

طراحی سامانهٔ هوشمند ساخت هستان‌نگار به کمک شبکهٔ عصبی ART و روش C-value

مریم حورعلی^۱، غلامعلی منتظر^۲

^۱دانشجوی دکتراپی در فناوری اطلاعات، گروه مهندسی فناوری اطلاعات، دانشکدهٔ فنی و مهندسی، دانشگاه تربیت مدرس

^۲دانشیار مهندسی فناوری اطلاعات، گروه مهندسی فناوری اطلاعات، دانشکدهٔ فنی و مهندسی، دانشگاه تربیت مدرس

چکیده

در سال‌های اخیر تلاش‌های زیادی برای طراحی روش‌های یادگیری و خودکارسازی فرآیند ساخت هستان‌نگار انجام شده است. ساخت انواع هستان‌نگار برای انواع قلمروها و کاربردهای گوناگون فرآیند پرهزینه و زمان‌بر بوده و خودکارسازی این فرآیند گامی مهم در رفع مشکل اکتساب دانش در سامانه‌های اطلاعاتی و کاهش هزینه ساخت آنهاست. در این مقاله روشی نوین برای یادگیری هوشمند هستان‌نگار ارائه شده که می‌توان از آن در کاربردها و حوزه‌های مختلف استفاده کرد. در این روش نیازی به وجود هستان‌نگارهای عمومی یا تخصصی اولیه و واژگان معنایی از پیش تعریف شده نیست و پایگاه دانش اولیه آن، تنها شامل مجموعه‌ای از متون ورودی است. سامانهٔ یادگیرندهٔ پیشنهادی با شروع از متون ورودی و با استفاده از رهیافت پیشنهادی در این مقاله قادر خواهد بود هستان‌نگار حوزه‌های مختلف را استخراج کند. در این روش از ترکیبی از روش‌های زبانی، آماری و روش‌های یادگیری ماشینی بر اساس روش TF-IDF.C-value، شبکه عصبی نظریهٔ تشدید وفقی و روش تحلیل هم‌رخدادی استفاده شده است؛ بدین ترتیب که ابتدا اسناد مرتبط با حوزهٔ مورد نظر گردآوری شده و سپس پردازش‌های متون زبان طبیعی روی اسناد انجام شده و واژه‌های اصلی با استفاده از روش C-value استخراج شده است. آنگاه با استفاده از شبکه عصبی ART اسناد مربوطه خوشه‌بندی شده و برای هر خوشه با محاسبه وزن واژه‌ها بر اساس روش TF-IDF، واژهٔ کلیدی مناسب استخراج شده است. در پایان با استفاده از روش تحلیل هم‌رخدادی، سلسه‌های مراتب مفاهیم استخراج شده و هستان‌نگار مربوطه ساخته شده است. نتایج حاصل شده نشان می‌دهند که این روش در مقایسه با روش‌های مشابه، دقت خوبی در یادگیری هستان‌نگار داشته است.

واژگان کلیدی: هستان‌نگار، یادگیری، شبکه عصبی ART، فراوانی واژه‌ها-محکوس، فراوانی اسناد (TF-IDF)، C-value.

۱- مقدمه

یا «دید صریح از جهان همراه با تشریح مفاهیم و روابط آنها» (Shamsfard, 2003). تحقیق در زمینهٔ هستان‌نگار رشد فزآیندهای در حوزهٔ رایانه داشته است. هستان‌نگارها در زمینه‌های زیادی نظیر وب‌معنایی^۱، جویش‌گرها^۲، تجارت الکترونیکی، پردازش زبان طبیعی، مهندسی دانش، استخراج و بازیابی اطلاعات، طراحی دادگان^۳، سامانه‌های چندکارگزاره و کتابخانه‌های دیجیتال کاربرد دارند. مشکل عمده در ساختن هستان‌نگارها، اکتساب دانش و زمان‌بر بودن ساخت آنها برای کاربردهای مختلف است. بنابراین

هستان‌نگارها^۴ ابزار بیان رسمی مفاهیم و روابط موجود در یک قلمروی خاص هستند و در قلب کاربردهای وب‌معنایی قرار می‌گیرند. واژهٔ هستان‌نگار دارای معانی و کاربردهای مختلف در متون گوناگون است. در فلسفهٔ و زبان‌نگار، هستان‌نگار «مطالعهٔ هستی و آنچه در این جهان وجود دارد و رده‌بندی مفاهیم جهان هستی» است. کاربردی ترین تعریف هستان‌نگار در حوزهٔ هوش مصنوعی و فناوری اطلاعات چنین است: «توصیف صریح یک ادراک مشترک»

¹Ontologies

²Semantic web

³Search engines

⁴Database design





منطق مرتبه اول^{۱۰}، یادگیری قوانین مبتنی بر منطق مرتبه اول^{۱۱} و یادگیری گزارهای^{۱۲} برای استخراج دانش هستان نگار از رودی استفاده می‌شود. این روش‌ها دانش جدید را با استنتاج یا استقرا به دست آورده و آن را توسط گزاره‌ها و منطق مرتبه اول یا مراتب بالاتر نمایش می‌دهند (Shamsfard, 2003).

سامانه‌های یادگیری مبتنی بر استنتاج^{۱۳} نظیر Hasti، استنتاج منطقی و قوانین استنتاج را برای استخراج دانش جدید از دانش موجود به کار می‌برند، در حالی که سامانه‌های یادگیری مبتنی بر استقرا^{۱۴} (مانند WEB → KB) فرضیه‌هایی را از مشاهده‌ها (مثال‌ها) ایجاد کرده و دانش جدید را از تجربه به دست می‌آورند. این سامانه‌ها الگوریتمهای یادگیری مرتبه اول را به منظور یادگیری قوانین برای دسته‌بندی کردن صفحات، تشخیص روابط بین صفحات متعدد و استخراج مตون مشخص در صفحات وب به کار می‌برند (Shamsfard, 2004). در روش‌های زبانی روش‌هایی نظیر: تحلیل نحوی^{۱۵}، تجزیه الگویی نحوی-لغوی^{۱۶}، پردازش معنایی و درک متن برای استخراج دانش از مตون زبان طبیعی استفاده می‌شود. این روش‌ها اغلب به زبان وابسته هستند و به منظور استخراج دانش و ساخت هستان نگار بر روی متن پیش‌پردازش انجام می‌دهند. روش‌های مبتنی بر الگو کاربرد زیادی در زمینه استخراج اطلاعات دارند و در زمینه یادگیری هستان نگار نیز استفاده شده‌اند. در این روش‌ها، رودی به منظور یافتن کلید واژه‌های از پیش تعریف شده و الگوهایی که نشان‌دهنده برخی روابط هستند (نظیر ترادف) جستجو می‌شود (Shamsfard, 2003).

ایده استفاده از الگوهای نحوی برای استخراج روابط معنایی (به‌ویژه روابط رده‌بندی) توسط هیرست^{۱۷} معرفی این روش‌ها بیشتر روش‌های ابتکاری هستند که از قواعد منظم استفاده می‌کنند. در این رویکردها، متن برای یافتن نمونه الگوهای نحوی که بیان‌گر روابط خاصی نظیر رده‌بندی

^{۱۰}FOL(First Order Logic)-based clustering

^{۱۱}FOL rule-learning

^{۱۲}Propositional learning

^{۱۳}Deduction-based learning systems

^{۱۴}Induction-based learning systems

^{۱۵}Syntactic analysis

^{۱۶}Lexico-syntactic pattern-parsing

^{۱۷}Hearst

ساخت خودکار هستان نگار، راه حل مناسبی برای چیره‌شدن بر مشکل زمان‌بودن ساخت آنها و اکتساب دانش است (Shamsfard, 2004). یادگیری هستان نگار روشی است که برای استخراج دانش و ساخت هر چه آسان‌تر هستان نگار به کار می‌رود. با توجه به گسترش سریع وب‌معنایی و نیاز فزآینده به یافتن ارتباط معنایی میان اطلاعات موجود در وب، یادگیری هستان نگار یکی از زمینه‌های مهم تحقیقات مرتبط با وب‌معنایی است (Gerd, 2006).

