

دریافت مقاله: ۹۴/۳/۲۰

پذیرش مقاله: ۹۵/۳/۳۰

کشف سرویس‌های ابری در زبان فارسی از طریق تکامل هستان‌شناسی

نفیسه هراتیان اول^۱، علی اصغر صفایی^۲

^۱ گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد بین‌الملل قشم، قشم، ایران

haratiannafiseh@yahoo.com

^۲ گروه انفورماتیک پزشکی، دانشکده علوم پزشکی، دانشگاه تربیت مدرس، تهران، ایران

aa.safaei@modares.ac.ir

چکیده

یکی از ملاحظات اصلی در حوزه رایانش ابری، مسئله جست‌وجو کردن و کشف سرویس‌ها به منظور بهره‌برداری از آنهاست. از آنجا که یافتن سرویس‌های مناسب به صورت دستی و غیر خودکار، وقت‌گیر و بعضاً غیرممکن است، انجام خودکار و در صورت امکان افزودن معنا به آن می‌تواند به افزایش کیفیت و سرعت فرایند کشف سرویس‌ها کمک کند؛ البته به کارگیری این رویکرد نیاز به ایجاد شرح‌های معنایی برای همه سرویس‌های موجود دارد که این کار نیز اگر به صورت غیر خودکار انجام شود، کاری بسیار وقت‌گیر خواهد بود. به همین دلیل، پژوهشگران معمولاً از روش‌های پردازش زبان طبیعی برای خودکارسازی این فرایندها استفاده کرده‌اند.

ایجاد سیستمی که بتواند کشف سرویس‌های ابری را برای زبان فارسی، با کیفیت قابل قبولی انجام دهد، همچنان یک مسئله تحقیقاتی باز است. استفاده از هستان‌شناسی‌ها، برای معنایی نمودن کشف سرویس‌های ابری، می‌تواند به بهبود نتایج سیستم‌های کشف سرویس ابری کمک کند.

در این مقاله، سامانه‌ای برای کشف سرویس‌های ابری مبتنی بر تکامل هستان‌شناسی، برای زبان فارسی ارائه شده است. بدین صورت که معانی کلماتی را که کاربر برای جست‌وجوی یک سرویس در رایانش ابری به کار می‌برد، با به کارگیری تکنیک‌های تفسیر، هستان‌شناسی، استخراج و دسته‌بندی کرده، و از آن جهت جست‌وجو و یافتن سرویس مورد نیاز کاربر استفاده می‌کند. در زبان‌های دیگر، ابزارها و داده‌های مناسبی برای ارزیابی وجود دارد که ما در زبان فارسی این داده‌ها را در دسترس نداریم. بدین منظور از ضرایب دقت و فراخوانی برای ارزیابی سامانه پیشنهادی استفاده شده است و نتایج قابل قبولی را در مقایسه با پژوهش‌های پیشین نشان می‌دهد.

واژه‌های کلیدی: کشف سرویس، سرویس ابری، تکامل هستان‌شناسی، سیستم معنایی، موتور جست‌وجوی معنایی.

۱. مقدمه

متفاوت است. در سیستم ارائه‌شده در این مقاله، از نرم‌افزار فارسی‌نت [۱] به‌منظور اعمال هستان‌شناسی، و از مجموعه داده‌های معرفی‌شده در وب‌سایت دادگان [۲]، برای شناسایی کلمات توقف فارسی استفاده شده است.

در ادامه مقاله، ابتدا در بخش ۲، کارهای مرتبط بیان شده که در آن، برخی از پژوهش‌های مشابه در زبان‌های فارسی و انگلیسی معرفی شده است. بخش ۳ شامل سامانه پیشنهادی است که به معرفی معماری سامانه ارائه‌شده، مراحل آن و تشریح اجزای هر مرحله پرداخته است. بخش ۴ به ارزیابی روش ارائه‌شده و چالش‌های موجود در زمینه ارزیابی سیستم‌های کشف سرویس ابری پرداخته است. در نهایت، در بخش ۵ نتیجه‌گیری و پیشنهاداتی برای توسعه و تکامل بیشتر فعالیت‌های انجام‌شده ارائه شده است.

۲. کارهای مرتبط

بررسی پژوهش‌های انجام‌شده در زمینه کشف سرویس‌های ابری می‌تواند برای انجام پژوهش‌های جدید در این زمینه مورد استفاده قرار بگیرد. در این مقاله، برخی از مهم‌ترین پژوهش‌های انجام‌شده در زمینه کشف سرویس‌های ابری بررسی می‌شود. این پژوهش‌ها در دو بخش پژوهش‌های انجام‌شده در زمینه کشف سرویس‌های ابری برای زبان فارسی و پژوهش‌های انجام‌شده در زمینه کشف سرویس‌های ابری برای زبان انگلیسی مورد بررسی قرار می‌گیرد.

تاکنون سیستم‌های متفاوتی برای کشف سرویس‌های ابری ارائه شده که تقریباً در تمامی آن‌ها زبان انگلیسی مبنای بوده و تلاش‌های مختصری نیز در سایر زبان‌ها همچون اسپانیایی صورت گرفته است [۳].

در [۳] کنار سایر مفاهیم مطرح در محیط ابری، به مفهوم جدیدی بنام Discovery as a service (Daas) پرداخته شده است. آن‌ها چهارچوبی را معرفی کرده‌اند که امکان کشف سرویس‌های ابری را در محیط موبایل فراهم می‌سازد. در چارچوب معرفی‌شده توسط آن‌ها، ترجیحات کاربر و زمینه کاری وی نیز در نظر گرفته شده است. در ارزیابی این سیستم

