه تعریف آنهاست؛ به‌عبارتی، واژه مکانی اعم از جغرافیایی‌است.

5. Abresch
6. Loglisci
7. Sanderson & Kohler
8. Jones, Alani, & Tudhope
9. McCurley

مقدمه

دهه‌های متوالی پایگاه‌های داده یا نظام‌های اطلاعات جغرافیایی[3] مهم‌ترین گردآورنده اطلاعات مکانی[4] قلمداد می‌شدند (آبرش[5]، ۲۰۰۸، ص ۳۸)، اما با ظهور وب در دهه ۱۹۹۰ این گزاره به چالش کشیده شد (لاگلیسکی[6] و همکاران، ۲۰۱۲). امروزه، بخش قابل توجهی از جستجوهای روزانه در بانک‌های اطلاعاتی و موتورهای جستجو حاوی ابعاد مکانی است. منابع اطلاعاتی متعددی اطلاعات را نه تنها در حوزه‌های موضوعی بلکه در یک بافت جغرافیایی ارائه می‌دهند. زمانی که کاربران وب در جستجوی اطلاعات مورد نیاز خود هستند احتمال پرس‌وجوی اطلاعاتی که دارای بار مکانی باشد دور از ذهن نیست. ساندرسن و کولر[7] (۲۰۰۳) و جونز و آلانی، و تودهاپ[8] (۲۰۰۱) تخمین می‌زنند که ۱۳ تا ۱۵ درصد پرس‌وجوهای کاربران در موتورهای جستجوی عمومی شامل اسم یک مکان می‌شود. بسیاری از منابع نیز اطلاعات مکانی دارند. مک‌کورلی[9] (۲۰۰۱) نشان داد که ۴/۵ درصد از تمامی صفحات وب دارای کد پستی، ۸/۵ درصد دارای شماره تلفن، و ۹/۵ درصد شامل حداقل یکی از این دو مورد هستند.

نشانی‌ها و اطلاعات مکان‌های مختلف در زندگی روزمره مردم نقش حیاتی دارند؛ آنها گاهی مجبور به جستجوی نشانی مکان‌های ناآشنا از طریق وب و استفاده از نقشه‌های مختلف برای پیدا کردن آن نشانی‌ها و جهت‌یابی‌های مورد نظر خود

www.SID.ir

Archive of SID

می‌شوند. بازیابی این اطلاعات مسئله چالش‌برانگیزی است که باید به شیوه‌ای کارآمد به آن پاسخ داد (هولمز[1]، ۱۹۹۰). موتورهای جستجوی عمومی می‌توانند تا حدی نیاز کاربران را در بازیابی اطلاعات مکانی برطرف کنند. به‌طور مثال، اگر شخصی به‌دنبال اطلاعات هتلی در تهران می‌گردد موتور جستجو براساس روش‌های آماری (برای نمونه، بسامد یک واژه در صفحه وب) می‌تواند نتایجی را بازیابی کند که پاسخگوی نیاز کاربران باشد (اسدی[2] و همکاران، ۲۰۰۷)، اما نه نیاز کاربران همیشه به این سادگی است و نه موتورهای جستجوی مذکور قابلیت‌های تخصصی برای بازیابی اطلاعات جغرافیایی دارند. بنابراین، موضوع بازیابی اطلاعات جغرافیایی[3] در دهه‌های اخیر مطرح شده است (بیزا ییتس و ریبیرو- نتو[4]، ۱۹۹۹) تا بتواند به پرسش‌های گوناگون در این زمینه پاسخ دهد. این حوزه به‌سرعت در حال پیشرفت است و متون مربوط به آن روز به‌روز غنی‌تر می‌شود (پروس[5] و همکاران، ۲۰۰۵). بازیابی اطلاعات جغرافیایی، حوزه مطالعاتی و کاربردی از بازیابی اطلاعات قلمداد می‌شود (پررا ـ اورتگا[6] و همکاران، ۲۰۱۳) که به‌صورت خلاصه آن را می‌توان به این شکل تعریف کرد:

بازیابی اطلاعات جغرافیایی زیرشاخه‌ای از علم بازیابی اطلاعات است که با نمایه‌سازی، جستجو، بازیابی[7]، و مرور[8] اطلاعات جغرافیایی سروکار دارد و به‌دنبال طراحی نظام‌هایی است تا بتواند این وظایف و فعالیت‌ها را به‌شکلی کارآمد و مؤثر اجرا کند.

بازیابی اطلاعات جغرافیایی از اوایل نیمه دوم قرن بیستم مورد توجه قرار گرفت، اما ظهور وب سبب رونق این حوزه پژوهشی شد و مطالعات جدی در این زمینه از اویل دهه ۱۹۹۰ کلید خورد. هیرست[9] (۱۹۹۲) از پژوهشگرانی است که پژوهش‌های مهمی در این زمینه انجام داده و مطالعات او بیشتر معطوف به استخراج الگو از متون بدون ساختار[10] بوده است. بازیابی اطلاعات جغرافیایی یک حوزه میان‌رشته‌ای است و جوانب مختلف آن در رشته‌های علم اطلاعات، علوم رایانه و برنامه‌نویسی، و جغرافیا بررسی می‌شود. تمامی این تخصص‌ها در کنار هم امروزه در یک رشته دانشگاهی با عنوان "علم اطلاعات جغرافیایی"[11] آموزش داده می‌شود. متخصصان جغرافیا برای ارائه داده‌های مکانی، متخصصان رایانه برای فراهم کردن زیرساخت‌های نرم‌افزاری، و متخصصان علم اطلاعات و دانش‌شناسی برای ارائه مدل‌های سازماندهی و مدل‌های داده‌ای مناسب برای پیوند این دو حوزه می‌توانند بازیابی داده‌های مکانی را کارآمدتر از گذشته کنند.

1. Holmes
2. Asadi
۳. این اصطلاح معادل Geographic Information Retrieval (GIR) است. ممکن است به‌جای بازیابی اطلاعات جغرافیایی از بازیابی اطلاعات مکانی نیز استفاده شود. اطلاعات مکانی حوزه گسترده‌تری نسبت به اطلاعات جغرافیایی دارد؛ با این حال، اصطلاح بازیابی اطلاعات جغرافیایی رایج‌تر بوده و به‌همین دلیل در این مقاله از این اصطلاح استفاده شده است.
4. Baeza-Yates & Ribeiro-Neto
5. Purves
6. Perea-Ortega
7. Retrieval
8. Browsing
9. Hearst
10. Unrestricted text
11. Geographic Information Science

www.SID.ir

Archive of SID

ورود ابزارهای موبایل مانند، گوشی‌های تلفن هوشمند، تبلت‌ها، و سایر ابزارهای دستی سیار سبب شده است تا موضوع بازیابی اطلاعات مکانی اهمیت بیشتری بیابد. تأکید بر فناوری‌های چندرسانه‌ای نیاز به بازنمایی اطلاعات مکانی را افزایش داده و امکانات مناسبی برای پاسخگویی به این نیازها فراهم کرده است (اجنهافر و کوهن[1]، ۱۹۹۹). با این حال، ابزارهای کارآمد در فضای وب برای جستجوی داده‌های مکانی فراوان نیستند. شرکت گوگل از مؤسسه‌هایی است که با راه‌اندازی دو طرح گوگل ارث[2] و گوگل مپز[3] توانسته موفقیت نسبی در جذب کاربران جویای داده‌های مکانی کسب کند. ویکی‌مَپیا[4] از دیگر طرح‌های مرتبط است که با استفاده از مشارکت اطلاعاتی کاربران در سراسر جهان، پایگاه داده‌ای مکانی را توسعه داده است. در حال حاضر، ابزارهای دیگری از قبیل Bing Maps، Here Map، MapQuest، TomTom، Waze، Map، Openstreetmap، Yahoo! Maps، Maps.randmcnally، و Yandex جستجوی داده‌های مکانی را در وب تسهیل بخشیده‌اند.