رویکردهای مختلفی برای یادگیری هستان نگار وجود دارد. از یک رویکرد روش‌های یادگیری هستان نگار به دو روش آماری و روش‌های نمادین^۱ دسته‌بندی می‌شود. روش‌های نمادین شامل روش‌های منطقی^۲، زبان‌شناختی^۳ و مبتنی بر الگو^۴ هستند (Shamsfard, 2003). از رویکردهای دیگر، روش‌های یادگیری به روش‌های آماری، مبتنی بر قاعده^۵ و ترکیبی^۶ تقسیم می‌شود. ضمن اینکه روش‌های ابتکاری^۷ برای تسهیل هر یک از رویکردها به کار می‌روند (Zhou, 2007).

در روش‌های آماری تحلیل آماری بر روی داده‌های ورودی اعمال می‌شود. برای مثال سامانه WEB → KB از رویکرد آماری بسته‌ای از لغات^۸ برای دسته‌بندی صفحات وب استفاده می‌کند، سامانه‌های DoDDE II و Text-To-Onto از تحلیل آماری داده‌ای هم‌رخداد برای یادگیری روابط مفهومی در استناد استفاده می‌کنند. ایده اصلی این روش‌ها در این است که معنای یک لغت از روی پراکندگی آن در اسناد مختلف مشخص می‌شود، بنابراین معنای یک لغت وابسته به لغات هم‌رخداد با آن است. در این روش‌ها ابتدا ساختاری نظیر ماتریس ایجاد و با استفاده از تحلیل آماری ساختار، روابط مفهومی بین مفاهیم استخراج می‌شود (Cimiano, 2006). در روش‌های منطقی از روش‌هایی نظیر: برنامه‌نویسی منطق استنتاجی^۹، خوشه‌بندی مبتنی بر

^۱Symbolic

^۲Logicalmethods

^۳Linguistic-basedmethods

^۴Pattern based/template drivenmethods

^۵Rule-basedmethods

^۶Hybridmethods

^۷Heuristic

^۸Bag-of-words

^۹Inductive Logic Programming (ILP)



جنبه‌های نوآوری این تحقیق از دو بعد قابل بررسی است: یکی معماری سامانه پیشنهادی ساخت هستان نگار و استفاده از روشی ترکیبی در مراحل طراحی، ساخت و ارزیابی هستان نگار و دیگری از بعد کاربرد که در آن سعی شده است هستان نگار حوزه بسط پرسمان، برای اولین بار، استخراج شود.

ادامه این مقاله به صورت زیر سازماندهی شده است: در بخش دوم، مسأله مقاله تبیین شده و مفاهیم مرتبط نظری شبکه عصبی ART، روش استخراج واژه‌ها و C-value، روش وزن‌دهی TF-IDF و استخراج سلسله‌مراتب مفاهیم ^۶براساس روش تحلیل هم‌رخدادی تشریح شده است. در بخش سوم، روش پیشنهادی برای یادگیری هستان نگار گام به گام تشریح شده و سپس در بخش چهارم نتایج تجربی و ارزیابی مختلف آنها ارائه شده است. بخش پایانی نیز به نتیجه‌گیری و تحقیقات آتی در این زمینه اختصاص یافته است.

۲- تبیین مسئله

تعدادی سند در حوزه بسط پرسمان داریم که می‌خواهیم مفاهیم اصلی بیان گر این حوزه و ارتباط معنایی بین آنها را مشخص کنیم تا بتوان در مراحل بعدی از آن برای جستجوی معنایی استناد مرتبط با این حوزه بهره برد. بدین منظور لازم است هستان نگاری در این حوزه طراحی شود که دربرگیرنده مفاهیم و ارتباط معنایی بین آنها باشد.

به‌منظور ساخت هستان نگار ابزارهایی نظری: شبکه عصبی نظریه تشدید وفقی، روش C-value، روش وزن‌دهی TF-IDF و تحلیل هم‌رخدادی استفاده شده که در ادامه به اختصار به مبانی هر یک اشاره می‌شود.

۱-۱- شبکه عصبی نظریه تشدید وفقی (ART)

شبکه عصبی ART در سال ۱۹۷۶ توسط استفان گراسبرگ^۷ و گیل کارپنتر^۸ ارائه شد. بعد از آن اشکال و مدل‌های مختلفی از شبکه ART هم از نوع بانتظارت و هم از نوع بدون نظارت ابداع شد. از انواع شبکه‌های ART بدون نظارت می‌توان به ART-1^۹ برای بردارهای

هستند پیمایش می‌شود(Shamsfard, 2003). در روش هیرست، الگوها به صورت دستی تعریف می‌شند که کاری زمان‌بر و همراه با خطای بود. مرین^۱ برای بهبود الگوهای نحوی از یادگیری ماشین استفاده کرد. همچنین از الگوریتم‌های خوشبندی مفهومی برای تشکیل مفاهیم و رده‌بندی در سامانه ASIUM استفاده شده است (Maedche, 2002). روش‌های ابتکاری به طور مستقل کار نمی‌کنند و برای پشتیبانی سایر روش‌ها استفاده می‌شوند؛ برای مثال سامانه InfoSleuth برخی از قوانین ابتکاری را برای قراردادن مفاهیم جدید در مکان مناسب خود در هستان نگار به کار برده و مفاهیم مرتبط با یک عبارت اسمی را در زیر مفاهیم اصلی (پدر) قرار می‌دهد (Shamsfard, 2003). سامانه‌هایی که بیش از یکی از اجزای هستان نگار را یاد می‌گیرند، اغلب روش‌های ترکیبی را به کار می‌برند و برای یادگیری اجزای مختلف از قوانین یادگیری متفاوتی استفاده می‌کنند. برای مثال سامانه Text-To-Onto از قوانین وابستگی، تحلیل رسمی مفاهیم و روش‌های خوشبندی استفاده می‌کند (Zhou, 2007) KB→WEB قوانین یادگیری، منطق مرتبه اول را همراه با یادگیری بیزی به کار می‌برد و Hasti ترکیبی از روش‌های منطقی، زبانی، مبتنی بر الگو و ابتکاری را استفاده می‌کند (Shamsfard, 2003).

در این مقاله روش جدیدی برای یادگیری هستان نگار براساس ترکیبی از روش‌های زبانی، آماری و خوشبندی در نظر گرفته شده و با به کارگیری آنها هستان نگاری در حوزه بسط پرسمان ساخته شده است. بدین منظور ابتدا اسناد مرتبط در این زمینه گردآوری شده است. در مرحله بعد پیش پردازش‌های اولیه متون زبان طبیعی نظری حذف کلمات توقف، پردازش زبانی و پردازش آماری روی اسناد انجام شده و به کمک روش C-value واژه‌های اصلی استخراج شده است. در مرحله بعد ماتریس «واژه‌ها-سنده» ساخته شده که با استفاده از شبکه عصبی نظریه تشدید وفقی (ART)^{۱۰} خوشبندی و به کمک روش «وزن‌دهی فراوانی واژه-معکوس فراوانی سنده» (TF-IDF)^{۱۱} واژه‌ای که بیشترین وزن را دارد، به عنوان نام خوش انتخاب شده است. در پایان با استفاده از روش تحلیل هم‌رخدادی^{۱۲} سلسله‌مراتب مفاهیم استخراج و هستان نگار مربوطه ساخته شده است. مهم‌ترین

¹Morin

²Query expansion

³Adaptive resonance theory network (ART)

⁴Term frequency-Inverse document frequency (TF-IDF)

⁵Co occurrence analysis

⁶Concept hierarchy induction

⁷Stephan Grossberg

⁸Gail Carpenter

می‌کند و هدف آن ایجاد فهرستی از واژه‌های است که به حوزهٔ مورد نظر مرتبط باشند (Syafullah, 2010).