در یک دهه اخیر، به دلیل مزایای فراوان رایانش ابری، استفاده از این محیط برای کاربران مورد توجه قرار گرفته است. به دلیل رشد سریع تعداد خدمات ارائه‌شده در محیط ابر، این مشکل برای کاربران پیش می‌آید که چگونه می‌توانند سرویس‌های ابری مورد نیاز خود را از میان انبوهی از سرویس‌های ارائه‌شده بیابند. این مسئله به نام کشف سرویس شناخته می‌شود. استفاده از هستان‌شناسی‌های به‌روزشده می‌تواند برای یافتن جدیدترین سرویس‌ها به کاربران کمک کند. به دلیل توسعه هستان‌شناسی با استفاده از کلمات کلیدی پرتکرار، پس از گذشت مدتی، یک هستان‌شناسی در دامنه خاص خواهیم داشت؛ برای مثال فرض کنیم تمامی سرویس‌های ابری اضافه‌شده به سیستم، در «حوزه فناوری‌های وب» باشند، در این صورت با توجه به فرایند تکامل هستان‌شناسی، پس از مدتی هستان‌شناسی سیستم دربرگیرنده مفاهیم و لغات «حوزه فناوری‌های وب» خواهد بود. بدین منظور بایستی هستان‌شناسی را به‌روز نگه داشت. عبارت «تکامل هستان‌شناسی» در پژوهش‌ها به موضوع به‌روزرسانی یک هستان‌شناسی جهت تطابق با تغییرات محیطی اشاره دارد.

در این مقاله یک سیستم معنایی برای کشف سرویس‌های ابری، در زبان فارسی پیشنهاد شده است که از طریق مفاهیم کلمات (معنی، مترادف، متضاد و...) و ارتباط بین آن‌ها فرایند جست‌وجو را انجام می‌دهد. همچنین با استفاده از هستان‌شناسی می‌توان شبکه‌ای از کلمات، معانی و مفاهیم و ارتباط بین آن‌ها را به‌صورت سلسله‌مراتبی سازماندهی کرده و با تکامل این هستان‌شناسی، امکان اعمال تغییرات جدید در هستان‌شناسی موجود را ایجاد کرد تا بتواند سازگاری بیشتری را با محیط ابر داشته باشد.

نکته حائز اهمیت دیگر آن است که در هر زبانی، باید از نرم‌افزارهای مخصوص آن زبان برای هستان‌شناسی، تکامل هستان‌شناسی و انجام پردازش‌های زبان طبیعی استفاده کرد (امکان استفاده نرم‌افزارهای زبان‌های دیگر برای زبان فارسی وجود ندارد). همچنین پایگاه داده لغات، تعداد لغات هر زبان و نحوه نگارش کلمات، چالش‌هایی است که برای هر زبان

دقت کشف سرویس نیز ارتقا یافته است. این مقاله فقط به محیط موبایل پرداخته و نرم‌افزارهای آن برای زبان فارسی مناسب نیست.

در [۴] به بررسی آخرین تکنولوژی‌ها و روش‌های کشف سرویس در محیط ابر پرداخته شده که کشف سرویس را از پنج دیدگاه مورد تجزیه و تحلیل قرار داده است: ۱. تکنیک‌های تصمیم‌گیری؛ ۲. مدل نمایش داده؛ ۳. پارامترها و ویژگی‌های سرویس‌های ابری؛ ۴. مفاهیم، اهداف. در پایان نیز پس از مقایسه و خلاصه‌سازی روش‌های کشف سرویس با استفاده از این پنج دیدگاه، مسائل و چالش‌های پژوهش در حوزه کشف سرویس بیان شده است.

در [۵] کشف سرویس مبتنی بر زمینه را معرفی کرده است. هدف این رویکرد کشف سرویس‌های باکیفیت با توجه به زمینه کاری و اطلاعات کاربر است. شرح درخواست کاربر و اطلاعات زمینه، تحلیل و بسط داده می‌شوند و سپس مورد جست‌وجو قرار می‌گیرند. در این مقالات از هستان‌شناسی و تکامل آن برای معنایی نمودن فرایند جست‌وجو استفاده نشده است.

مقاله‌ای که ایده اصلی این پژوهش از آن گرفته شده، در [۶] محیطی را برای کشف سرویس‌های ابری در زبان انگلیسی با استفاده از تکامل هستان‌شناسی‌ها، روش‌های معنایی و برخی از ابزارها و روش‌های پردازش زبان طبیعی معرفی کرده است.

در این مقاله، از هستان‌شناسی WordNet انگلیسی، ابزار GoodRelations و دانشنامه ویکی پدیا برای تکامل هستان‌شناسی، و برای انجام پردازش‌های زبان طبیعی از مجموعه ابزارهای GATE استفاده شده است که هیچ‌یک قابلیت استفاده برای زبان فارسی را ندارند.

تاکنون برای زبان فارسی نیز پژوهش‌هایی برای کشف سرویس‌های ابری انجام شده، ولی تعداد پژوهش‌های انجام‌شده برای زبان فارسی، در مقایسه با موارد مشابه انگلیسی بسیار اندک است.

در [۷] یک رویکرد نوین برای کشف سرویس آگاه به زمینه در شبکه‌های محاسبات فراگیر معرفی شده است. برای کشف سرویس‌ها، از اطلاعات زمینه کاربر و اطلاعات زمینه سرویس،

همچنین از هستان‌شناسی‌ها برای توصیف سرویس‌ها استفاده شده است.

در [۸] رهیافتی برای کشف سرویس‌ها در شبکه‌های موقت سیار MANET پیشنهاد شده است. راهکار پیشنهادی به صورت غیرمتمرکز بوده است و گروه‌های موجود دسته‌بندی شده‌اند. سپس برای هر دسته سرگروهی تعیین شده که با پوشه مرکزی در ارتباط است.

در [۹] برای کشف سرویس، علاوه بر ویژگی‌های عملکردی سرویس‌ها، خصوصیات غیر عملکردی آن‌ها (QoS) نیز در نظر گرفته شده است که تولیدکننده باید اطلاعات مربوط به خصوصیات غیر عملکردی سرویس‌ها را نیز ارائه کرده باشد. در هر یک از این مقالات، پژوهش‌های صورت گرفته منحصر به یک زمینه‌اند و از هستان‌شناسی و تکامل آن به صورت گسترده استفاده نشده است.