در ایران، طـرح ساماندهی جی‌آی‌اس ملی از معدود طرح‌هایی است که در زمینه سازماندهی داده‌های مکانی از سوی شورای ملی کاربران نظام‌های اطلاعات جغرافیایی در سال ۱۳۷۷ پیشنهاد و اجرا شد. این شورا به‌منظور سیاست‌گذاری، برنامه‌ریزی، و هماهنگ‌سازی فعالیت‌ها در زمینه جی‌آی‌اس، تحلیل نیازمندی‌ها، و نیز بهره‌برداری شایسته از تمامی ظرفیت‌های علمی، فنی، و نیروی انسانی در راستای ایجاد و به‌کارگیری جی‌آی‌اس در دی ماه ۱۳۷۲ تأسیس شد (صدیقی، ۱۳۸۳).

با اینکه پیشینه پژوهش‌های فارسی در حوزه جی‌آی‌اس و کاربردهای آن غنی به‌نظر می‌رسد (برای مثال، پرهیزکار، ۱۳۷۶؛ علی‌گلی، ۱۳۷۸؛ بهبودی، ۱۳۸۰؛ رنجبران، ۱۳۸۰؛ صدیقی، ۱۳۸۳)؛ اسدی، ۱۳۸۹؛ فیروزی، سجادیان، و سجادیان، ۱۳۹۰؛ پورنقی، ۱۳۹۲)، اما بحث درباره بازیابی اطلاعات مکانی به‌نسبت مغفول مانده است. پژوهش حاضر می‌کوشد با مرور پژوهش‌های موجود در حوزه بازیابی اطلاعات جغرافیایی و بررسی انواع مدل‌ها و الگوهای رایج، مدل مفهومی بازیابی اطلاعات مکانی را ترسیم کند.

مدل‌های بازیابی اطلاعات جغرافیایی

تاکنون، مدل‌های گوناگونی درباره بازیابی اطلاعات جغرافیایی ارائه شده است. با توجه به پیچیدگی‌های خاصی که به‌دلیل نامتجانس بودن اطلاعات جغرافیایی در وب وجود دارد (ماتا و کلارامانت[5]، ۲۰۱۱)، هیچ‌کدام از این مدل‌ها به‌صورت کامل

1. Egenhofer & Kuhn
2. Google Earth (https://earth.google.com)
3. Google Maps (https://maps.google.com)
4. Wikimapia (http://wikimapia.org)
5. Mata & Claramunt

www.SID.ir

Archive of SID

پذیرفته نشده است و ارائه مدل‌های جدید هنوز ادامه دارد. براساس ماهیت این مدل‌ها و ابزارهای مورد استفاده می‌توان آنها را به سه دسته عام تقسیم‌بندی کرد: مدل‌های الگو- محور[1]، هستی‌شناسی‌مدار[2]، و ماشین‌آموز[3]. البته مدل‌هایی هم وجود دارند که نمی‌توان آنها را در یکی از این سه دسته طبقه‌بندی کرد و موفقیت چندانی نیز کسب نکرده‌اند. به‌طور مثال، کازما[4] (۲۰۱۲) مدل خود را براساس استخراج نام‌های جغرافیایی از پرسمان[5]های کاربران پیشنهاد کرده است. کیتامایا[6] و همکارانش (۲۰۱۳) روشی را برای استخراج ویژگی‌های مرتبط به‌هم براساس تاریخچه استفاده کاربر از یک نظام بازیابی اطلاعات پیشنهاد کرده‌اند.

۱. مدل‌های الگو- محور

مدل‌های الگو- محور از نخستین رویکردهای بازیابی اطلاعات جغرافیایی در وب هستند؛ پیش‌فرض آنها این است که الگوهای مشترکی در اطلاعات جغرافیایی مانند نشانی‌ها یا کدهای پستی وجود دارد که با کشف آنها می‌توان به ایجاد نظامی کارآمد برای بازیابی اطلاعات جغرافیایی دست زد (ناگابوشان، انگدی، و انامی[7]، ۲۰۰۶). این مدل در آغاز موفقیت‌هایی کسب کرد و برخی موتورهای جستجو اقدام به پیاده‌سازی این روش کردند، اما با چالش‌های مهمی روبه‌رو شدند. از جمله این چالش‌ها می‌توان به طبیعت پویای نام‌های جغرافیایی (برای مثال، تغییر نام خیابان‌ها)، حالت‌های مختلف نشانی‌نویسی (اختصارنویسی، نوشتن نشانی پستی از جزء به کل یا برعکس، نقشه، و غیره)، کمبود سامانه‌های اطلاعات مکانی در کشورهای کمتر توسعه‌یافته، تشابه اسامی جغرافیایی، و شباهت نام‌های جغرافیایی و نام‌های دیگر به‌ویژه نام اشخاص (اسدی و همکاران، ۲۰۰۶) اشاره کرد.

● شناسایی موجودیت جغرافیایی

شناسایی موجودیت جغرافیایی[8] (موسلی[9]، ۱۹۹۹) یکی از رویکردهای رایج در بازیابی اطلاعات جغرافیایی است که بر استخراج نام‌های جغرافیایی از متون و اسناد تمرکز می‌کند. این روش زمان‌بر است و در دنیای واقعی نمی‌توان از آن بهره زیادی گرفت. اساس این مدل، شناخت هر موجودیت جغرافیایی در وب براساس اطلاعات پیرامونی است. برای مثال، اینکه میزبان یک وبگاه در کدام شهر یا کشور قرار دارد یک نشانی جغرافیایی برای آن وبگاه محسوب می‌شود. واضح است که چنین رویکردی دقت کافی ندارد، زیرا ممکن است میزبان وبگاهی در کشور کانادا باشد، اما موضوع و مخاطب آن ارتباطی با کانادا نداشته باشد.

1. Pattern-based models
2. Ontology-driven models
3. Machine learning methods
4. Kazma
5. Query
6. Kitayama
7. Nagabhushan, Angadi, & Anami
8. Named Entity Recognition (NER)
9. Muslea

www.SID.ir

Archive of SID

● رویکرد جاینامه‌ای

رویکرد جاینامه‌ای[1] بر ایجاد یک فرهنگ جغرافیایی و استفاده از آن در بازیابی اطلاعات مکان‌های جغرافیایی متمرکز است (کن[2] و همکاران، ۲۰۰۵). فکر اصلی این مدل بر تشخیص اسامی مکان‌ها و نشانی‌ها در متن و برچسب زدن آنها استوار شده است. در این روش، منابعی که پیشتر در یک پایگاه اطلاعاتی گردآوری شده‌اند به‌صورت غیربرخط[3] تجزیه و تحلیل می‌شوند تا داده‌های مکانی موجود در آنها مشخص شود. سپس با استفاده از برچسب‌زن جغرافیایی[4] به هر کدام از نام‌های جغرافیایی موجود در منبع برچسب متناسب زده می‌شود و نام یکسان آن، طول و عرض جغرافیایی، درختواره ارتباطی، و دیگر اطلاعات آن براساس جاینامه‌ای که در ورای نظام قرار دارد مشخص خواهد شد. به‌طور مثال، اگر در یک صفحه وب کلمه "طهران" مشاهده شود، براساس اطلاعات موجود در جاینامه مشخص می‌شود که نخست، این واژه اسم یک مکان جغرافیایی است و دوم، اسم یک شهر در استان تهران در ایران و صورت برگزیده آن "تهران" است. بنابراین، نظام بازیابی اطلاعات جغرافیایی از اسامی مکان‌ها و نشانی‌ها ابهام‌زدایی می‌کند و نتایج مکانی قابل قبولی ارائه می‌دهد.

این روش مستلزم جاینامه‌های اثربخش جغرافیایی است که در بسیاری از کشورها وجود ندارد (اسدی[5] و همکاران، ۲۰۰۵). به نظر المیدائویی[6] و همکاران (۲۰۱۵)، با توجه به پیشرفت‌های اخیر در حوزه روش‌های نمایه‌سازی معنایی خودکار، ایجاد تاکسونومی[7]های جامع در این زمینه می‌تواند به پیشرفت این حوزه کمک کند.