در مطالعات انجام‌شده روش‌های زیادی به منظور یادگیری واژه‌ها مطرح شده است که می‌توانند به عنوان اولین قدم در یادگیری هستان‌نگار استفاده شوند (شیخ اسماعیلی، ۱۳۸۵). بسیاری از روش‌ها بر مبنای روش‌های زبانی و تعدادی دیگر بر مبنای روش‌های آماری هستند. در روش‌های زبانی بیشتر از روش‌های پردازش متن نظری: توکن‌ساز^۶، برچسب‌گذاری بخشی از کلام^۷ و تحلیل‌گر نحوی^۸ استفاده می‌شود. به عنوان مثال Text-to-Onto یکی از سامانه‌هایی است که روش‌های زبانی را برای استخراج فهرستی از واژه‌ها به کار می‌برد. سامانهٔ SVETLAN از تحلیل‌گر نحوی با نام Sylex بدین منظور استفاده می‌کند. در روش‌های آماری، تحلیل آماری بر روی داده‌های ورودی اعمال شده و واژه‌ها بر مبنای رتبه‌بندی آماری مشخص می‌شوند. بسیاری از این روش‌ها از روش‌های بازیابی اطلاعات^۹ برای استخراج و رتبه‌بندی واژه‌ها استفاده می‌کنند (Syafullah, 2010). روش یادگیری واژه‌ها، تأثیر بهزیابی در کیفیت هستان‌نگار نهایی خواهد داشت. یک واژه مناسب باید دو ویژگی توصیف و تمایز را داشته باشد. توصیف، به این معنی که واژه مناسب باید محتوای اطلاعاتی یک سند را به درستی بیان کند و تمایز، به این معنی که واژه‌ای مناسب است که یک سند را از سندهای دیگر متمایز کند (شیخ اسماعیلی، ۱۳۸۵).

روش C-value از جمله روش‌هایی است که ترکیبی از روش‌های زبانی و آماری را برای استخراج واژه‌های چند کلمه‌ای استفاده می‌کند. در این مقاله از این روش به منظور استخراج واژه‌ها استفاده شده که در ادامه جزئیات آن ارائه شده است.

دو دویی^{۱۰} (Grossberg, 1987) ART-2، ART-2A (Carpenter, 1987) نسخه سریع‌تری از ART-2 (Carpenter, 1989)، ART-3 (Carpenter, 1991) ART-2 Fuzzy ART (Grossberg, 1991) و از انواع شبکه‌های ART با نظرات می‌توان ARTMAP (Grossberg, 1991b) Fuzzy ARTMAP (Carpenter, 1992) Gaussian ARTMAP (Carpenter, 1992) و ARTMAP (Williamson, 1996) را نام برد. در این شبکه‌ها در ابتدا نمونه‌های آزمایشی وارد شبکه می‌شوند، سپس شبکه یاد می‌گیرد چگونه قوانین خوشبندی را شکل دهد. اگر داده جدید در خوشبندی موجود نباشد، شبکه خوشة جدیدی را ایجاد خواهد کرد (Chen, 2008). شبکه‌های ART دارای دو خصوصیت پایداری^{۱۱} و انعطاف‌پذیری^{۱۲} هستند. پایداری به معنی عدم نوسان یک الگو در مراحل مختلف آموزش بین خوشبندی مختلف و انعطاف‌پذیری به معنی توانایی شبکه در یادگیری الگوهای جدید در هر یک از مراحل یادگیری است (Gour, 2008).

این شبکه‌ها به گونه‌ای طراحی شده‌اند که به کاربر امکان می‌دهند تا درجهٔ شباخت الگوهایی را که در یک خوشه قرار می‌گیرند با تنظیم پارامتری به نام پارامتر مراقبت^{۱۳} کنترل کند. همچنین در این شبکه‌ها نیازی نیست تعداد خوشه‌ها از قبل تعیین شده باشند و می‌توان از پارامتر مراقبت برای تعیین تعداد مناسب خوشه‌ها استفاده کرد؛ هر چه این پارامتر بیشتر شود، تعداد خوشه‌ها نیز افزایش و اندازه آنها کاهش می‌یابد و بر عکس (Chen, 2008). در این مقاله از شبکه ART-1 برای خوشبندی اطلاعات استفاده شده و به همین دلیل این شبکه با توضیح بیشتری تشریح می‌شود: شبکه ART-1 برای خوشبندی نمونه‌هایی با مقادیر دودویی (صفر و یک) به کار می‌رود. این شبکه از دو لایه مقایسه^{۱۴} و بازشناسی^{۱۵} تشکیل شده که در شکل (۱) نشان داده شده است (Kumar, 2007). روند نمای الگوریتم شبکه عصبی ART-1 در شکل (۲) آورده شده است.

۲-۲-استخراج واژه‌ها

استخراج واژه‌ها یکی از لایه‌های یادگیری هستان‌نگار است که واژه‌های موجود در اسناد را به طور خودکار استخراج

⁶Tokenizer

⁷Part of speech (POS) tagging

⁸Syntactic analyzer (parser)