۳. سامانه پیشنهادی

ایده اصلی این مقاله، ارائه سیستمی مشابه [۶] برای زبان فارسی بوده است. در [۶] از WordNet برای هستان‌شناسی، از ابزار GoodRelations و دانشنامه ویکی پدیا برای تکامل هستان‌شناسی، از مجموعه ابزارهای GATE برای انجام پردازش‌های زبان طبیعی، در زبان انگلیسی استفاده شده است، درحالی‌که چنین ابزارهایی در زبان فارسی وجود ندارد.

در این پژوهش از WordNet فارسی یا فارسن‌ت برای ایجاد و تکامل هستان‌شناسی استفاده شده است. این مجموعه کامل‌ترین مجموعه برای زبان فارسی بوده و در نسخه مورد استفاده (نسخه ۲,۱) شامل بیش از ۳۰ هزار مدخل است. همچنین به جای استفاده از ابزار GoodRelations ابزاری شبیه به آن تولید شده است. در این ابزار همچون GoodRelations کلمات از طریق رابطه is_a یا is_subset_of (گره پدر و فرزند) با یکدیگر در ارتباط‌اند و در مجموع، یک ساختار سلسله‌مراتبی را ایجاد می‌کنند. در این پژوهش، ابزارهای مورد نیاز برای پردازش زبان طبیعی، همچون تقطیع کلمات، یافتن تشابه آن‌ها و... نیز تولید شده است.

الف. ثبت اطلاعات سرویس توسط کاربر

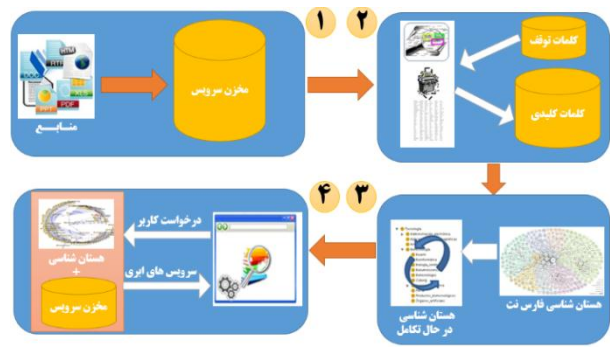
در این روش، اطلاعات سرویس‌های ابری توسط یک کاربر، در مخزن سرویس درج می‌شود. برای درج سرویس امکان خواندن اطلاعات با استفاده از قالب‌های استاندارد RDF وجود دارد. در زیر نمونه‌ای از اطلاعات ذخیره‌شده به وسیله RDF آورده شده است:

```
<?xml version="1.0"?>
rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-
"#rdf-syntax-ns
<"#xmlns:cd="http://www.sthsite.com/cd
rdf:Description>
rdf:about="http://www.sthsite.com/cd/google
">
<cd:title/>گوگل درایو</cd:title>
```

```
<cd:company/>گوگل</cd:company>
<cd:type>IaaS</cd:type>
<cd:details>
با وجودی که گوگل درایو پدیده‌ای جز همان سرویس اسناد
گوگل با ظاهری جدیدتر و امکان هماهنگ‌سازی با یک پوشه
خاص در سیستم عامل نیست، قابلیت‌هایی دارد که ارزش
بررسی کردن را به ما می‌دهد. نخست اینکه گوگل درایو
می‌تواند تا ۳۰ نوع فایل با فرمت‌های گوناگون را در محیط
مرورگر باز کند. این در حالی است که سرویس‌های دیگری
چون Dropbox و برخی دیگر تنها این کار را در محیط گالری
وب صورت می‌دهند.
```

```
<cd:details/>
از دیگر ویژگی‌های گوگل درایو امکان تعریف یک پوشه
جدید در سیستم عامل برای همگام‌سازی فایل‌ها، پشتیبانی از
قابلیت بگپرو رها کن (یا همان Drag and Drop خومان) است.
نقطه قوت گوگل درایو در محیط وب و یکپارچگی آن با دیگر
سرویس‌های گوگل یعنی جیمیل، سیستم عامل اندروید، شبکه
اجتماعی گوگل پلاس و... است.
```

```
<rdf:Description/>
rdf:Description>
rdf:about="http://www.sthsite.com/cd/drop
">
<cd:title/>دراپ باکس</cd:title>
<cd:company/>دراپ باکس</cd:company>
<cd:type>IaaS</cd:type>
<cd:details>
```



شکل (۱): معماری سامانه ارائه شده

در ادامه به معرفی سامانه ارائه‌شده، معماری و اجزای آن پرداخته می‌شود.

۱.۳. معماری سامانه ارائه‌شده

در این سیستم، زیرسیستم‌های زیر وجود دارد:

- زیرسیستم ۱: جمع‌آوری اطلاعات سرویس‌های ابری
- زیرسیستم ۲: استخراج کلمات کلیدی
- زیرسیستم ۳: تکامل هستان‌شناسی
- زیرسیستم ۴: موتور جست‌وجوی معنایی

که در ادامه هرکدام از این زیرسامانه‌ها به اختصار معرفی می‌شوند.

۱.۱.۳. زیرسیستم جمع‌آوری اطلاعات سرویس‌های ابری

در [۶] از ابزار آماده GoodRelations برای دستیابی به اطلاعات سرویس‌های ابری استفاده شده است. در این ابزار پس از ثبت اطلاعات یک سرویس یا محصول، با استفاده از ابزارهای پردازش زبان طبیعی برای هر سرویس حاشیه‌نویسی انجام می‌شود.