● روش مبتنی بر الگو

روش مبتنی بر الگو[8] یکی از روش‌های مورد توجه در بازیابی اطلاعات جغرافیایی، استخراج اسامی، و نشانی‌های مکانی است. در این روش، منابع موجود برای یافتن نشانی‌های پستی تجزیه و تحلیل می‌شوند. وبگاه‌های خدمات مکانی و محلی، نشانی‌هایی برای مراجعه حضوری یا مکاتبه در اختیار کاربران قرار می‌دهند. هرچند وجود صفحاتی با عنوان‌هایی مثل "تماس با ما" و "اسامی شعبه‌ها" می‌تواند سرنخ خوبی برای یافتن نشانی‌ها در یک منبع وبی باشد، اما همیشه و همه نشانی‌ها و اسامی جغرافی به‌سادگی قابل تشخیص نیستند. نمونه مدل بازیابی اطلاعات مکانی مبتنی بر الگو توسط اسدی و همکاران (۲۰۰۸) معرفی شده است. این مدل بدون نیاز به دخالت انسان و استفاده از جاینامه‌های آماده، به‌صورت خودکار، نشانی‌های

1. Gazetteer approach
2. Can
3. Offline
4. Geotagger
5. Asadi
6. El Midaoui
7. Taxonomy
8. Pattern-Based Model

www.SID.ir

Archive of SID

جغرافیایی را استخراج و تجزیه و تحلیل می‌کند. در این روش، یک نشانی به اجزای معنادار تقسیم می‌شود و نیازی به جاینامه‌های بزرگ نیست و مشکل اسامی جدید نیز برطرف می‌شود (اسدی و همکاران، ۲۰۰۹). نظام یادشده با یک مجموعه کوچک اولیه از اسامی جغرافیایی و دستورات زبانی شروع به‌کار می‌کند و اسامی جغرافیایی، نشانی‌ها، و الگوهای نشانی‌دهی جدید را در فرایند جستجو یا کشف دانش به پایگاه‌هایی می‌افزاید که به‌صورت خودکار تولید و روزآمد می‌شوند. در این نظام، استخراج نشانی از پنج جزء مهم (پیش‌ـ پردازشگر اچ‌تی‌ام‌ال، تجزیه‌کننده، جستجوگر دانش، تصمیم‌گیرنده، و انباشت‌گر دانش)[1] تشکیل شده است.

- **پیش پردازشگر اچ‌تی‌ام‌ال**: این جزء برچسب‌های اچ‌تی‌ام‌ال را تحلیل و آنها را به کدهای ایکس‌ام‌ال تبدیل می‌کند. سپس، تحلیل عمیق‌تری روی کدهای تبدیل‌شده انجام می‌شود تا محتوای اطلاعاتی به‌دست آید. نظام، نتایج را در یک توالی خطی بیان می‌کند و آنها را با شماره گره از همدیگر متمایز می‌سازد و نمایه گره‌ها ساخته می‌شود.
- **تجزیه‌کننده**: این واحد تمام جمله‌هایی را که احتمالاً اطلاعات جغرافیایی دارند، شناسایی و آنها را به اجزای کوچک‌تر تجزیه می‌کند. هر بخشی که در این گام به‌دست می‌آید، یک واحد جستجوی پیش‌فرض در جستجوگر دانش تلقی خواهد شد. جملات احتمالی براساس مجموعه‌ای دستورالعمل زبانی و یک پایگاه کوچک از اسامی کشورها و واژگان جغرافیایی مشخص می‌شوند. به‌طور مثال، تمام جمله‌هایی که در آنها چنین واژگان یا عبارت‌هایی باشند انتخاب می‌شوند: گیرنده، کوچه، نشانی پستی، شماره تلفن، پلاک، شعبه‌ها، مراجعه حضوری، و مانند آنها.
- **جستجوگر دانش**: این جزء، عناصر یک نشانی جغرافیایی بالقوه را بخش‌بندی و حالت‌های ممکن آن را استخراج می‌کند و آنها را در سه مرحله به‌صورت فهرستی از الگوهای ممکن درمی‌آورد:

۱. استاندارد کردن واژه‌ها: ابتدا واژه‌هایی که املاهای مختلف دارند (طهران/ تهران)، یا به‌صورت اختصاری نوشته شده‌اند (USA)، یا مترادف‌های مختلف (Holland/ Netherland)، همه استاندارد می‌شوند؛

۲. تطبیق با نام مکان‌ها در پایگاه دانش[2]: در این مرحله، عناصر تجزیه‌شده در مراحل قبلی در تطبیق با اسامی جغرافیایی موجود در پایگاه دانش، در سطوح جزئی و دقیق‌تر تفکیک می‌شوند؛ و

1. HTML Pre-processor, Parser, Knowledge Searcher, Decision Maker, & Knowledge Accumulator
2. Knowledge-base

www.SID.ir

Archive of SID

۳. رفع ابهام: در این مرحله، ابهام‌هایی که در تطبیق نام‌های جغرافیایی به‌وجود آمده است رفع می‌شود.

- **تصمیم‌گیرنده**: این مرحله به‌دنبال تصمیم‌گیری در مورد ارزش جغرافیایی عبارت‌هایی است که به‌صورت بالقوه به‌عنوان اطلاعات جغرافیایی (به‌طور مثال، یک نشانی) تلقی شده‌اند. تصمیم‌گیرنده با تطبیق اطلاعات بالقوه با الگوهای نشانی‌نویسی که از قبل در پایگاه اطلاعاتی ذخیره شده‌اند، سرانجام تصمیم می‌گیرد که این اطلاعات، مکانی هستند یا خیر. در این مرحله، ابهام‌ها و تضادهایی که در مورد نام‌ها وجود دارد رفع می‌شود. این ابهامات و تضادها می‌تواند هم به‌شکل نحوی باشد و هم به‌شکل معنایی. به‌طور مثال، "شهریار" هم می‌تواند نام شهری در استان البرز باشد و هم می‌تواند نام یک شخص باشد. این موارد در زمره ابهامات جغرافیایی/ غیرجغرافیایی[1] قرار می‌گیرند. از سوی دیگر، "کویت" هم اسم شهر است و هم اسم کشور. در مرحله بعدی، هر کدام از این نشانی‌های احتمالی به اجزای ریزتری تقسیم خواهند شد و اگر نشانی تشخیص داده شوند، قسمت‌های گم‌شده در نشانی (طبق درخت سلسله‌مراتبی نام‌های جغرافیایی) به جمله اضافه خواهد شد. ممکن است نشانی مندرج در منبعی به این شکل باشد: "تهران، خیابان آزادی، هشتم غربی، پلاک ۲۰" که اگر نظام بتواند با موفقیت عمل کند نشانی براساس اطلاعات موجود در پایگاه دانش به‌شکل زیر تکمیل خواهد شد:

"(ایران)، (استان تهران)، تهران، (منطقه ۲)، خیابان آزادی، (کوچه) هشتم غربی، پلاک ۲۰"

موفقیت این کار به میزان بسامد نشانی و اسامی یادشده در وب بستگی دارد.

- **انباشت‌گر دانش**: این جزء دارای دو وجه است: انباشت‌گر اسامی جغرافیایی و انباشت‌گر الگوهای نشانی‌نویسی. وظیفه انباشت‌گر اسامی جغرافیایی، گردآوری نام‌های جغرافیایی موجود در منابع است؛ نام شهرها، کشورها، خیابان‌ها، کوچه‌ها، مناطق، و غیره. وظیفه انباشت‌گر الگوهای نشانی‌نویسی گردآوری الگوهای مختلفی است که مردم در کشورهای مختلف از آنها برای نوشتن نشانی استفاده می‌کنند. حساس‌ترین جزء نظام بازیابی اطلاعات جغرافیایی همین بخش است و عمکرد مناسب نظام تا حد زیادی متکی به همین جزء است. برای شروع به کار باید انباشت‌گر دانش توسط انسان تا حدی توصیف شود. هرچه این توصیف دقیق‌تر باشد عملکرد کلی نیز بهتر می‌شود. یکی از مهم‌ترین چالش‌های این

1. Geo/non-geo ambiguities

www.SID.ir

Archive of SID

مدل نیز در مورد همین جزء پایانی است که عملکرد مدل را در دنیای واقعی با مشکلاتی مواجه می‌سازد، از جمله عدم یکدستی اطلاعات و شیوه‌های نگارش در وب. در مثال بالا، دو فرد ممکن است نشانی مذکور را به دو شیوه متفاوت بنویسند و حتی یک فرد ممکن است الگوی یکسانی برای نوشتن همه نشانی‌ها نداشته باشد و یک نشانی را کامل و یکی را ناقص بنویسد.