⁹Information retrieval

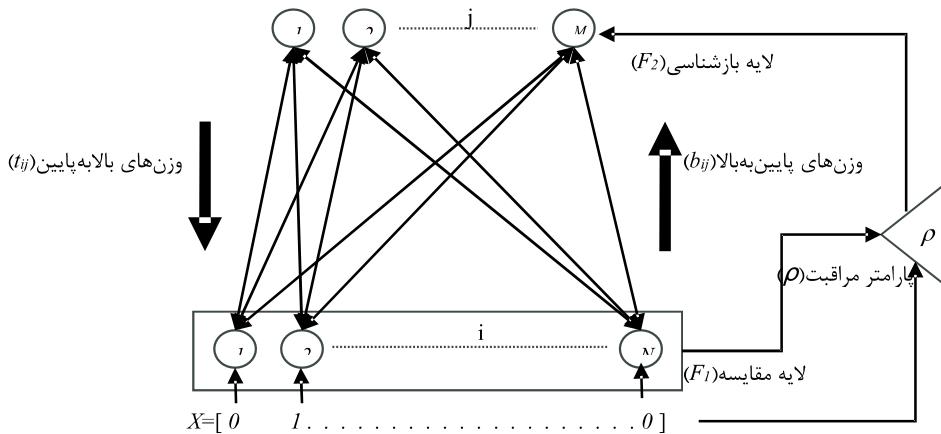
¹Stability

²Plasticity

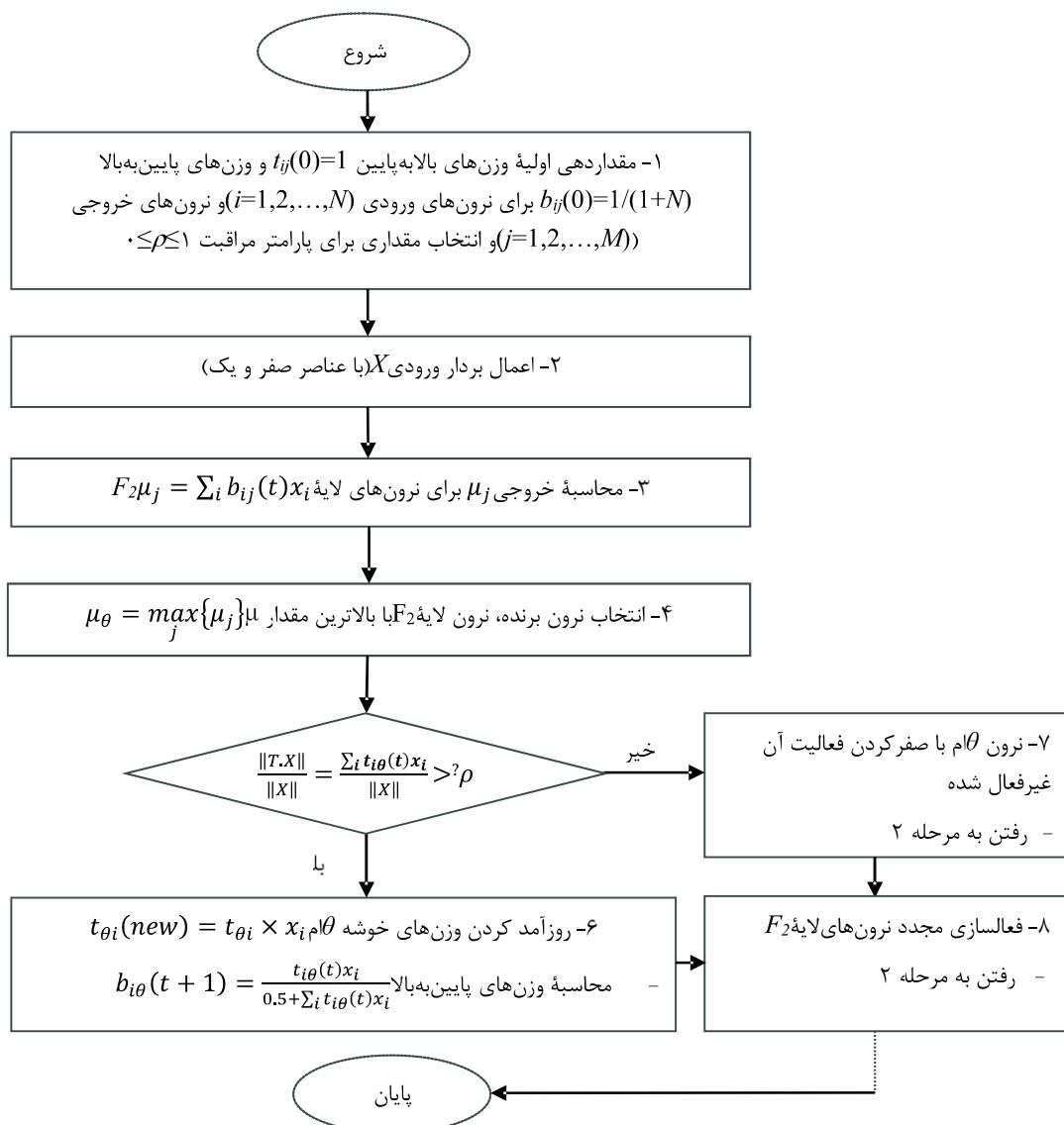
³Vigilance

⁴Comparison

⁵Recognition



شکل ۱- معماری شبکه عصبی ART-۱



شکل ۲- روند نمای الگوریتم شبکه عصبی ART-۱

که در آن $|a|$ طول رشته a ، $f(a)$ فراوانی رخداد رشته a در متن، T_a مجموعه رشته‌های نامزد استخراج شده شامل a ، $P(T_a)$ تعداد عناصر T_a و $\sum_{b \in T_a} f(b)$ مجموع فراوانی‌های (Frantzi, 2000) است که a در رشته‌های طولانی‌تر رخ می‌دهد.

به منظور درک بهتر نحوه محاسبه C-value، جدول (۱) را در نظر می‌گیریم که در آن چند رشته مربوط به حوزه فناوری اطلاعات و فراوانی آنها آورده شده است:

جدول ۱- رشته‌های شامل information technology

رشته	فراوانی
information technology	۲۰
information technology system	۸
information technology management	۷

در جدول (۱) رشته‌های information technology system و information technology management بیشترین طول را دارد و تودرتو هم نیست؛ بنابراین مقدار C-value آنها بر اساس رابطه (۱) به صورت زیر محاسبه می‌شود:

$$C\text{-value}(information\ technology\ system) = \log_2 |a| \cdot f(a) = \log_2 |3| \cdot 8 = 12.68$$

$$C\text{-value}(information\ technology\ management) = \log_2 |a| \cdot f(a) = \log_2 |3| \cdot 7 = 11.09$$

از آنجا که رشته a تو در تو بوده و در رشته‌های information technology system و information technology management آمده است، مقدار C-value آن بر اساس رابطه (۲) به صورت زیر محاسبه می‌شود:

$a = information\ technology, |a| = 2$

$$T_a = \{ information\ technology\ system, , information\ technology\ management \}$$

$$P(T_a) = 2, \sum_{b \in T_a} f(b) = 8 + 7 = 15$$

$$C\text{-value}(information\ technology) = \log_2 |a| \cdot (f(a) - \frac{1}{P(T_a)} \sum_{b \in T_a} f(b)) = \log_2 |2| \cdot (20 - \frac{1}{2} (8 + 7)) = 12.5$$

۳-۲- روشن C-value

روشی برای استخراج واژه‌های چندکلمه‌ای است که هدف آن بهبود استخراج واژه‌های تودرتو است. واژه‌های تودرتو آنها بیانی هستند که با واژه‌های طولانی تر ظاهر می‌شوند و ممکن است به تنهایی در متن رخ ندهند. به عنوان مثال واژه real time تودرتو است که با واژه‌های نظیر real time output, real time clock جمله مزیت‌های این روش در این است که دقت بیشتری از روش‌هایی که تنها از مقدار فراوانی به منظور استخراج واژه‌ها استفاده می‌کنند دارد و مزیت عمده دیگر آن توانایی استخراج واژه‌های تودرتو است. این روش واژه‌های چندکلمه‌ای استاد انگلیسی را با استفاده از روش‌های زبانی و آماری استخراج می‌کند. روش زبانی شامل برچسب‌گذاری بخشی از کلام، پالایه زبانی^۱ و سیاهه کلمات توقف^۲ است. روش آماری خصیصه‌های آماری را در نظر گرفته و C-value بر اساس آنها استخراج می‌کند (Frantzi, 2000). در ادامه جزئیات این روش تشریح شده است.

۲-۱- روشن آماری

مقدار آماری C-value، به رشته‌ها مقادیری را نسبت می‌دهد و بر این اساس، آنها را در سیاهه خروجی رتبه‌بندی می‌کند. به منظور محاسبه مقدار C-value رشته a ، دو حالت زیر در نظر گرفته می‌شود:

الف- اگر a رشته‌ای با بیشترین طول باشد یا تودرتو نباشد، مقدار C-value با استفاده از فراوانی کلی آن در متن و طول آن بر اساس رابطه (۱) بدست می‌آید که در آن $|a|$ طول رشته a و $f(a)$ فراوانی رخداد آن در متن است:

$$C\text{-value}(a) = \log_2 |a| \cdot (f(a)) \quad (۱)$$

ب- اگر a رشته‌ای تو در تو باشد، باید بررسی شود که آیا بخشی از واژه‌ها با طول بلندتر است. اگر چنین باشد، برای محاسبه مقدار C-value باید فراوانی آن به عنوان یک رشته تودرتو و تعداد واژه‌ها طولانی‌تر محاسبه شود. در این حالت مقدار C-value بر اساس رابطه (۲) محاسبه می‌شود:

$$C\text{-value}(a) = \log_2 |a| \cdot (f(a) - \frac{1}{P(T_a)} \sum_{b \in T_a} f(b)) \quad (۲)$$

¹Nested terms

²Linguistic filter

³Stop word



۵-۲- تحلیل هم‌رخدادی

سلسله‌مراتب مفاهیم، اطلاعات را به رده‌هایی، ساختاردهی می‌کند تا امکان جستجو، استفاده مجدد و درک آنها تسهیل شود. سامانه‌های دانش‌بنیاد^۳ با مشکل اکتساب دانش و بهویژه مدل‌سازی دانش حوزه مواجه هستند که در این موارد استخراج سلسله‌مراتب مفاهیم می‌تواند راه‌گشا باشد. نتایج برخی پژوهش‌ها نشان می‌دهد که رخداد برخی کلمات به معنی رخداد دیگر کلمات در جملات، پاراگراف یا اسناد مشابه است و رابطه مستقیمی بین آن دو کلمه وجود دارد.