این ابزار برای زبان انگلیسی، ابزار بسیار مناسبی است و درخصوص سیستم‌های کشف سرویس معنایی بسیار مورد استفاده قرار می‌گیرد؛ ولی متأسفانه ابزار مناسبی در این زمینه برای زبان فارسی ارائه نشده است. در [۱۰] چارچوبی به نام FLOPPIES برای ایجاد یک مخزن معنایی از اطلاعات محصولات، با استفاده از داده‌هایی که ساختار جدولی دارند، معرفی شده که در این پژوهش، روشی ترکیبی شامل روش‌های ارائه‌شده در [۵ و ۱۰] پیاده‌سازی شده است. در این بخش اطلاعات مربوط به سرویس‌های ابری در یک پایگاه داده رابطه‌ای وارد می‌شود. این زیرسیستم می‌تواند به دو روش زیر، اطلاعات سرویس‌ها را جمع‌آوری کند:

۳.۱.۲.۱. تقطیع

اولین فعالیتی که در زیرسیستم استخراج کلمات کلیدی انجام می‌شود، تقطیع عنوان سرویس ابری و شرح آن است. تقطیع متون فارسی، یکی از چالش‌های اصلی پردازش زبان طبیعی است. برخی از مشکلاتی که تقطیع متون فارسی با آن درگیر است، عبارت‌اند از: وجود فاصله در برخی از کلمات مثل «درخت‌ها». این کلمات ممکن است به صورت درخت + ها تقطیع شوند. برای رفع این مشکل می‌توان هنگام درج اطلاعات در سیستم، کلماتی را که در آن‌ها فاصله وجود دارد، با نیم‌فاصله جایگزین نمود. وجود کلمات ترکیبی همچون «مانده‌مرخصی» یا «نرم‌افزار» که ممکن است به صورت دو کلمه مجزا در نظر گرفته شوند. در این زیرسیستم عنوان سرویس و شرح آن بر مبنای محل قرارگیری، فاصله، نقطه، ویرگول، پرانتز و... تقطیع می‌شود.

۳.۱.۲.۲. حذف کلمات توقف

در این بخش عنوان و شرح هر سرویس تجزیه می‌شود و سپس کلمات توقف یافت‌شده حذف می‌شوند. در این بخش، پس از تقطیع، به‌ازای هر سرویس ابری، لیستی از کلمات در اختیار خواهیم داشت. بسیاری از این کلمات، کلماتی هستند که هیچ تأثیری در فرایند معنایی نمودن سیستم ندارند؛ به این کلمات کلمات توقف گفته می‌شود. برخی از کلمات توقف رایج در زبان فارسی عبارت‌اند از: از، به، با، در، این، آن، و، است، که، ...

در پیاده‌سازی این بخش لیستی از کلمات توقف زبان فارسی، به کمک مجموعه داده‌های معرفی‌شده در وب‌سایت «دادگان»، به صورت دستی تهیه و از آن استفاده شده است.

۳.۲.۱.۳. استخراج کلمات کلیدی و اندیس‌گذاری

پس از حذف کلمات توقف، کلمات باقی‌مانده به عنوان کلمات کلیدی مربوط به سرویس ابری در نظر گرفته می‌شود. این کلمات در کنار سایر اطلاعات سرویس ابری درج می‌شود. سپس تعداد تکرار هر یک از آن‌ها در عنوان و شرح سرویس، محاسبه شده و در یک پایگاه داده مجزا ثبت می‌شود. این پایگاه داده‌ها برای شناسایی کلمات پر تکرار دامنه بوده و از آن برای تکامل هستان‌شناسی استفاده می‌شود.

سرویس کلاد دراپ باکس برای همگام‌سازی فایل‌ها در ابزارهای گوناگون به کار گرفته می‌شود. شما می‌توانید فایل‌های خود را در رایانه رومیزی درون پوشه مخصوص انداخته و آن را در تبلت، گوشی و لپ‌تاپ خود نیز داشته باشید. استفاده از خدمات Dropbox بسیار ساده و در حد کپی و جایگزاری ساده خودمان است. از این گذشته، برنامه رومیزی‌اش تنظیمات خاص و پیچیده‌ای ندارد به همین دلیل، بسیاری از کاربران حرفه‌ای تا مبتدی به سادگی می‌توانند با آن کار کنند. نقاط قوت Dropbox بعد از سادگی و سراسری‌اش، API آن به حساب می‌آید؛ که در دسترس بسیاری از توسعه‌دهندگان قرار گرفته که موجب خلق نرم‌افزارهای جانبی متعددی شده است.

```
<cd:details/>
<rdf:Description/>
<rdf:RDF/>
```

برای هر سرویس ابری اطلاعاتی دریافت می‌شود؛ مانند عنوان سرویس ابری، شرکت ارائه‌دهنده، نوع سرویس ابری (این فیلد، نوع سرویسی را که ارائه می‌شود مشخص می‌کند. برخی از مقادیر ممکن برای این فیلد عبارت‌اند از: PaaS, SaaS و IaaS) و شرح سرویس ابری.

شرح سرویس یک توضیح متنی درباره سرویس ابری است. این توضیح در بخش‌های بعدی تجزیه شده و کلمات کلیدی آن برای ایجاد هستان‌شناسی مورد استفاده قرار می‌گیرد.

ب. استفاده از خزنده

در حال حاضر اطلاعات سرویس‌های ابری بایستی به صورت دستی به سیستم اضافه شوند، درحالی‌که امکان اضافه کردن بخشی به نام خزنده به این زیرسیستم پیش‌بینی شده است. با اضافه شدن خزنده به سیستم، امکان جمع‌آوری خودکار اطلاعات سرویس‌ها و محصولات فراهم می‌شود. در این روش با استفاده از استاندارد OWL اطلاعات سرویس‌ها به صورت اتوماتیک جمع‌آوری می‌شود.

۳.۱.۲. زیرسیستم استخراج کلمات کلیدی

این زیرسیستم خود شامل بخش‌های مختلفی است که در ادامه به شرح آن‌ها پرداخته شده است.

۳.۱.۲.۱. شرح الگوریتم تکامل هستان‌شناسی

در الگوریتم ارائه‌شده برای تکامل هستان‌شناسی، از هستان‌شناسی فارسی نت استفاده شده است. نحوه کار به این صورت است که ابتدا لیست کلمات کلیدی استخراج‌شده در زیرسیستم استخراج کلمات کلیدی، دریافت می‌شود. سپس به‌ازای هر یک از این کلمات کلیدی، ابتدا در هستان‌شناسی سیستم جست‌وجو می‌شود. اگر کلمه کلیدی قبلاً به هستان‌شناسی سیستم اضافه شده بود، فعالیت دیگری برای آن کلمه انجام نمی‌شود؛ ولی اگر کلمه کلیدی از قبل به هستان‌شناسی سیستم اضافه نشده باشد، آن کلمه در هستان‌شناسی فارسی نت جست‌وجو می‌شود. اگر کلمه کلیدی در هستان‌شناسی فارسی نت وجود نداشت، خود کلمه کلیدی به‌عنوان یک گره جدید به هستان‌شناسی سیستم اضافه می‌شود. با اضافه کردن بخشی به فرایند تکامل هستان‌شناسی، این کلمه کلیدی به یک مفهوم در هستان‌شناسی تبدیل خواهد شد.