۲. مدل مبتنی بر هستی‌شناسی

بازیابی مبتنی بر هستی‌شناسی یکی از روش‌هایی است که برای این نوع بازیابی اطلاعات استفاده شده است (بورگس[1]، ۲۰۰۶؛ بورگس و همکاران، ۲۰۰۷؛ کای[2] و همکاران، ۲۰۰۵). کاربرد هستی‌شناسی در نظام‌های بازیابی اطلاعات بحث به‌نسبت جدیدی قلمداد می‌شود که متخصصان علوم رایانه، زبان‌شناسی، و علم اطلاعات در مطالعات خود به آن توجه کرده‌اند (اندرسن، نیلسن، و تامسن[3]، ۲۰۰۱). محدودیت‌های ابزارهای سنتی در فضای وب معنایی، عدم پاسخگویی دقیق اصطلاحنامه‌ها و ابزارهای مشابه به پرسش‌ها، بسط کلیدواژه‌های جستجو در نظام‌های هستی‌شناسی، تعریف دقیق مفاهیم، و روابط میان آنها باعث رواج هستی‌شناسی‌ها شده است و به نظر می‌رسد این ابزارها می‌توانند بسیاری از چالش‌های ابزارهای سنتی را در بازیابی اطلاعات برطرف کنند (نپی[4] و همکاران، ۲۰۰۷؛ بوگال و همکاران، ۲۰۰۷). همچنین، درک نیاز اطلاعاتی کاربران، نمایه‌سازی مفهومی، شناخت حوزه موضوعی، بسط جستجو، مصورسازی معنایی، شخصی‌سازی، و سازماندهی مدارک بازیابی‌شده براساس روابط معنایی از کاربردهای مهم هستی‌شناسی‌ها در بازیابی اطلاعات هستند (صنعت‌جو، ۱۳۹۱).

بر همین اساس، رویکرد نوینی در بازیابی اطلاعات جغرافیایی توسعه یافت که مبتنی بر هستی‌شناسی بود. به‌طور کلی، ارجاع به مکان‌های جغرافیایی در صفحات وب می‌تواند به دو صورت باشد: مستقیم یعنی اسامی مکان‌ها یا نشانی‌های کامل پستی و غیرمستقیم مانند کدهای پستی، کدهای تلفن، یا نشانی‌های نسبی (به‌طور مثال، هتل لاله پایین‌تر از پارک لاله قرار گرفته است). مدل‌های الگو- محور در تفسیر این اطلاعات ناتوانند، در حالی که برخی روش‌های مبتنی بر هستی‌شناسی در این زمینه موفقیت‌هایی داشته‌اند، از جمله مدل معرفی‌شده توسط بورگس و همکاران (۲۰۰۷) که در سه مرحله عمل می‌کند:

۱. استخراج: در این مرحله، محتوای مرتبط در محیط وب شناسایی و استخراج می‌شود.

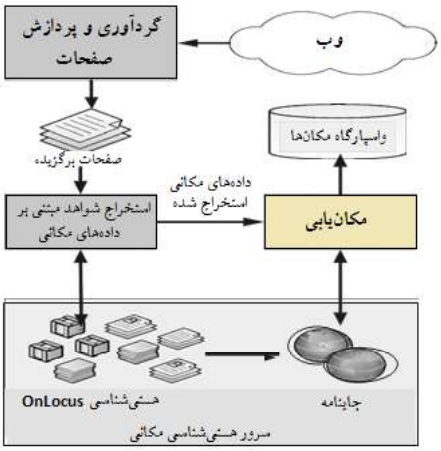
1. Borges
2. Cai, Wang, & Jiang
3. Andreasen, Nilsen, & Thomsen
4. Knappe

www.SID.ir

Archive of SID

یک نظام هستی‌شناسی استخراج قادر به شناسایی موجودیت‌ها و ارتباط‌های موجود خواهد بود. هستی‌شناسی باید شامل قوانین و دستورالعمل‌هایی برای شناسایی عناصر موجود در صفحات وب باشد؛

۲. شناسایی یا تشخیص: وظیفه مرحله شناسایی این است که اطلاعات جغرافیایی موجود در صفحات وب را شناسایی و متمایز کند (شکل ۱).



شکل ۱. عملکرد مرحله شناسایی یا تشخیص در روش بازیابی مبتنی بر هستی‌شناسی (بورگس و همکاران، ۲۰۱۱)

در این مرحله، به مسائل مرتبط با رفع ابهام‌ها نیز توجه می‌شود. همچنین اصطلاحات و اظهارات باید به‌عنوان اسامی مکان‌های جغرافیایی تشخیص داده شوند. مقایسه این اصطلاحات با جاینامه‌های موجود (مانند دو جاینامه مهم Alexandria و GeoNames) عملکردی است که در مرحله شناسایی انجام می‌شود؛ و

۳. مکان‌یابی: پس از شناسایی اطلاعات جغرافیایی، در مرحله بعد با استفاده از جاینامه‌های مناسب یا پایگاه‌های اطلاعاتی، فرایند مکان‌یابی[1] یا نشانی‌دهی تکمیل می‌شود. یافتن نشانی‌های واقعی جغرافیایی با کمک جاینامه‌ها یا انجام فرایندی به‌نام کددهی جغرافیایی[2] صورت می‌گیرد. این فرایند در مورد نشانی‌های مستقیم ساده‌تر است. در مورد نشانی‌های غیرمستقیم دو حالت وجود دارد: اگر این نشانی‌ها رسمی هستند (مانند کدهای پستی یا کد شماره‌های تلفن) باید از طریق مقایسه این کدها با پایگاه‌های اطلاعاتی مناسب بازیابی شوند، اگر غیررسمی هستند (مثل عبارت "ترسیده به چهارراه") باید از فنون پردازش زبان طبیعی استفاده کرد.

در نهایت، پس از اجرای این سه مرحله، فعالیت‌های مربوط به نمایه‌سازی،

1. Locating
2. Geocoding

www.SID.ir

Archive of SID

رتبه‌بندی، و سایر مراحل در نظام بازیابی اطلاعات اجرا خواهد شد و کاربر از این طریق اطلاعات مورد نظر خود را خواهد یافت. نظام بازیابی اطلاعات OnLocus یکی از نظام‌های مهم در این زمینه است.

۳. مدل ماشین‌آموز

مدل ماشین‌آموز یا یادگیری ماشینی (فریتگ[1]، ۱۹۹۸؛ دیتریش[2]، ۲۰۰۲) مبتنی بر کاربرد متن‌کاوی در صفحات وب است و از این طریق، اسامی مکان‌های جغرافیایی را در دسته‌های مختلف طبقه‌بندی می‌کند (به‌طور مثال، "کامیاران" در دسته شهرها، یا "قشم" در دسته جزیره‌ها). یکی دیگر از اهداف متن‌کاوی تعیین این موضوع است که یک مکان خاص در کجا واقع شده است (برای مثال، "کامیاران" در استان کردستان کشور ایران است؛ یا "کیش" جزیره‌ای در خلیج فارس است).