این نظریه با نظریه هم‌مکانی مرتبط است و «تحلیل هم‌رخدادی کلمات»^۴ نامیده می‌شود. این تحلیل یکی از روش‌های استخراج سلسله‌مراتب مفاهیم است. دو کلمه را در صورتی «هم‌مکان» می‌گویند که در یک پاراگراف، جمله یا سند با هم رخ‌دهند یا نزدیک به هم بیشتر از حد تصادف ظاهر شوند. سندرسون^۵ و کرفت^۶ در سال ۱۹۹۲ تعریف رده‌بندی^۷ را بدین صورت ارائه کردند: واژه t_1 خاص‌تر از واژه t_2 است اگر t_2 در همه اسنادی که t_1 رخ می‌دهد، ظاهر شود. این رویکرد را می‌توان به شکل زیر تعمیم داد: واژه x شامل واژه y است اگر داشته باشیم:

$$p(y|x) < p(x|y) \quad (6)$$

که $P(x|y)$ به شکل زیر تعریف می‌شود:

$$P(x|y) = n(x,y)/n(y) \quad (7)$$

$n(x,y)$ تعداد اسنادی است که x و y هم رخ می‌دهند و $n(y)$ تعداد اسنادی است که شامل y هستند (Cimiano, 2006). در ادامه جزئیات ساخت هستن نگار با استفاده از مفاهیم مطرح شده در بالا آورده می‌شود.

۳- ساخت هستن نگار

فرآیند ساخت هستن نگار شامل مراحل: (۱) تحلیل اسناد، (۲) خوشه‌بندی اسناد، (۳) استخراج سلسله‌مراتب مفاهیم و ساخت هستن نگار و (۴) ارزیابی هستن نگار است. معماری سامانه پیشنهادی در شکل (۳) نشان داده شده که جزئیات آن در ادامه تشریح می‌شود:

۴-۲- روش وزن‌دهی فراوانی واژه-معکوس

(TF-IDF)

به‌طور کلی اهمیت یک کلمه در مجموعه اسناد با دو شاخص مشخص می‌شود: یکی فراوانی نسبی رخداد آن کلمه در سند که فراوانی واژه^۸ نامیده می‌شود و دیگری تعداد اسنادی دربرگیرنده آن سند که فراوانی سند^۹ نام دارد. بدیهی است اگر کلمه‌ای با فراوانی بالا در سندی رخ دهد، آن کلمه مهم‌تر از سایر کلمات در آن سند بوده و به عنوان کلمه کلیدی آن سند محسوب می‌شود. DF بیان گر نسبت اسناد دربرگیرنده آن کلمه در بین تمامی اسناد است. اگر فراوانی رخداد یک کلمه در تمامی اسناد نسبت به سند موجود کمتر باشد، بیان گر این است که آن کلمه، سند موجود را بهتر از دیگر اسناد متمایز می‌کند. برای محاسبه آنها ابتدا فراوانی واژه i در سند j (F_{ij}) محاسبه می‌شود و با هنجارکردن آن در تمامی مجموعه، مقدار TF به دست می‌آید یعنی:

$$TF_{ij} = F_{ij}/\max(F_{ij}) \quad (3)$$

اساس IDF بر این است که واژه‌هایی که در اسناد زیادی ظاهر می‌شوند، کمتر بیانگر موضوع کلی هستند. به همین دلیل برای محاسبه آن، ابتدا تعداد اسنادی که در برگیرنده واژه i هستند (n_i) و تعداد کل اسناد در مجموعه (N) مشخص شده، سپس IDF به صورت زیر محاسبه می‌شود:

$$IDF_i = \log(N/n_i) \quad (4)$$

در انتها، روش وزن‌دهی TF-IDF به صورت زیر محاسبه می‌شود (Mihalcea, 2009):

$$TF-IDF = TF_{ij} * IDF_i \quad (5)$$

این رابطه، نشان‌دهنده حاصل ضرب TF در IDF و بیان گر اهمیت یک کلمه در سند بوده و می‌توان بر اساس آن کلمه‌های موجود در اسناد را بر حسب میزان اهمیت آنها رتبه‌بندی کرد. در این مقاله از این روش به منظور تعیین میزان اهمیت کلمات در مجموعه اسناد استفاده شده که با به کارگیری آن می‌توان پس از خوشه‌بندی اسناد توسط شبکه عصبی از کلمه‌ای که بیشترین وزن را دارد برای انتخاب نام متناظر با آن خوشه استفاده کرد. همچنین از میزان اهمیت کلمات در ساخت سلسله‌مراتب هستن نگار نیز بهره برده‌ایم که جزئیات آنها در بخش ۳ تشریح شده است.

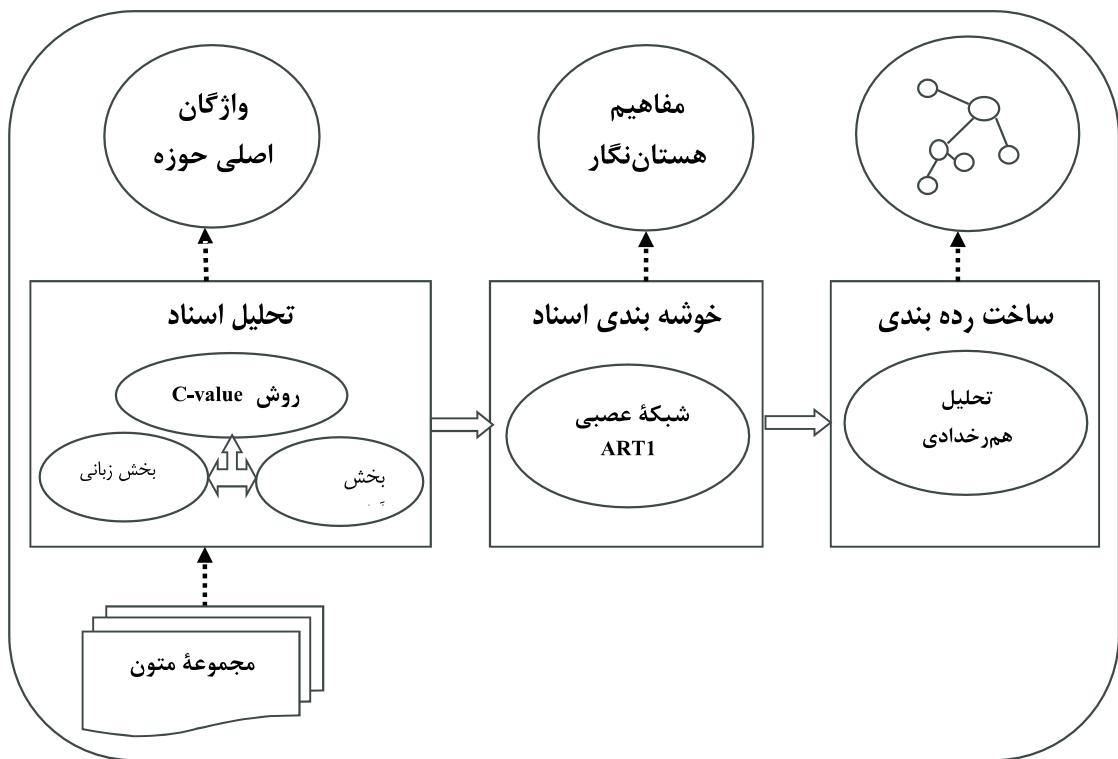
³Knowledge-based system

⁴ Terms co-occurrence analysis

⁵Sanderson

⁶Croft

⁷Taxonomy



شکل ۳- معما ری سامانه ساخت هستان نگار

منتشر شده در سال های ۲۰۰۷-۲۰۱۰ در نشریات معتبر مرتبی با این حوزه و شامل حداقل یک کلید واژه با عنوان "query expansion" مدت نظر قرار گرفته شد. پس از حذف مقالات غیر مرتبی، از حدود یکصد مقاله برای ادامه کار استفاده شد. هدف از این مرحله استخراج واژه های اصلی برای ساخت هستان نگار است که این کار به کمک روش C-value در دو بخش زبانی و آماری انجام شده است. در بخش زبانی ابتدا با استفاده از روش برچسب گذاری بخشی از کلام هر یک از اسناد برچسب گذاری شد تا برچسب های گرامری هر واژه در متن (نظیر اسم، صفت، فعل، حرف اضافه، ضمیر و ...) مشخص شود، سپس به منظور استخراج نوع خاصی از واژه های مورد نظر، پالیه های زبانی ارائه شده در بخش زبانی به متن برچسب گذاری شده اعمال شد. از آنجاکه پالیه شامل اسم یا صفت همراه با اسم هستند، نظیر user profile و automatic text processing شناسایی شوند. در مرحله بعد به منظور جلوگیری از استخراج کلماتی که به عنوان واژه یا کلمه مطرح نیستند، سیاهه کلمات توقف در نظر گرفته شد. در این سیاهه کلماتی نظیر several, great, the, of, a، ...، شناخته شده اند.