اگر کلمه کلیدی در هستان‌شناسی فارسی نت وجود داشت، آنگاه گره پدر آن (یعنی گرهی که نسبت به این گره، در سطح بالاتری قرار دارد و کلی‌تر است، به عبارت دیگر یعنی گرهی که مفهوم یافت‌شده، زیرمجموعه آن است). نیز جست‌وجو می‌شود. اگر گره پدر وجود داشت، گره یافت‌شده به همراه پدرش به هستان‌شناسی سیستم اضافه می‌شود و در صورتی که گره پدر یافت نشود، خود گره یافت‌شده به‌تنهایی به هستان‌شناسی سیستم اضافه می‌شود.

پس از اضافه شدن گره یا گره‌های جدید به هستان‌شناسی سیستم، فرایند به‌روزرسانی روابط کلمه‌های جدید، در هستان‌شناسی سیستم انجام می‌شود. در این مرحله، وجود ارتباط میان کلمات اضافه‌شده به هستان‌شناسی سیستم نسبت به سایر گره‌های قبلی آن بررسی می‌شود و روابط کشف‌شده اضافه می‌شوند. هدف از اضافه کردن گره پدر در الگوریتم تکامل هستان‌شناسی، بهبود فرایند توسعه هستان‌شناسی و فراهم نمودن بستری است که طی آن، روابط بین مفاهیم موجود در هستان‌شناسی سیستم سریع‌تر کشف شوند؛ برای مثال چنانچه کلمه‌ای همچون «نرم‌افزارهای سیستمی» در هستان‌شناسی سیستم وجود نداشته باشد، پس از جست‌وجو در هستان‌شناسی

۳.۱.۲.۴. بسط کلمات کلیدی و حاشیه‌نویسی

یکی از بخش‌های مهم سیستم ارائه‌شده، بسط کلمات کلیدی مربوط به هر سرویس است. برای این منظور، هر یک از کلمات کلیدی استخراج‌شده از سرویس‌های ابری، با استفاده از هستان‌شناسی فارسی نت بسط داده می‌شوند؛ به این صورت که کلمه کلیدی مورد نظر در فارسی نت جست‌وجو شده و در صورت وجود کلمه در فارسی نت، تمامی کلمات مترادف آن استخراج می‌شوند، سپس کلمات استخراج‌شده به همراه کلمات کلیدی مربوط به سرویس ابری، به‌عنوان حاشیه، برای سرویس‌ها ثبت می‌شود. از این حاشیه‌نویسی (ثبت کلمات کلیدی هر سرویس)، در زیرسیستم جست‌وجو، برای جست‌وجوی سریع‌تر استفاده شده است.

در این بخش، امکان اضافه کردن قابلیت دسته‌بندی سرویس‌ها براساس کلمات کلیدی آن‌ها، نیز پیش‌بینی شده است؛ به گونه‌ای که با اضافه کردن این قابلیت، فرایند جست‌وجوی سرویس با سرعت و کیفیت بیشتری صورت گیرد.

۳.۱.۲.۳. زیرسیستم تکامل هستان‌شناسی

در این زیرسیستم، فرایند تکامل هستان‌شناسی انجام می‌شود. برای این منظور کلمات کلیدی که در تمامی سرویس‌ها تعداد تکرار بیشتری داشته باشند، شناسایی شده و به هستان‌شناسی اضافه می‌شوند؛ که الگوریتم آن به‌صورت زیر است:

```
TermList = GetKeywordsList();
For (i = 0; i < TermList.Count; i++)
{
    Term = Term List.GetTerm(i);
    TermConcept= SystemOntology. Search (Term);
    if (TermConcept != null)
        return;
    else
    {
        TermConcept= FarsNetOntology. Search (Term);
        if (TermParent !=null)
        {
            TermParent=
            FarsNetOntology.SearchForParent(Term);
            if (TermParent !=null)
                SystemOntology.AddEntries(TermConcept
                , TermParent);
        }
        else
            SystemOntology.AddEntry(TermConcept);
    }
    else
        SystemOntology.AddRntry(Term);
}
SystemOntology.UpdateForNewNodes();
}
```

فارس‌نت، مفهوم «نرم‌افزار» که پدر «نرم‌افزارهای سیستمی» است نیز شناسایی می‌شود و هر دو گره به هستان‌شناسی سیستم اضافه می‌شود. در این صورت، پس از اضافه شدن این کلمات ممکن است ارتباطات جدید بیشتری کشف و اضافه شود. به دلیل توسعه هستان‌شناسی با استفاده از کلمات کلیدی پر تکرار، پس از گذشت مدتی، یک هستان‌شناسی در دامنه خاص خواهیم داشت؛ برای مثال فرض کنیم تمامی سرویس‌های ابری اضافه‌شده به سیستم، در «حوزه فناوری‌های وب» باشند، در این صورت با توجه به فرایند تکامل هستان‌شناسی، پس از مدتی هستان‌شناسی سیستم دربرگیرنده مفاهیم و لغات «حوزه فناوری‌های وب» خواهد بود.

۳.۱.۳. زیرسیستم موتور جست‌وجوی معنایی

کار اصلی این زیرسیستم دریافت درخواست کاربر و ارائه بهترین سرویس‌های ابری متناظر با نیاز اوست. این زیرسیستم خود شامل بخش‌های زیر است:

۱.۳.۱.۳. بخش دریافت، فهم و بسط درخواست کاربر

این بخش پس از دریافت درخواست کاربر، فعالیت‌های زیر را انجام می‌دهد:

- آن را تقطیع می‌کند؛
- کلمات توقف موجود در آن را حذف می‌کند؛
- با استفاده از هستان‌شناسی سیستم، کلمات کلیدی موجود در درخواست کاربر را بسط می‌دهد؛
- با استفاده از هستان‌شناسی فارس‌نت، سایر کلمات کلیدی را بسط می‌دهد.