ساختن جاینامه‌های جامع خودکار یکی از ویژگی‌های مهم این مدل است. طوری که ماشین از طریق الگوریتم‌های متنوع این کار را به‌صورت خودکار انجام می‌دهد. در این روش، میان شش دسته از اسامی جغرافیایی تمایز وجود خواهد داشت: شهر، منطقه، کشور، جزیره، رودخانه، و کوه[3] (اوریوپینا[4]، ۲۰۰۲). نظام بازیابی اطلاعات برای هر دسته اسامی جغرافیایی مجموعه‌ای از الگوریتم‌ها را به‌کار خواهد بست. به‌طور مثال، دو مورد از این الگوها به این شکل هستند: KEYWORD + of + X یا X + KEYWORD. شمار این کلیدواژه‌ها برای هر دسته از اسامی از سه (برای جزیره) تا ۱۰ مورد (برای کوه) است. مثال زیر می‌تواند به فهم این بخش از مدل کمک کند؛ در این مثال، برای "شهر" شش کلیدواژه مجزا وجود دارد و بر این اساس هفت الگو شکل خواهد گرفت:

Class CITY uses 4 keywords (“city”, “town”, “mayor”, “streets”) and 7 corresponding patterns (“city + of + X”, “X + city”, “town + of + X”, “mayor + of + X”, “X + mayor”, “streets + of + X”, and “X + streets”)

با توجه به اینکه کلیدواژه‌ها و الگوها باید توسط انسان و به‌صورت دستی انتخاب و به نظام تزریق شوند (اوریوپینا، ۲۰۰۳) این مدل کاملاً خودکار نیست و یکی از چالش‌های آن به‌شمار می‌رود؛ زیرا نوع، تعداد، و الگوهای این عبارت‌ها در زبان‌های مختلف بسیار متنوع و متفاوت است.

1. Freitag
2. Dietterich
3. CITY, REGION, COUNTRY, ISLAND, RIVER, & MOUNTAIN
4. Uryupina

مدل مفهومی نظام‌های بازیابی اطلاعات جغرافیایی

کاربرد مدل‌های یادشده در گرو مقرون به‌صرفه بودن آنها از نظر اقتصادی و زمانی

www.SID.ir

Archive of SID

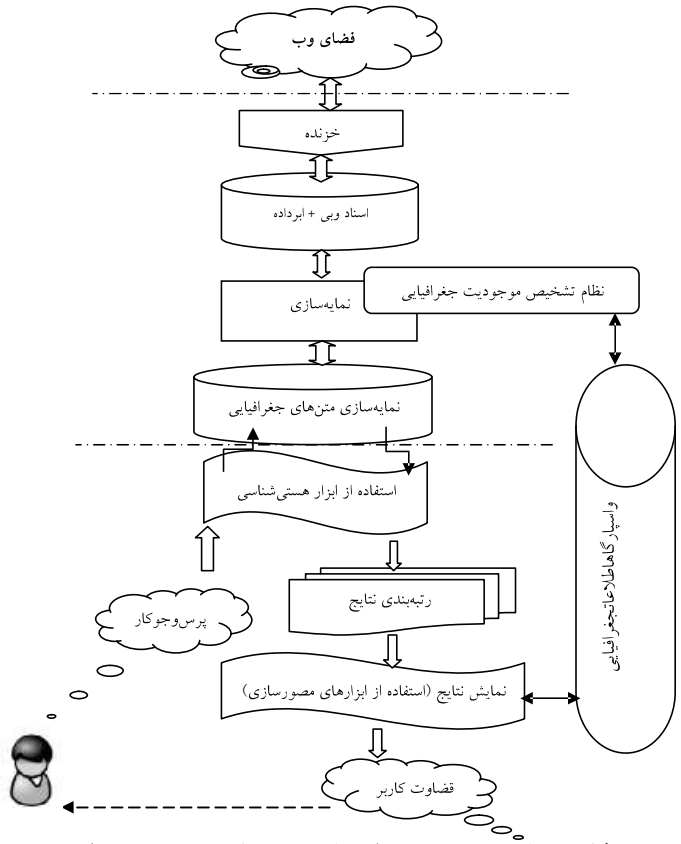
است. برخی مدل‌ها فقط در یک مورد پیاده‌سازی شده‌اند و برخی دیگر فقط در سطح نظری مطرح شده‌اند و احتمالاً فقط روی یک نمونه کنترل‌شده آزمایش شده‌اند. هدف نهایی از طراحی مدل‌های بازیابی اطلاعات جغرافیایی به‌کارگیری آنها در یک نظام بازیابی اطلاعات جغرافیایی است. موتورهای جستجوی وب از رایج‌ترین نظام‌های بازیابی اطلاعات هستند و بازیابی اطلاعات جغرافیایی می‌تواند در موتورهای جستجوی عمومی و اختصاصی اجرا شود (اسدی و همکاران، ۲۰۰۷). بسته به اینکه از چه مدلی برای بازیابی استفاده خواهیم کرد ساختار و معماری نظام بازیابی اطلاعات متفاوت خواهد بود، اما به‌طور کلی می‌توان معماری یک نظام عمومی بازیابی اطلاعات را به این صورت ترسیم کرد که خزنده، وب را جستجو می‌کند. سپس صفحات، بازیابی و نمایه‌سازی می‌شوند. از سوی دیگر، جستجوی کاربر در میان این صفحات نمایه‌سازی‌شده به بازیابی اطلاعات منجر می‌شود. بر این اساس، می‌توان چنین مدلی را برای بازیابی اطلاعات جغرافیایی نیز به‌کار برد، با این تفاوت که این نظام با موجودیت‌هایی سروکار دارد که متفاوت از موجودیت‌های دیگر در بازیابی اطلاعات است. شکل ۲ یک مدل بازیابی اطلاعات جغرافیایی را نشان می‌دهد که به‌طور اختصاصی برای این نوع اطلاعات طراحی شده است.

یک "ابزار هستی‌شناسی" هم در مرحله ارسال پرسش به نظام اطلاعات جغرافیایی به‌کار می‌رود و هم در مرحله بازیابی. از سوی دیگر، "نظام شناسایی موجودیت جغرافیایی" وظیفه تشخیص اصطلاحات جغرافیایی را در اسناد برعهده دارد و نیز می‌تواند از این اصطلاحات ابهام‌زدایی کند و در نهایت اسناد را براساس مقیاس‌های جغرافیایی طبقه‌بندی کند. موتور رتبه‌بندی با کمک اعمالی که توسط سایر اجزاء از قبل روی اسناد صورت گرفته، اسناد را براساس اهمیت و اولویت منظم و رتبه‌بندی می‌کند. "واسپارگاه اطلاعات جغرافیایی" جزء دیگری است که در نظام بازیابی اطلاعات جغرافیایی می‌تواند به‌کار گرفته شود. با استفاده از پیوندهایی که میان این بخش و اجزای شناسایی موجودیت‌های جغرافیایی و نمایش نتایج وجود دارد می‌توان دقت بازیابی اطلاعات را افزایش داد. پس از آنکه بازیابی این اطلاعات با دقت انجام شد نوبت به نحوه رتبه‌بندی نتایج می‌رسد. کومار[1] (۲۰۱۱) با اشاره به اینکه هدف غایی نظام‌های بازیابی اطلاعات، بازیابی اسناد مرتبط با پرس‌وجوی کاربر از نظام است رتبه‌بندی نتایج و اندازه‌گیری میزان ربط را از مهم‌ترین چالش‌های این قبیل نظام‌ها می‌داند.

1. Kumar

www.SID.ir

Archive of SID



شکل ۲. مدل مفهومی معماری یک نظام بازیابی اطلاعات جغرافیایی/ مکانی

با توجه به شکل ۲، مدل مفهومی در سه سطح طراحی شده است: رابط کاربری، نمایه‌سازی، و وب. پس از آنکه موتور جستجو صفحات موجود در وب را شناسایی و گردآوری کرد، نمایه‌سازی این صفحات و اطلاعات موجود در آنها آغاز می‌شود. "نظام تشخیص موجودیت جغرافیایی" که اطلاعاتش از "واسپارگاه اطلاعات جغرافیایی" تأمین می‌شود جزئی از فرایند نمایه‌سازی است که با کمک این جزء داده‌های مکانی شناسایی و نمایه‌سازی می‌شوند. پس از نمایه‌سازی این اطلاعات، ابزار هستی‌شناسی که حاوی داده‌های مکانی و اطلاعات جغرافیایی این مکان‌هاست وارد عمل می‌شود. کاربر پرسمان خود را به این ابزار می‌دهد، این ابزار با کمک اطلاعات نمایه‌سازی‌شده پرسمان کاربر را پردازش و صفحات وب را براساس اطلاعات موجود برای کاربر رتبه‌بندی می‌کند. علاوه بر نتایج متنی، از ابزارهای مصورسازی نیز می‌توان کمک گرفت تا کارآمدی موتور جستجو بیشتر شود. برای مثال، این اطلاعات را می‌توان روی یک نقشه جغرافیایی پیاده‌سازی کرد.

www.SID.ir

Archive of SID

به نظر تبلر[1] (۱۹۷۰) "در فضای جغرافیایی هریک از اشیاء با اشیائی دیگر مرتبط است، اما اشیائی که نزدیک یکدیگرند بیشتر از اشیائی که از هم دور هستند با یکدیگر ارتباط دارند". مصورسازی جغرافیایی[2] امکانی است که در طراحی رابط کاربری نظام‌های بازیابی اطلاعات جغرافیایی به‌کار می‌رود و در آن از نمایش تصویری نقاط جغرافیایی برای کاوش اطلاعات استفاده می‌شود (مک‌ایچرن[3]، ۲۰۰۱). امروزه، بسیاری از نظام‌های بازیابی اطلاعات سعی بر استفاده از این روش‌ها دارند (درودی و سلیمانی‌نژاد، ۱۳۹۲؛ درودی و محمد علیپور، ۱۳۹۲). مصورسازی می‌تواند کاربران را در یافتن مکان مورد نظر به‌خوبی راهنمایی کند.