۳-۱-۳- تحلیل اسناد

این مرحله شامل سه زیر مرحله است: پیش پردازش اسناد، پردازش زبانی و پردازش آماری. هدف از این مرحله استخراج واژه های اصلی حوزه است. بدین منظور از روش C-value استفاده شده است. از جمله مزیت های این روش آن است که دقیق بیشتری از روش هایی که تنها از مقدار فراوانی به منظور استخراج واژه ها استفاده می کنند، دارد و مزیت عدم دیگر آن توانایی استخراج واژه های تودر تو است (Frantzi, 2000). علاوه بر این، این روش واژه های چند کلمه ای را استخراج می کند که واژه های چند کلمه ای معنای متمايز تر و مشخص تری نسبت به واژه های یک کلمه ای برای یادگیری واژگان هستان نگار هستند و برای مدل سازی دانش یک حوزه مناسب تر هستند (Drymonas, 2009). همچنین اطلاعات آماری و زبانی بیشتری برای استخراج واژه های چند کلمه ای وجود دارد (Nghia, 2011). بدین منظور در ادامه از این روش به منظور استخراج واژه های اصلی حوزه استفاده شده است.

از آنجاکه در این مقاله به دنبال ساخت هستان نگار در حوزه بسط پر سمان هستیم، نخست اسناد مرتبی در این حوزه، گردآوری شده است. بدین منظور مقاله های انگلیسی



۲-۳- خوشبندی اسناد

هدف از این مرحله، خوشبندی اسناد و یافتن گروههایی از اسناد است که دارای مشابهت با یکدیگر هستند. بدین منظور از شبکه عصبی ART1 استفاده شده است. این شبکه قادر است الگوهای ورودی با ترتیب دلخواه را به شکل پایدار، سریع و خودسازمانده یاد بگیرد؛ بنابراین می‌تواند بر مشکل یادگیری غیر پایدار که شبکه‌های رقابتی را دچار مشکل می‌کند غلبه کند. همچنین ART یک نظریه یادگیری است که در یادگیری شبکه تشدید ایجاد کرده و باعث یادگیری سریع می‌شود(Xu,2005). به همین دلیل مقادیر ماتریس واژه‌ها-سند به عنوان ورودی شبکه مورد استفاده قرار گرفته و خروجی شبکه مطابق روند نمای شبکه عصبی ART که در شکل (۲) نشان داده شده به دست آمده است.

به منظور تعیین مقدار مناسب پارامتر مراقبت، مقادیر آستانه^۱ مختلف آن از ۰/۱ تا ۰/۹ با ضریب افزایش ۰/۱ درنظر گرفته شد که میزان پارامتر مراقبت در آستانه ۰/۳ بهترین کیفیت خوشبندی را ارائه می‌داد. به این منظور این مقدار آستانه برای خوشبندی انتخاب شد که به استخراج ۲۴ خوشه منجر شده، ارائه شده است.

بدین ترتیب چند خوشه حاصل شد که هر کدام در برگیرنده تعدادی سند هستند. برای هر یک از واژه‌های موجود در اسناد خوشه‌ها وزن TF-IDF مطابق رابطه ۵ محاسبه شد و واژه‌ای که بیشترین مقدار را دارا بود، به عنوان نام آن خوشه (مفهوم هستان نگار) انتخاب شد. در جدول (۳) مفاهیم متناظر با خوشه‌های استخراج شده در جدول (۲) نشان داده است.

همان‌گونه که جدول (۳) نشان می‌دهد، برای برخی از خوشه‌ها مانند خوشه‌هایی که، شائزده و بیست مفهوم یکسان (relevance feedback) انتخاب شده است. به منظور انتخاب مفاهیم یکتا، خوشه‌های دارای نام یکسان ادغام شدند. درنهایت پانزده مفهوم منحصر به فرد به دست آمد که در مرحله بعد برای ساخت هستان نگار استفاده شد. در جدول (۴) خوشه‌های ادغام شده و مفاهیم نهایی استخراج شده آورده شده است.

numerous just year year قرار داده شد. در بخش آماری مقدار C-value واژه‌های استخراج شده در بخش زبانی بر اساس اینکه واژه‌های استخراج شده تودرتو بودند یا نبودند، بر اساس رابطه‌های (۱) یا (۲) محاسبه شد. بدین ترتیب در این مرحله ۴۱۱۶ واژه استخراج شد. به منظور استخراج واژه‌های اصلی، مقدار C-value را در بازه [۰,۱] هنجار کرده و واژه‌های دارای C-value هنجارشده بیش از ۰/۵ در محاسبات در نظر گرفته شدند. بدین ترتیب ۲۰۶۰ واژه در ساخت هستان نگار مذکور نظر قرار گرفت. تعدادی از واژه‌های استخراج شده در این مرحله در جدول (۲) آورده شده است. پس از استخراج واژه‌های اصلی، ماتریس واژه‌ها-سند ساخته شد که ردیف‌های آن بیان گر اسناد و ستون‌های آن بیان گر واژه‌های استخراج شده است. عناصر این ماتریس ۰ یا ۱ و بیان گر این هستند که واژه مورد نظر در آن سند بوده است یا خیر.

جدول ۲- نمونه‌ای از واژه‌های استخراج شده براساس C-value روش

واژه	C-value
query expansion	۰/۹۶۱
relevance feedback	۰/۹۲۲
semantic network	۰/۸۸۸
information processing	۰/۸۴۴
knowledge model	۰/۷۸۶
retrieval system	۰/۷۵۴
original query	۰/۷۵۳
query term	۰/۶۷۷
expansion term	۰/۶۷۷
information space	۰/۶۴۶
interactive query expansion	۰/۵۸۹
wordnet synset	۰/۵۸۳
knowledge resource	۰/۵۷۱
term co-occurrence	۰/۵۷۰
initial query	۰/۵۴۱
automatic query expansion	۰/۵۴۱
information retrieval	۰/۵۳۸
Term weighting	۰/۵۳۶
query reformulation	۰/۵۲۲

^۱ Threshold

query term reweighting	۱۹	natural language query expansion	۵,۲۳
content-based query expansion	۲۱	probabilistic query expansion	۶,۱۲
complex fuzzy query	۲۴	fuzzy query expansion	۷
		term weighting	۸