۲.۳.۱.۳. بخش جست‌وجو

این بخش، از بخش قبلی دو لیست دریافت می‌کند:

- لیست کلمات کلیدی به‌دست‌آمده از هستان‌شناسی سیستم
- لیست کلمات کلیدی به‌دست‌آمده از هستان‌شناسی فارس‌نت

سپس باید این دو لیست را با حاشیه‌های نوشته‌شده برای هر سرویس مقایسه نمود و شباهت‌ها را اندازه‌گیری کرد. برای این کار دو بار جست‌وجو انجام می‌شود: یک بار جست‌وجوی کلمات کلیدی به‌دست‌آمده از هستان‌شناسی سیستم با حاشیه نوشته‌شده بر تمامی سرویس‌ها مقایسه می‌شود و در مرحله بعدی، کلمات کلیدی به‌دست‌آمده از هستان‌شناسی فارس‌نت با حاشیه‌های نوشته‌شده بر تمامی سرویس‌ها مقایسه می‌شود.

امتیازهای به‌دست‌آمده در محاسبه شباهت‌ها، برای کلمات کلیدی به‌دست‌آمده از هستان‌شناسی سیستم مقدار ۲،۱ و برای هستان‌شناسی فارس‌نت، ۱ است. به این معنا که «کلمات کلیدی موجود در هستان‌شناسی سیستم، در مقایسه با کلمات کلیدی موجود در هستان‌شناسی فارس‌نت، امتیاز بیشتری خواهند داشت.»

این شباهت از طریق محاسبه کلمات مشترک بین دو جمله و از طریق معیار فاصله Dice و به‌صورت زیر محاسبه می‌شود:

$$Dice(S1, S2) = (2 \times |S1 \cap S2|) / (|S1| + |S2|)$$

در این رابطه، S1 و S2 دو مجموعه کلمه‌ای هستند که می‌خواهیم میزان شباهت آن‌ها را با یکدیگر محاسبه کنیم. صورت کسر تعداد کلمات مشترک را مشخص می‌کند و در مخرج نیز تعداد کل کلمات موجود در دو مجموعه، قرار می‌گیرد؛ برای مثال، می‌خواهیم میزان شباهت دو مجموعه کلمه زیر را از طریق همین رابطه به‌دست آوریم:

- مجموعه کلمات اول: بهترین سرویس ارائه‌شده در وب برای نگهداری اطلاعات کاربران.
- مجموعه کلمات دوم: برای نگهداری اطلاعات خود در وب از این سرویس استفاده کنید.
- تعداد کلمات مجموعه اول: ۱۰
- تعداد کلمات مجموعه دوم: ۱۱
- تعداد کلمات مشترک: ۵

$$Dice(S1, S2) = \frac{2 \times 5}{10 + 11} = \frac{10}{21} = 0.476$$

بنابراین شباهت این دو جمله برابر ۰،۴۷۶ است. و این دو جمله بسیار به هم شبیه‌اند.

۳.۱.۳.۳. بخش رتبه‌بندی نتایج جست‌وجو

در این بخش سرویس‌های ابری یافت‌شده براساس امتیاز

۱. منظور از بسط دادن در این بخش، اضافه کردن کلمات مترادف آن کلمه، با استفاده از هستان‌شناسی است.

- برای زبان فارسی مراکزی به‌منظور ارائه مجموعه داده‌های استاندارد به کاربران راه‌اندازی شده‌اند:
 - مهم‌ترین این مراکز برای زبان فارسی مرکز دادگان است. این مرکز یک مرکز جامع جهت معرفی و در اختیار دادن فهرست دادگان زبان فارسی است. ولی در زمینه کشف سرویس‌های ابری، مجموعه داده استاندارد وجود ندارد.
- به دلیل عدم وجود مجموعه داده استاندارد برای کشف سرویس‌های ابری، از ضریب دقت، فراخوانی و معیار f جهت مقایسه سیستم‌های مختلف استفاده می‌شود.
- بدین منظور برای ارزیابی سامانه ارائه‌شده از فرمول زیر استفاده شد.

$$Precision^N(q) = \frac{|Rel_q \cap Rank_q^N|}{|Rank_q^N|}$$

$$Recall^N(q) = \frac{|Rel_q \cap Rank_q^N|}{|Rel_q|}$$

در این روابط Rel_q مجموعه سرویس‌های متناظر با درخواست q است. همچنین $Rank$ مجموعه‌ای از N سرویس بازیابی شده برای درخواست q است. برای معیار F هم از فرمول زیر استفاده شد. F معیاری است که تعادل بین دقت و صحت را نشان می‌دهد.

$$F - Measure = 2 * \frac{Precision * Recall}{Precision + Recall}$$

در پژوهش‌هایی که برای زبان انگلیسی انجام شده است، برای ارزیابی سیستم ارائه شده، از ابزارهایی همچون JDOM و SAWDL4J استفاده می‌شود. این ابزارها با در اختیار داشتن مجموعه داده‌های استاندارد، به ارزیابی خروجی سیستم‌های کشف سرویس می‌پردازند، ولی برای زبان فارسی چنین ابزارهایی وجود ندارد؛ به همین دلیل برای ارزیابی سامانه ارائه‌شده، عملکرد نسبی این سیستم‌ها شبیه‌سازی شده است. برای این منظور مجموعه‌ای از ۳۰

مشابهت خود با درخواست کاربر، به‌صورت صعودی مرتب می‌شوند. سپس تعداد مورد نیاز از این نتایج انتخاب می‌شوند تا در بخش بعدی در اختیار کاربر قرار بگیرند.