ارزیابی نظام‌های بازیابی اطلاعات جغرافیایی

با اینکه موضوع بازیابی اطلاعات جغرافیایی هنوز در مراحل اولیه است و مطالعات اندکی درباره ارزیابی این نظام‌ها انجام شده است، همچنان موضوع مهمی تلقی می‌شود. به نظر کلاف و سندرسان[4] (۲۰۰۴) و لیندر[5] (۲۰۰۴) پیشرفت در حوزه بازیابی اطلاعات جغرافیایی نیازمند یک روش ارزیابی برای اندازه‌گیری و مقایسه فنون گوناگون در این زمینه است. کمپین GeoCLEF[6] یکی از مجموعه فعالیت‌هایی است که به‌منظور ارزیابی نظام‌های بازیابی اطلاعات جغرافیایی در سال‌های ۲۰۰۵ تا ۲۰۰۸ فعالیت داشته است.

هر نظام بازیابی اطلاعات جغرافیایی چندین جزء را دربرمی‌گیرد که می‌توانند بر همدیگر تأثیرگذار باشند. حین ارزیابی، بهتر است میان این اجزاء تمایز قائل شد و آنها را به‌طور جداگانه بررسی کرد. برخی چالش‌های مهم در این زمینه عبارت است از: ۱) ساخت جاینامه‌ها و هستی‌شناسی‌های جغرافیایی، ۲) سروکار داشتن با ارجاع‌های جغرافیایی در متن، ۳) رتبه‌بندی رکوردهای بازیابی‌شده براساس ربط جغرافیایی، و ۴) وجود رابطهای کاربری قدرتمند (مارتینز، سیلوا، و چیوز[7]، ۲۰۰۵). به‌طور سنتی، ارزیابی نظام‌های بازیابی اطلاعات غالباً بر معیار "ربط" تکیه دارد. بر این اساس، رکوردهای بازیابی‌شده یا مرتبط هستند یا غیرمرتبط. ربط با دو مقیاس "دقت" و "جامعیت" سنجیده می‌شود. صحت و خطا[8] دو معیار دیگرند که کمتر رواج دارند. جاینامه‌ها یا پایگاه‌های دانش جغرافیایی نقش مهمی در نظام‌های بازیابی ایفا می‌کنند. پتاسیس[9] و همکاران (۲۰۰۰) به محدودیت دست یافتن به جاینامه‌های جامع اشاره کرده‌اند. باید میان جاینامه‌ها و هستی‌شناسی‌های جغرافیایی تمایز قائل شد (مارتینز و همکاران، ۲۰۰۵). جاینامه‌های جغرافیایی فقط می‌توانند به ارجاع‌های

1. Tobler
2. Geo-visualization
3. MacEachren
4. Clough & Sanderson
5. Leindner
6. Geographical Conference & Labs of the Evaluation Forum
7. Martins, Silva, & Chaves
8. Accuracy & Error
9. Petasis

www.SID.ir

Archive of SID

موجود در مدارک اختصاص یابند، اما هستی‌شناسی‌ها ارتباط‌های معناشناسی/ سلسله مراتبی را نیز تشخیص می‌دهند و از این رو، می‌توانند کار نظام بازیابی اطلاعات را دقیق‌تر کنند. شناسایی و تشخیص ارجاع‌های جغرافیایی یکی دیگر از اجزای مهم نظام بازیابی اطلاعات است. این وظیفه نظام در ارتباط تنگاتنگ با جزء قبلی است (میکو، مونز، و گروور[1]، ۱۹۹۹).

یکی از فرایندهای ارزیابی نظام‌های بازیابی اطلاعات مکانی ارزیابی طرح اولیه است. یکی از معیارهای این مرحله، ارزیابی رابط کاربری است که قرار است رابط میان کاربر و پرس‌وجوی او با نظام بازیابی اطلاعات باشد. ارزیابی بخش‌هایی مانند خزنده و نمایه‌ساز نظام نیز باید در همین مرحله انجام شود. برای آگاهی از نیاز کاربران نهایی باید پس از راه‌اندازی طرح اولیه مطالعات نیازسنجی انجام شود. مسائل قابل ارزیابی در این مرحله عبارت‌اند از: ۱) اهمیت استفاده از رابط‌های کاربری چندمدلی تا به کاربران اجازه دهد نیاز اطلاعاتی خود را هم به‌صورت متنی و هم به‌صورت گرافیکی از نظام پرس‌وجو کنند، ۲) وظیفه دسترسی به منابع به‌صورت سلسله مراتبی برطبق مفاهیم جغرافیایی، ۳) اهمیت ارتباط‌های مختلف میان مفاهیم مختلف جغرافیایی برای بازیابی (برای مثال، نزدیک...، بالاتر از...، یا کنار...)، و ۴) استفاده از مزایای شماهای گوناگون نمایش نتایج (به‌طور مثال، رتبه‌بندی نتایج خوشه‌ای براساس میزان ربط) (مارتینز و همکاران، ۲۰۰۵).

نتیجه‌گیری

نظام‌های بازیابی اطلاعات مکانی روز به‌روز محبوبیت و رواج بیشتری پیدا می‌کنند و یکی از مهم‌ترین ابزارها برای برآورده کردن نیازهای اطلاعاتی کاربران به‌شمار می‌آیند. اگرچه این حوزه از بازیابی اطلاعات عمومی متمایز است، اما بسیاری از موضوع‌ها و چالش‌های این دو حوزه با هم مشترک هستند. این مقاله با طرح بازیابی اطلاعات جغرافیایی به‌عنوان حوزه‌ای مهم در دنیای امروز به بررسی انواع مدل‌ها و روش‌های بازیابی این حوزه پرداخت. سه شیوه رایج در این زمینه روش‌های الگومحور، مبتنی بر هستی‌شناسی، و ماشین‌آموز هستند که هر کدام چالش‌های خود را دارند. مهم‌ترین چالش در بیشتر روش‌های بازیابی اطلاعات جغرافیایی وابستگی آنها به یک پایگاه اطلاعات جغرافیایی مانند یک جاینامه یا پایگاه دانش جغرافیایی است که عموماً باید توسط انسان ساخته شود. به نظر می‌رسد این ابزار تقریباً برای تمامی روش‌ها در راه‌اندازی یک نظام بازیابی ضروری است. ساخت، نگهداری،

1. Mikheev, Moens, & Grover

www.SID.ir

Archive of SID

و به‌روزرسانی این ابزارها مشکلات خاص خود را دارد و متعاقباً الگوهایی نیز که براساس آنها شکل خواهند گرفت این مشکلات را خواهند داشت. طراحی الگوهایی که این نیاز را محدودتر کند می‌تواند سبب موفقیت‌های بیشتر در این حوزه شود. طراحی رابطهای کاربری، شیوه‌های رتبه‌بندی، و ارزیابی این نظام‌ها از دیگر مسائل بازیابی اطلاعات جغرافیایی هستند. در کنار این عوامل، دشواری‌های زبان فارسی در بازیابی، به‌ویژه پردازش خودکار، نباید مغفول بماند (ستوده و هنرجویان، ۱۳۹۱؛ کاظم‌پور و فهیم‌نیا، ۱۳۹۱).