۳-۳ ساخت رده‌بندی

پس از استخراج مفاهیم هستان نگار، رده‌بندی مطابق روابط ۶ و ۷ استخراج می‌شود. ترتیب استخراج رده‌بندی بر حسب وزن مفاهیم است. بدین‌منظور، ابتدا وزن مفاهیم استخراج شده جدول (۴) مطابق رابطه (۵) محاسبه شده است. سپس مفهومی که بیشترین وزن را دارد، احتمال شرطی سایر مفاهیم به شرط آن مفهوم محاسبه می‌شود. اگر احتمال شرطی محاسبه شده از میزان آستانه بیشتر باشد، آن مفاهیم در زیر مفهوم با بیشترین افزوده می‌شود؛ در غیراین صورت دو مفهوم با بیشترین وزن در نظر گرفته می‌شود و این رویه تا زمانی ادامه می‌یابد که همه مفاهیم در سلسه‌مراتب هستان نگار جای گیرند. از آنجاکه میزان آستانه ۰/۷ بیشترین ساختار رده‌بندی را ارائه می‌داد، بنابراین این میزان آستانه برای استخراج رده‌بندی در نظر گرفته شده است. در بین مفاهیم استخراج شده، مفهوم query expansion technique مطابق زیر محاسبه می‌شود:

$$\begin{aligned} P(\text{queryreformulation} | \text{queryexpansiontechnion}) \\ = \\ (\text{تعداد استناد شامل مفاهیم query reformulation} + \text{تعداد استناد شامل مفاهیم query expansion technion}) \\ / (\text{تعداد استناد شامل مفهوم query reformulation}) \\ = \\ 6/8 = 0.75 \end{aligned}$$

به دلیل این که میزان این احتمال شرطی بیشتر از حد آستانه ۰/۷ است، مفهوم query reformulation به عنوان فرزند مفهوم query expansion technique در سلسه مراتب هستان نگار قرار می‌گیرد. بدین ترتیب سایر روابط نیز استخراج شده و سلسه‌مراتب هستان نگار کامل می‌شود. در

جدول ۳- مفاهیم هستان نگار متناظر با خوشها

مفهوم استخراج شده	خوش	مفهوم استخراج شده	خوش
query reformulation	۱۳	relevance feedback	۱
ontological query	۱۴	query expansion technique	۲
query expansion technique	۱۵	information retrieval system	۳
relevance feedback	۱۶	ontology-based query expansion	۴
word sense disambiguation	۱۷	natural language query expansion	۵
information retrieval system	۱۸	probabilistic query expansion	۶
query term reweighting	۱۹	fuzzy query expansion	۷
relevance feedback	۲۰	term weighting	۸
content-based query expansion	۲۱	query expansion technique	۹
information retrieval system	۲۲	search engine	۱۰
natural language query expansion	۲۳	ontology-based query expansion	۱۱
complex fuzzy query	۲۴	probabilistic query expansion	۱۲

جدول ۴- مفاهیم انتخاب شده نهایی برای

ساخت هسته‌ای نگار

مفهوم هستان نگار	خوشهاي ادغام شده	مفهوم هستان نگار	خوشهاي ادغام شده
Search engine	۱۰	relevance feedback	۱,۱۶,۲۰
query reformulation	۱۳	query expansion technique	۲,۹,۱۵
ontological query	۱۴	information retrieval system	۳,۱۸,۲۲
word sense disambiguation	۱۷	ontology-based query expansion	۴,۱۱



جدول ۵- مشخصات خبرگان شرکت کننده در تحقیق

رتبه علمی	سمت	رشته تخصصی	تحصیلات	خبره
دانشیار	استاد دانشگاه	کامپیوتر	دکترا	۱
دانشیار	استاد دانشگاه	مهندسی کامپیوتر-فناوری اطلاعات	دکترا	۲
دانشیار	استاد دانشگاه	کامپیوتر	دکترا	۳
استادیار	استاد دانشگاه	کامپیوتر	دکترا	۴
استادیار	استاد دانشگاه	کامپیوتر	دکترا	۵
-	دانشجو	کامپیوتر	دانشجوی دکترا	۶
-	مدیر پژوهشی	مهندسی کامپیوتر-فناوری اطلاعات	دانشجوی دکترا	۷
-	مدیر اجرایی	فناوری اطلاعات	کارشناس ارشد	۸
-	کارشناس ارشد	کامپیوتر	کارشناس ارشد	۹
-	کارشناس ارشد	کامپیوتر	کارشناس ارشد	۱۰

جدول ۶- نتایج ارزیابی هستان نگار توسط خبرگان

متغیرهای ارزیابی نظر خبرگان				
D	C	B	A	
۴	۱۱	۲	۱۳	خبره ۱
۲	۱۳	۲	۱۳	خبره ۲
۲	۱۳	۱	۱۴	خبره ۳
۵	۱۰	۴	۱۱	خبره ۴
۲	۱۳	۱	۱۴	خبره ۵
۲	۱۳	۲	۱۳	خبره ۶
۴	۱۱	۳	۱۲	خبره ۷
۲	۱۳	۱	۱۴	خبره ۸
۳	۱۲	۲	۱۳	خبره ۹
۵	۱۰	۰	۱۵	خبره ۱۰
۳.۱	۱۱.۹	۱۸	۱۲.۲	Miangchin نظرات
Average C_L_P= ۰/۷۹۳	Average C_P= ۰/۸۸۰			

شکل(۴) هستان نگار نهایی استخراج شده در حوزه بسط پرسمان نمایش داده شده است.

۴- ارزیابی هستان نگار

برای ارزیابی نتایج این تحقیق و مقایسه هستان نگار حاصل با نتایج واقعی از ده نفر از خبرگان این حوزه برای ارزیابی دقت هستان نگار ساخته شده، استفاده شد. دلیل انتخاب ده خبره، محدودیت افراد مسلط به حوزه بسط پرسمان و زمان بر بودن تکمیل پرسشنامه است. در جدول (۵) مشخصات خبرگان شرکت کننده در این تحقیق آورده شده است.

بدین منظور دو نوع روش ارزیابی دقت مد نظر قرار گرفت. دقت مفاهیم^۱ که بیان گر دقت کلمات کلیدی است که سامانه انتخاب می کند و دقت مکانی مفاهیم^۲ هم بیان گر دقت کلمات کلیدی (مفاهیم) انتخابی و هم نشان دهنده دقت مکان کلمات کلیدی در سلسه مراتب روابط است. روابط مورد استفاده برای این دو شاخص به شرح زیر است(Chen, 2008)

$$Precision (C_P) = A/A+B \quad (8)$$

$$Precision (C_L_P) = C/C+D \quad (9)$$

که متغیرهای A، B، C، D، مطابق زیر تعریف می شود:
A: واژه ها (مفاهیم) که سامانه تولید و خبره آنها را تأیید می کند.

B: واژه ها (مفاهیم) که سامانه تولید کرده، اما خبره آنها را تأیید نمی کند.

C: واژه ها (مفاهیم) که سامانه تولید و خبره مکان آنها را در سلسه مراتب هستان نگار تأیید می کند.

D: واژه ها (مفاهیم) که سامانه تولید و خبره مکان آنها را در سلسه مراتب هستان نگار تأیید نمی کند.

به منظور ارزیابی دقت سامانه، از هر یک از خبرگان درخواست شد مقادیر بالا را تعیین کنند و سپس میانگین نظرات آنها ملاک ارزیابی دقت هستان نگار قرار گرفت. در جدول (۶) نتایج ارزیابی خبرگان آورده شده است. به این ترتیب میانگین دقت مفاهیم (Average C_P) ده خبره برابر با ۰/۸۸۰ و میانگین دقت مکانی مفاهیم (Average C_L_P) ده خبره برابر با ۰/۷۹۳ به دست آمد.

¹Concept precision(C_P)

²Concept location precision(C_L_P)



شکل ۴- هستان نگار نهایی استخراج شده در حوزه بسط پرسمان

بین مفاهیم و استخراج هستان نگار فازی در سایر حوزه‌ها بهبود دهیم.

قدرتانی

بخشی از این تحقیق با حمایت مالی مرکز تحقیقات مخابرات ایران (طی قرارداد شماره ۲۰۱۲۷/۵۰۰) انجام شده که نگارندگان بر خود لازم می‌دانند از حمایت‌های آن نهاد پژوهشی تقدیر کنند.

مراجع

Carpenter, G., Grossberg , S., Invariant pattern recognition and recall by an attentive self-organizing ART architect-ture in a nonstationary world, The IEEE First International Conference on Neural Networks, 1987, PP. 737- 745.

Carpenter, G., Neural Network Models for Pattern Recognition and Associative Memory, Neural Networks, 1989, Vol.2, PP.243-257.