۴.۳.۱.۳. بخش ارائه نتایج به کاربر

در این بخش، نتایج مرتب‌شده از بخش رتبه‌بندی دریافت شده و در اختیار کاربر قرار می‌گیرد. در این بخش می‌توان تعداد نتایج نمایش داده‌شده به کاربر را براساس تعداد سرویس‌های ابری نمایش داده‌شده یا امتیاز مربوط به سرویس‌ها محدود کرد. در نسخه اولیه پیاده‌سازی‌شده، در صورتی که تعداد نتایج جست‌وجو بیشتر از ۱۰ عدد باشد، سرویس‌های ابری با امتیاز کمتر نمایش داده نمی‌شوند؛ به عبارت دیگر حداکثر ده سرویس دارای بیشترین امتیاز به کاربر نمایش داده خواهد شد. همچنین در صورتی که امتیاز یک سرویس از مقدار مشخصی کمتر باشد، نمایش داده نخواهد شد.

۴. ارزیابی روش پیشنهادی

۱.۴. معرفی محیط ارزیابی

چنانچه دو سیستم، مجموعه داده یکسانی داشته باشند، می‌توان آن‌ها را از طریق معیارهای گفته‌شده مقایسه کرد. به همین دلیل معمولاً در حوزه‌های مختلف پژوهشی، مجموعه داده استاندارد ارائه شده است تا پژوهشگران بتوانند با اجرای سیستم‌های خود بر روی آن‌ها، در پایان سیستم خود را با سایر سیستم‌ها مقایسه کنند.

برای زبان انگلیسی تاکنون مجموعه‌های داده استاندارد فراوانی تولید شده است:

- SEER یک مجموعه داده استاندارد برای پژوهش‌های حوزه پزشکی است.
- DUC و TAC مجموعه داده‌های استاندارد برای پژوهش‌های متن کاوی به‌خصوص خلاصه‌سازی خودکار متون است.
- مراکزی همچون data.gov, StatLib, datahub و gcind.nasa.gov به ارائه مجموعه داده‌های استاندارد می‌پردازند.

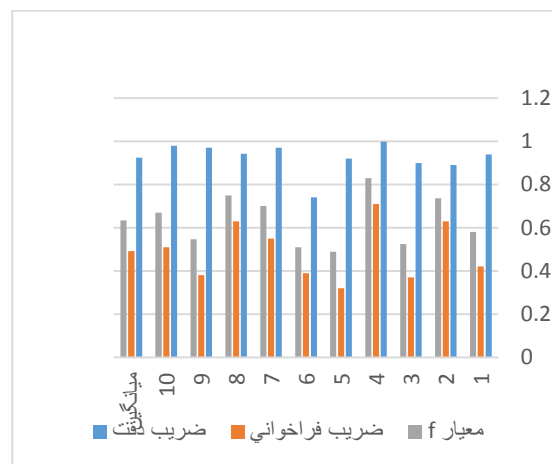
درخواست کاربر جمع آوری شد. برخی از این درخواست‌ها در جدول زیر نمایش داده شده است.

جدول (۱): درخواست‌های ایجاد شده برای ارزیابی سامانه

ذخیره اطلاعات در وب	سرویس خدمات حسابداری
نرم افزارهای کاربردی مورد استفاده در وب	فروشگاه خدمات نرم افزاری
سرویس انجام محاسبات ابری	سرویس مدیریت اسناد و مدارک
سرویس رندرینگ 3D Max	سرویس ارسال پیام کوتاه
سرویس ارسال ایمیل تبلیغاتی	پایگاه داده به عنوان سرویس
سرویس هاستینگ	سرویس ماشین مجازی
سرویس های خدمات حضور و غیاب	سرویس مدیریت ایمیل
خدمات مدیریت ارتباط با مشتری	خدمات برگزاری وبینار
سرویس راه اندازی فروشگاه مجازی	خدمات نظرسنجی پیامکی
برگزاری نظرسنجی ایمیلی	ارائه خدمات ثبت نام

۲.۴. نتایج ارزیابی

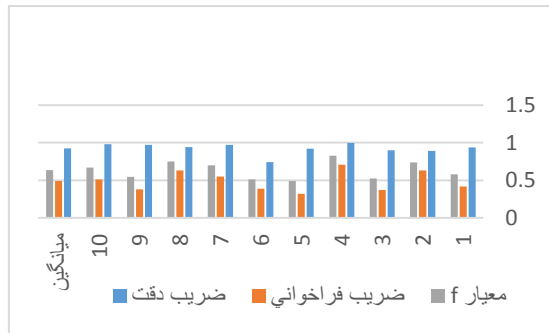
بدین منظور دو سناریو برای ارزیابی در نظر گرفتیم. در سناریوی اول N برابر ۷ است. میانگین نتایج به دست آمده در نمودار زیر آورده شد:



نمودار (۱): میانگین نتایج ارزیابی سامانه ارائه شده برای N=7

برای عنوان سرویس ابری اول دقت به میزان ۰/۹۴ به دست آمده که بدین معنی است که ۹۴ درصد سرویس های رایانش ابری پیشنهاد شده در لیستی هستند که متخصصان و کارشناسان آن را برای جست و جوی مرتبط، رشد و گسترش داده اند. امتیاز فراخوانی برای این عنوان، ۰/۴۲ است که بدین معناست که حدود ۵۸ درصد خدمات و سرویس هایی که در واقع نزدیک به جست و جوی صادر شده بودند، به وسیله سیستم

مورد بهره برداری قرار نگرفته اند. مقیاس F1 که برای این عنوان به دست آمده، ۰/۵۸ است. و در انتها میانگین این درخواست ها با ضریب دقت ۰/۹۲۵۱، نشان دهنده دقت بالاست. در سناریو دوم N را برابر ۳۰ در نظر گرفتیم. میانگین نتایج ارزیابی برای سامانه ارائه شده در نمودار زیر آمده است.



نمودار (۲): میانگین نتایج ارزیابی سامانه ارائه شده برای N=30

از مقایسه این دو نمودار، این نتیجه به دست می آید که هر چه مقدار N (تعداد درخواست هایی که به سیستم داده ایم) کمتر باشد (مثلاً ۵ یا ۱۰)، ضریب دقت بالاتر و ضریب فراخوانی پایین تری به دست می آید و هر چه N بزرگ تر باشد (مثلاً ۳۰ یا ۵۰)، ضریب دقت پایین و ضریب فراخوانی بالاتری به دست می آید؛ یعنی صفحات نامربوط را هم آورده است.