پژوهش در زمینه بازیابی اطلاعات جغرافیایی بیش از گذشته شده و فناوری‌های نوین امیدهای بیشتری برای بهبود روش‌های کارآمد در این حوزه به بشر داده‌اند. دِاندراده[1] و همکاران (۲۰۱۲) معتقدند "زیرساخت‌های داده مکانی"[2] یکی از این فناوری‌های به‌نسبت جدید است که می‌تواند به بازیابی اطلاعات کمک کند. هوش مصنوعی می‌تواند در حوزه‌های دیگری به بازیابی اطلاعات جغرافیایی کمک کند. به‌طور مثال، استفاده از این فن برای استخراج متن از تصاویر (مظفری، اروجی، و مرادی، ۱۳۹۲) می‌تواند یکی از این حوزه‌های کاربردی باشد. شبکه‌های اجتماعی نیز در بازیابی اطلاعات مکانی مفیدند. لی، واکامیا، و سومیا[3] (۲۰۱۱) روشی را توسعه داده‌اند که به کمک میکروبلاگ‌نویسی[4]، شبکه‌های اجتماعی و برچسب‌های جغرافیایی در این شبکه‌ها می‌تواند بازیابی اطلاعات جغرافیایی و مرتبط با مکان را از وب آسان کند.

نیاز انسان به اطلاعاتی که ابعاد مکانی و جغرافیایی دارد رو به افزایش است و به نظر می‌رسد که ابزارهای جستجو و بازیابی اطلاعات در آینده از نظر درک پرسش‌ها و اطلاعات جغرافیایی هوشمندتر و کارآمدتر خواهند بود. با وجود این، یکی از اساسی‌ترین چالش‌ها یعنی نمایه‌سازی پیشرفت چندانی پیدا نکرده است (ویدیارسی، یاداو، و یاداو[5]، ۲۰۱۳). پژوهش‌های بیشتر در زمینه نمایه‌سازی اطلاعات جغرافیایی می‌تواند برنامه‌ریزی ایجاد ابزارهای کارآمد را برای بازیابی این نوع اطلاعات متحول کند.

مآخذ

اسدی، سعید (۱۳۸۹). کاربرد سامانه اطلاعات جغرافیایی (جی‌آی‌اس) در خدمات کتابداری و اطلاع‌رسانی. *کتاب ماه کلیات*، ۱۳ (۷)، ۳۸-۴۱.

درودی، فریبرز؛ محمدعلی‌پور، نرگس (۱۳۹۲). نقش مصورسازی در فرایند بازیابی اطلاعات.

1. De Andrade
2. Spatial data infrastructures
3. Lee, Wakamiya, & Sumiya
4. Microblogging
5. Vidyarthi, Yadav, & Yadav

www.SID.ir

Archive of SID

مطالعات ملی کتابداری و سازماندهی اطلاعات، ۲۴ (۴)، ۱۵۴-۱۷۳.

درودی، فریبرز؛ سلیمانی‌نژاد، عادل (۱۳۹۲). دیداری کردن نتایج جست‌وجو در فرایند بازیابی اطلاعات. *تحقیقات اطلاع‌رسانی و کتابخانه‌های عمومی*، ۷۵ (۲)، ۵۸۳-۶۰۷.

ستوده، هاجر؛ هنرجویان، زهره (۱۳۹۱). مروری بر دشواری‌های زبان فارسی در محیط دیجیتال و تأثیرات آنها بر اثربخشی پردازش خودکار. *کتابداری و اطلاع‌رسانی*، ۱۵ (۴)، ۵۹-۹۲

صدیقی، مهری (۱۳۸۳). بررسی کاربرد سیستم اطلاعات جغرافیایی (جی‌آی‌اس) در ساماندهی مدارک علوم زمین موجود در مرکز اطلاعات و مدارک علمی ایران. *علوم اطلاع‌رسانی*، ۲۰ (۱ و ۲)، ۲۹-۴۹.

صنعت‌جو، اعظم (۱۳۹۱). عملکرد هستی‌شناسی‌ها در نظام‌های بازیابی اطلاعات. *کتاب ماه کلیات*، ۱۷ (۶)، ۴۳-۴۷.

کاظم‌پور، زهرا؛ فهیم‌نیا، فاطمه (۱۳۹۱). وزن‌دهی به اصطلاحات و نقش آن در بازیابی اطلاعات. *نظام‌ها و خدمات اطلاعاتی*، ۱ (۴)، ۹۵-۱۰۴.

فیروزی، محمدعلی؛ سجادیان، ناهید؛ و سجادیان، مهیار (۱۳۹۰). سیستم پشتیبانی تصمیم‌گیری فضایی مدیریت ریسک بحران‌های طبیعی در روستاها با بهره گیری از جی.آی.اس، گامی در راستای توسعه پایدار: مطالعه موردی روستاهای استان مازندران. *فصلنامه روستا و توسعه*، ۵۴ (۲)، ۹۳-۱۱۵.

پورنقی، رویا (۱۳۹۲). مدیریت مجموعه منابع کتابخانه با استفاده از سیستم اطلاعات مکانی (مطالعه موردی: کتابخانه مرکزی و مرکز اسناد دانشگاه تهران). *کتابداری و اطلاع‌رسانی*، ۴۷ (۳)، ۲۵۱-۲۷۲.

مظفری، سعید؛ اروجی، علی‌اصغر؛ و مرادی، محی‌الدین (۱۳۹۲). آشکارسازی و تعیین مکان متون فارسی - عربی در تصاویر ویدئویی. *دوفصلنامه پردازش علائم و داده‌ها*، ۱۰ (۲)، ۸۷-۱۰۴.

Abresch, J. (Ed.). (2008). *Integrating geographic information systems into library services*: *A guide for academic libraries*. Hershey: IGI Global.

Andreasen, T., Nilsson, J. F., & Thomsen, H. E. (2001). *Ontology-based querying*. Physica-Verlag HD. Retrieved July 04, 2015, from http://ceur-ws.org/Vol-71/Andreasen.pdf

Asadi, S., Chang, C. Y., Zhou, X., & Diederich, J. (2005). Searching the world wide web for local services and facilities: A review on the patterns of location-based queries. In *Advances in web-age information management* (pp. 91-101). Springer Berlin Heidelberg.

www.SID.ir

Archive of SID

Asadi, S., Xu, J., Shi, Y., Diederich, J., & Zhou, X. (2006). Calculation of target locations for web resources. In *Web information systems–WISE* 2006 (pp. 277-288). Springer Berlin Heidelberg.

Asadi, S., Yang, G., Zhou, X., Shi, Y., Zhai, B., & Jiang, W. W. R. (2008). Pattern-based extraction of addresses from web page content. In *Progress in WWW Research and Development* (pp. 407-418). Springer Berlin Heidelberg.

Asadi, S., Zhou, X., & Yang, G. (2009). Using local popularity of web resources for geo-ranking of search engine results. *World Wide Web*, *12* (2), 149-170.

Asadi, S., Zhou, X., Jamali, H., & Mofrad, H. (2007). Location-based search engines tasks and capabilities: a comparative study. *Webology*, *4* (4). Retrieved July 10, 2015, from http://www.webology.org/2007/v4n4/a48.html

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval* (Vol. 463). New York: ACM press.

Bhogal, J., Macfarlane, A., & Smith, P. (2007). A review of ontology based query expansion. *Information processing & management*, *43* (4), 866-886.

Borges, K. A. (2006). *Use of an ontology of urban places for recognition and extraction of geospatial evidences on the web*. Unpublished doctoral dissertation, Belo Horizonte (MG), Brazil, Federal University of Minas Gerais.

Borges, K. A., Laender, A. H., Medeiros, C. B., & Davis Jr, C. A. (2007). *Discovering geographic locations in web pages using urban addresses*. In Proceedings of the 4th ACM workshop on Geographical information retrieval, November 09-09, (pp. 31-36). ACM New York, NY, USA.