Carpenter, G., Grossberg, S., Rosen, D.B., ART2-A: An Adaptive Resonance Algorithm for Rapid Category Learning and Recognition, Neural Networks, 1991, Vol. 4, PP.493-504.

Carpenter, G., Grossberg, S., Markuzon, N., Reynolds, J.H., Rosen, D.B., Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps, IEEE Transactions on Neural Networks, 1992, Vol. 3, No. 5, PP.698-713.

۵- نتیجه‌گیری و ادامه تحقیق

این مقاله با هدف ارائه روش نوین در یادگیری هوشمند هستان نگار نگارش شده است. بدین منظور سامانه جدیدی طراحی و پیاده‌سازی شده که می‌تواند به منظور مدل‌سازی دانش حوزه‌های مختلف به کار برد شود. بدین ترتیب که با استفاده از روش‌های مختلف همچون روش C-value، شبکه عصبی نظریه تشدید وفقی، روش وزن دهی TF-IDF و تحلیل هم‌رخدادی امکان یادگیری هستان نگار حوزه‌های مختلف ایجاد شده و مدل‌سازی دانش آنها در قالب مفاهیم و روابط میان آنها امکان‌پذیر شود. از روش C-value به منظور استخراج واژه‌های اصلی، از شبکه عصبی ART1 به منظور خوشه‌بندی اسناد، از روش TF-IDF برای انتخاب مفهوم متناظر با خوشه‌های استخراج شده توسط شبکه عصبی و از تحلیل هم‌رخدادی در استخراج سلسه‌مراتب و ساخت هستان نگار استفاده شده است. در ارزیابی هستان نگار ساخته شده دو رویکرد بهره‌گیری از خبرگان حوزه و مقایسه با روش‌های مشابه در نظر گرفته شد و میانگین نظرات خبرگان ملاک ارزیابی دقت هستان نگار قرار گرفته شد. به این ترتیب میانگین دقت مفاهیم برابر با ۸۸٪ و میانگین دقت مکانی مفاهیم برابر با ۷۹٪ بود. هستان نگار استخراج شده ضعف‌هایی از این قبیل دارد که تنها روابط is_a را تولید می‌کند و سایر روابط رده‌بندی و غیر رده‌بندی هستان نگار را برای یافتن نمونه‌های مفاهیم، روابط چندگانه

Maedche, A., Staab, S.. Ontology Learning Part One - On Discovering Taxonomic Relations from the Web, Web Intelligence, Springer, 2002.

Mihalcea, R. Introduction to Information Retrieval, CSCE 5200 Information Retrieval and Web Search, 2009.

Shamsfard, M., Abdollahzadeh Barforoush, A., The state of the art in ontology learning: a framework for comparison, The Knowledge Engineering Review, 2003, Vol. 18, PP. 293–316.

Shamsfard, M., Abdollahzadeh Barforoush, A., Learning ontologies from natural language texts, Human-Computer Studies, 2004, Vol. 60, PP. 17–63.

Syafullah, M., Salim, N., Improving Term Extraction Using Particle Swarm Optimization techniques, Journal of computing, 2010, Vol. 2.

Williamson, J.R., Gaussian ARTMAP: A Neural Network for Fast Incremental Learning of Noisy Multidimensional Maps, Neural Networks, 1996, Vol. 9, No. 5, PP. 881-897.

Zhou, L. Ontology learning: state of the art and open issues, Inf Technol Manage, 2007, Vol. 8, PP. 241–252.

شیخ اسماعیلی، ک.، بازیابی اطلاعات در استناد وب معنایی، پایان نامه کارشناسی ارشد مهندسی کامپیوتر گرایش مهندسی نرم افزار دانشگاه صنعتی شریف، ۱۳۸۵.



مریم حور علی دوره کارشناسی خود را در سال ۱۳۷۸ در رشته ریاضی کاربردی در دانشگاه تهران گذرانده و مدرک کارشناسی ارشد خود را در سال ۱۳۸۵ در رشته مهندسی فناوری اطلاعات از دانشگاه تربیت مدرس اخذ کرده است. زمینه‌های علمی مورد علاقه‌وى وبمعنایی و بادگیری هستان نگار است.

نشانی رایانمأه ایشان عبارت است از:

Hourali@modares.ac.ir

Chen, R.C., Liang, J.Y., Pan, R.H., Using recursive ART network to construction domain ontology based on term frequency and inverse document frequency, Expert Systems with Applications, 2008, Vol. 34, PP. 488–501.

Chuang, C.H., Chen, R.C., Automating construction of a domain ontology using a projective adaptive resonance theory neural network and Bayesian network, Expert Systems, 2008, Vol. 25, No. 4.

Cimiano, P. “Ontology Learning and Population From Text Algorithms. Evaluation and Applications. Computational Linguistics”, The Association for Computer Linguistics, 2006, Vol. 45, PP. 888–895.

Frantzi, K., Ananiadou, S. , Mima. H., Automatic recognition of multi- word terms, International Journal of Digital Libraries, 2000, Vol. 3, No. 2, PP.115-130.

Gerd, S., Hotho , A., Berendt ,B., Semantic Web Mining-State of the art and future directions. Web Semantics: Science, Services and Agents on the World Wide Web, 2006, Vol. 4, PP. 124–143.

Gour, B., Bandopadhyaya, T. K., Sharma, S., ART Neural Network Based Clustering Method Produces Best Quality Clusters of Fingerprints in Comparison to Self Organizing Map and K-Means Clustering Algorithms, 2008, IEEE 978-1-4244-3397-1/08 .

Grossberg, S., Carpenter, G., A Massively Parallel Architecture for a Self-Organizing Neural Pattern Stephen Recognition Machine, Computer Vision, Graphics and Image Processing, 1987, Vol. 37, PP. 54-115.

Grossberg, S., Carpenter, G., Rosen, D.B., Fuzzy ART: Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System, Neural Networks, 1991, Vol. 4, PP. 759-771.

Grossberg, S., Carpenter, G., Reynolds, J.H., ART-MAP: Supervised Real-Time Learning and Classification of Nonstationary Data by a Self-Organizing Neural Network, Neural Networks, 1991b, Vol. 4, PP.565-588.

Kumar,N., Joshi R.S., Data Clustering Using Artificial Neural Networks, National Conference on Challenges & Opportunities in Information Technology (COIT-2007).



غلامعلی منتظر در سال ۱۳۴۸ در کازرون (فارس) به دنیا آمد. او در سال ۱۳۷۰ مدرک کارشناسی خود را در رشته مهندسی برق از دانشگاه صنعتی خواجه نصیرالدین طوسی و

سپس در سال ۱۳۷۳ و ۱۳۷۷ مدارک کارشناسی ارشد و دکتری خود را در همین رشته از دانشگاه تربیت مدرس اخذ کرد. وی پس از اتمام تحصیل به عضویت هیئت علمی دانشگاه تربیت مدرس درآمد و در حال حاضر دانشیار مهندسی فناوری اطلاعات در این دانشگاه است. حوزه‌های تخصصی وی شامل نرم‌رایانش (نظریه مجموعه‌های فازی، شبکه‌های عصبی مصنوعی، نظریه مجموعه‌های نادقيق) و کاربرد آن در سیستمهای اطلاعاتی (همچون سیستم یادگیری الکترونیکی و تجارت الکترونیکی) است. وی تاکنون بیش از ۷۰ مقاله در نشریات معتبر علمی و بیش از ۱۵۰ مقاله در کنفرانس‌های معتبر علمی ملی و بین‌المللی منتشر کرده است. علاوه بر این حائز دریافت جوایزی معتبر علمی از جمله «برگزیده جشنواره بین‌المللی خوارزمی»، « برنده کتاب سال دانشگاهی ایران »، « پژوهش گر برگزیده آیسیکو » و « متخصص برجسته فناوری اطلاعات ایران » شده است.

نشانی رایانمۀ ایشان عبارت است از:

montazer@modares.ac.ir