۵. نتیجه گیری و پیشنهاد کارهای آینده

کشف سرویس های ابری در حال حاضر، یکی از مهم ترین چالش های موجود در زمینه ارائه خدمات در محیط ابر است. در محیط ابر، برای موضوعات مختلف، سرویس های متعددی توسط ارائه دهندگان سرویس مختلف عرضه شده است. وجود سرویس های متعدد، برای این چالش را برای کاربران به وجود آورده است که چگونه می توانند سرویس های مورد نیاز خود را از میان انبوه سرویس های ارائه شده بیابند؟ تاکنون پژوهش های زیادی انجام شده که هدف اصلی آن ها بهبود کیفیت محیط های کشف سرویس ابری بوده است.

بسیاری از پژوهش های انجام شده برای زبان انگلیسی بوده است و تنها تعداد کمی از آن ها کشف سرویس های ابری را برای زبان فارسی، مدنظر قرار داده اند.

بخش‌ها داشته باشد، می‌توان عملکرد کلی سیستم را مورد بررسی قرار داد.

محدودیت‌های موجود در زمینه هستان‌شناسی‌های فارسی، عدم وجود ابزارهای آماده برای استفاده در سیستم، عدم وجود مجموعه داده استاندارد و...، نیز از جمله محدودیت‌هایی بوده که تأثیر مستقیمی در پیاده‌سازی سیستم داشته است. با تقویت این بخش‌ها می‌توان نتایج حاصل از سامانه ارائه‌شده را بهبود بخشید.

به‌منظور تکامل سامانه ارائه‌شده، اضافه کردن مؤلفه‌هایی به زیرسیستم‌ها پیشنهاد می‌شود. طراحی خزنده برای استخراج اطلاعات سرویس‌های ابری به‌صورت خودکار، پیاده‌سازی و استفاده از ابزارهای پردازش زبان طبیعی برای افزایش دقت زیرسیستم‌ها و تکامل پایگاه داده‌های کلمات توقف زبان فارسی و افزودن قابلیت یادگیری به آن، بومی‌سازی کامل ابزاری GoodRelations برای زبان فارسی و پیاده‌سازی ابزاری جهت استفاده از دانشنامه ویکی پدیا و سایر سیستم‌های مشابه برای زبان فارسی و استفاده از سایر روش‌های شباهت‌سنجی برای امتیازدهی.

پژوهش حاضر، محیطی برای کشف سرویس‌های ابری از طریق تکامل هستان‌شناسی معرفی کرده است. این محیط برای زبان فارسی تولید شده است و می‌تواند با دریافت درخواست کاربر، سرویس‌های مورد نیاز او را از مخزن سرویس‌بازایی نموده، و در یک رتبه‌بندی مشخص ارائه کند. چنانچه بخواهیم محیطی برای کشف سرویس‌های ابری برای زبان فارسی تولید کنیم، با چالش‌های گوناگونی روبه‌رو خواهیم شد که از آن جمله می‌توان به نبود هستان‌شناسی و مجموعه داده استاندارد برای این موضوع نام برد. همچنین سیستم‌های کشف سرویس‌های ابری از ابزارهای قدرتمندی استفاده کرده‌اند که متأسفانه نمونه‌ای برای آن‌ها در زبان فارسی وجود ندارد.

سامانه ارائه‌شده، یک محیط برای کشف سرویس‌های ابری است. مهم‌ترین نقطه قوت این سامانه، پیاده‌سازی آن برای کشف سرویس‌های ابری برای زبان فارسی است. همچنین استفاده از روش تکامل هستان‌نگاری و قابلیت‌های شاخص‌گذاری بر روی اطلاعات سرویس‌ها، سرعت کشف سرویس را بهبود بخشیده است.

پیاده‌سازی به‌صورت مرحله به مرحله و معماری پیمانه‌ای این سامانه از دیگر نقاط قوت آن است؛ چراکه با تغییر در یک بخش از سیستم، بدون آنکه کوچک‌ترین تأثیری بر سایر

مراجع

- [1] <http://nlp.sbu.ac.ir/farsnet/> Access Date: 1393/12/10
- [2] <http://dadegan.ir> Access Date:1393/12/10
- [3] K. Elgazzar, et al, "DaaS: Cloud-based Mobile Web Service Discovery", *Pervasive and Mobile Computing*, Volume 13, pp. 67–84, 2014.
- [4] A. Le Sun, et al, "Cloud service selection: State-of-the-art and future research directions". *Journal of Network and Computer Applications*, Volume 45, pp. 134 – 150, 2014.
- [5] M. Shang-Pin, et al, "Contextual service discovery using term expansion and binding coverage analysis", *Future Generation Computer Systems*, Volume 48, pp. 73 – 81, 2014.
- [6] M. Ángel Rodríguez-García, et al, "Creating a semantically-enhanced cloud services environment through ontology evolution", *Future Generation Computer Systems journal*, Volume 32, pp. 295–306, 2014.
- [7] بتول محمدی و همکاران، «بررسی روش‌های کشف سرویس آگاه به زمینه در شبکه‌های محاسبات فراگیر»، نهمین همایش کامپیوتر و سیستم‌های هوشمند، اسفند ۱۳۹۰
- [8] علی گلزاده و همکاران، «رهیافتی جدید مبتنی بر گروه‌بندی برای کشف سرویس در شبکه‌های فراگیر»، اولین کنفرانس ملی دانش‌پژوهان کامپیوتر و فناوری اطلاعات، ۱۳۹۰.
- [9] ندا محمدی، مهران محسن‌زاده و ماشاءالله عباسی دزفولی، «کشف سرویس مبتنی بر پارامترهای QoS و اعتبار سرویس»، سومین کنفرانس مهندسی برق و الکترونیک ایران، ۱۳۹۰
- [10] L.J. Nederstigt et al. "FLOPPIES: A framework for large-scale ontology population of product information from tabular data in e-commerce stores, *Decision Support Systems*", Volume 59 - Issue 1 p. 296- 311, 2014.