Can, L., Qian, Z., Xiaofeng, M., & Wenyin, L. (2005, April). Postal address detection fromweb documents. In *Web Information Retrieval and Integration*, 2005. *WIRI*'05. Proceedings International Workshop on Challenges in Web Information Retrieval and Integration (pp. 40-45). IEEE Washington, DC, USA.

Clough, P., & Sanderson, M. (2004). A proposal for comparative evaluation of automatic annotation for geo-referenced documents. In *proceedings of Workshop on Geographic Information Retrieval SIGIR*. Sheffield. Retrieved July

www.SID.ir

Archive of SID

01, 2015, from http://www.geo.uzh.ch/~rsp/gir/abstracts/clough.pdf

De Andrade, F. G., de Souza Baptista, C., & Schiel, U. (2012). A Temporal Search Engine to Improve Geographic Data Retrieval in Spatial Data Infrastructures. In ICEIS Proceedings of the 14th *International Conference on Enterprise Information Systems*, Volume 1. Wroclaw, Poland, 28 June - 1 July, (pp. 56-65). SciTePress.

Dietterich, T. G. (2002). Machine learning for sequential data: A review. In *Structural*, *syntactic*, *and statistical pattern recognition* (pp. 15-30). Springer Berlin Heidelberg.

Egenhofer, M., & Kuhn, W. (1999). Interacting with geographic information systems. In *Geographical information systems*: *principles*, *techniques*, *applications*, *and management* (*pp*. 401-412). D. Rhind, Wiley: New York,.

Freitag, D. (1998). *Information extraction from HTML*: *Application of a general machine learning approach*. In AAAI/IAAI Proceedings of the Fifteenth National Conference on Artificial Intelligence, Madison, Wisconsin, July 26–30, (pp. 517-523). California: AAAI Press.

Hearst, M. A. (1992, August). *Automatic acquisition of hyponyms from large text corpora*. In Proceedings of the 14th conference on Computational linguistics Nantes, France, August 23-28, 1992 (pp. 539-545). Association for Computational Linguistics.

Holmes, D. O. (1990). Computers and geographic information access. *Meridian*, *4*, 37-49.

Jones, C. B., Alani, H., & Tudhope, D. (2001). Geographical information retrieval with ontologies of place. In *Spatial information theory* (pp. 322-335). Springer Berlin Heidelberg.

Kazama, K. (2012). Distribution pattern analysis of associated geographical names on transportation Network. *Transactions of the Japanese Society for Artificial Intelligence*, *27*, 34-39.

Kitayama, D., Matsuo, J., & Sumiya, K. (2013). Extracting relations among search properties based on the operational context of geographical information

www.SID.ir

Archive of SID

retrieval systems. In *Database Systems for Advanced Applications* (pp. 179-192). Springer Berlin Heidelberg.

Knappe, R., Bulskov, H., & Andreasen, T. (2007). Perspectives on ontology-based querying. *International Journal of Intelligent Systems*, *22* (7), 739-761.

Kumar, C. (2011). Relevance and ranking in geographic information retrieval. In *Proceedings of the Fourth BCS-IRSG conference on future directions in information access*, Koblenz, Germany, August 31, (pp. 2-7). UK: British Computer Society.

Lee, R., Wakamiya, S., & Sumiya, K. (2011). Discovery of unusual regional social activities using geo-tagged microblogs. *World Wide Web*, *14* (4), 321-349.

Loglisci, C., Ienco, D., Roche, M., Teisseire, M., & Malerba, D. (2012). An unsupervised framework for topological relations extraction from geographic documents. In *Database and expert systems applications* (pp. 48-55). Springer Berlin Heidelberg.

MacEachren, A. M. (2001). An evolving cognitive-semiotic approach to geographic visualization and knowledge construction. *Information Design Journal*, *10* (1), 26-36.

Martins, B., Silva, M. J., & Chaves, M. S. (2005). Challenges and resources for evaluating geographical IR. In *Proceedings of the* 2005 *workshop on Geographic information retrieval*, Bremen, Germany, October 31 - November 05, (pp. 65-69). New York: ACM.

Mata, F., & Claramunt, C. (2011). GeoST: Geographic, thematic and temporal information retrieval from heterogeneous web data sources. In *Web and Wireless Geographical Information Systems* (pp. 5-20). Springer Berlin Heidelberg.

McCurley, K. S. (2001, April). Geospatial mapping and navigation of the web. In *Proceedings of the* 10*th international conference on World Wide Web* (pp. 221-229). ACM.

Midaoui, O. E., Qadi, A. E., Rahmani, M. D., & Aboutajdine, D. (2015). A new approach to build a geographical taxonomy of adjacency automatically using

www.SID.ir

Archive of SID

the latent semantic indexing method. In *Intelligent Systems and Computer Vision* (*ISCV*), (pp. 1-6). IEEE.

Mikheev, A., Moens, M., & Grover, C. (1999). Named entity recognition without gazetteers. In *Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics*, Bergen, Norway, June 08-12, (pp. 1-8). Stroudsburg: Association for Computational Linguistics.

Muslea, I. (1999). Extraction patterns for information extraction tasks: A survey. In *The AAAI-99 Workshop on Machine Learning for Information Extraction*, Orlando, Florida, July 19. Retrieved July 15, 2015, from http://www.isi.edu/info-agents/RISE/ML4IE/ml4ie.muslea.ps

Nagabhushan, P., Angadi, S. A., & Anami, B. S. (2006). A fuzzy symbolic inference system for postal address component extraction and labelling. In *Fuzzy Systems and Knowledge Discovery* (pp. 937-946). Springer Berlin Heidelberg.

Perea-Ortega, J. M., Lloret, E., Ureña-López, L. A., & Palomar, M. (2013). Application of Text Summarization techniques to the Geographical Information Retrieval task. *Expert Systems with Applications*, *40* (8), 2966-2974.

Petasis, G., Cucchiarelli, A., Velardi, P., Paliouras, G., Karkaletsis, V., & Spyropoulos, C. D. (2000). *Automatic adaptation of Proper Noun Dictionaries through cooperation of machine learning and probabilistic methods*. In Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, Athens, Greece, July 24-28, (pp. 128-135). New York: ACM.

Purves, R., Clough, P., & Joho, H. (2005). *Identifying imprecise regions for geographic information retrieval using the web*. In Proceedings of the 13th Annual GIS Research UK Conference (pp. 313-18) Glasgow: University of Glasgow. Retrieved July 25, 2015, from http://ir.shef.ac.uk/cloughie/papers/gisuk05.pdf.

Sanderson, M., & Kohler, J. (2004). Analyzing geographic queries. In *SIGIR Workshop on Geographic Information Retrieval*. Ritrieved December 22, 2014,

www.SID.ir

Archive of SID

from: http://www.geo.uzh.ch/~rsp/gir/abstracts/sanderson.pdf (accessed).

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, *46*, 234-240.

Uryupina, O. (2003). *Semi-supervised learning of geographical gazetteers from the internet*. In Proceedings of the HLT-NAACL 2003 workshop on Analysis of geographic references Edmonton, Alberta, May 31, (pp. 18-25). Stroudsburg: Association for Computational Linguistics. Ritrieved December 25, 2014, from http://delivery.acm.org/10.1145/1120000/1119397/p18-uryupinaf?ip=194.225.91.24&id=1119397&acc=OPEN&key=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E6D218144511F3437&CFID=696778619&CFTOKEN=26626363&__acm__=1479917663_e26515b40095a24c92dba564eaade09c

Vidyarthi, V., Yadav, A., &Yadav, D. (2013). *New methodology in GIR systems*: *Improving web document searching*. In Sixth International Conference on Contemporary Computing, Noida, India, August 8-10, (pp. 208-2012). IEEE.

استناد به این مقاله:

رسولی، بهروز؛ اسدی، سعید (۱۳۹۵). بازیابی اطلاعات جغرافیایی در محیط وب: مروری بر نوشته‌ها و ترسیم مدل مفهومی. *مطالعات ملی کتابداری و سازماندهی اطلاعات*، ۲۷ (۴)، ۱۰۵-۱۲۶.

www.SID.ir