**نام مقاله: رده بندي خودكار: رويكردها و چالشهاي رده‌بندي خودكار**

**نام نشريه: فصلنامه كتابداري و اطلاع رساني (اين نشريه در www.isc.gov.ir نمايه مي شود)**

**شماره نشريه: 38 \_ شماره دوم، جلد 10**

**پديدآور: رضيه اسماعيل پور**

**مترجم:**

**چكيده**

**سازماندهي و مديريت منابع اينترنتي، همواره يكي از دغدغه‌هاي دانشمندان علوم رايانه و اطلاع‌رساني است. امروزه همين متخصصان به اين نتيجه رسيده‌اند كه علم كتابداري در زمينة رويارويي با وضعيت آشفتة اطلاعات مي‌تواند نقش برجسته‌اي ايفا كند. از آنجا كه رده‌بندي، فعاليتي وقت‌گير و پرهزينه بوده و رده‌بندي منابع اينترنتي براي فهرست‌نويسان دشوار مي‌باشد، تحقيقات زيادي براي انجام رده‌بندي به صورت خودكار صورت گرفته است. كاربرد شبكه‌هاي عصبي مصنوعي با توجه به ضعف روشهاي آماري و زبانشناسي رايانه‌اي در عملي كردن فعاليت رده‌بندي خودكار، رو به افزايش است.**

**كليدواژه‌ها: رده‌بندي خودكار، شبكه عصبي مصنوعي، اصطلاحنامه، سازماندهي منابع اينترنتي.**

**مسائل و مشكلات بازيابي اطلاعات در عصر جديد**

**عدم كنترل ورود اطلاعات به محيط وب و سرعت افزايش منابع اطلاعاتي از يك‌سو، و نبود راهكارهاي مناسب براي سازماندهي آنها، سبب شده بازيابي اين منابع با مشكلاتي از جمله ربط ضعيف، ريزش كاذب فراوان و صرف وقت زياد از كاربران، همراه باشد. فراواني توليد اطلاعات، آن‌قدر تأثيرگذار بوده كه باعث ورود اصطلاحات و لغات جديدي به زبان شده است: سيل اطلاعات، طوفان اطلاعات و اخيراً «Infoglut»[2] يك بيماري شناخته شده است كه جهانيان از آن رنج مي‌برند.**

**وقتي با انبوهي از كتابها روبرو بوديم، براي سازماندهي آنها به يك سيستم رده‌بندي نياز داشتيم تا بتواند با منظم كردن كتابها در هر ردة موضوعي، به بازيابي آنها سرعت بدهد. اكنون نيز با شكلهاي ديگر كتاب روبرو هستيم كه همان «فرمت الكترونيكي» است. اين‌گونه اطلاعات به علت قابليتهاي خاص خود سرعت نشر بيشتري نسبت به مواد چاپي دارند. همچنين، اكثر مواد چاپي، قبل از انتشار، به شكل الكترونيكي در دسترس هستند. زماني كه براي سازماندهي انبوهي از كتابها و بازيابي سريع آنها به يك سيستم رده‌بندي نياز داريم، ضرورت يك سيستم رده‌بندي مناسب براي اطلاعات روبه رشد ديجيتالي، بيشتر مشخص مي‌شود.**

**با انتشار رو به رشد انواع منابع اطلاعاتي به شكل الكترونيكي (مجله‌ها و روزنامه‌هاي الكترونيكي و پايگاههاي اطلاعاتي) و همچنين گرايش رو به رشد ميان رشته‌اي شدن علوم، همواره جستجوگران اطلاعات، با انبوه وسيعي از اطلاعات روبرو مي‌شوند. براي مثال وقتي كاربر، كليدواژه‌اي را در موتور جستجوي گوگل جستجو مي‌كند، حداقل بين 300 تا 400 هزار مدرك بازيابي مي‌شود كه جستجوي وي با خاص كردن آن كليدواژه، به يافتن 300 تا 400 مدرك مي‌انجامد. كاربر با وقت اندك خود، محال است كه بيشتر از 30 تا 40 مورد آن را ببيند؛ چون از نظر زماني مقرون به صرفه نيست كه وقت خود را صرف ديدن اطلاعات بسيار زيادي كند كه ممكن است به كار وي نيايد. در اين زمان، كاربر مطمئن هم نيست كه جستجوي وي كامل بوده يا نه، زيرا موفق به ديدن همة اطلاعات بازيابي شدة كليدواژه خود نبوده است. آمارها نشان مي‌دهند كاربران از ميان صفحات بازيابي شده در موتورهاي جستجو هميشه صفحات كمتري را مي‌بينند (Jansen & Spink, 2006). همچنين در يك مجموعه از صفحات بازيابي شده، مدخلهاي تكراري فراواني ديده مي‌شوند.**

**اهميت و ضرورت رده‌بندي اطلاعات در عصر جديد**

**با توليد روزافزون اطلاعات نياز مبرم به رده‌بندي كردن آنها براي بهينه‌سازي بازيابي اطلاعات، پررنگ‌تر مي‌شود. يافتن اطلاعات مورد نياز، تنها با جستجوي كليد واژه‌اي از طريق موتورهاي كاوش امكان‌پذير نيست. دانشمندان معمولاً با تورّق نشريات معتبر در زمينة علمي خود، بيشتر و راحت‌تر به اطلاعات مورد نياز خود دست مي‌يابند. زيرا اكثر مواقع، فردي كه به دنبال اطلاعات است، تعريف مشخصي از آنچه مي‌خواهد، ندارد و نمي‌تواند كليد واژه‌اي معيّن و مشخص را انتخاب و سپس بر اساس آن جستجو كند. بنابراين، افراد با تورّق مطالب بيشتر، امكان يافتن اطلاعات مورد نياز خود را دارند. امروزه فهرستهاي راهنما[3] به خوبي امكان تورّق را فراهم مي‌آورند؛ مثلاً فهرست راهنماي [4] ODP كه در نوع خود، بزرگترين فهرست جهان به شمار مي رود، توانسته با يك رده‌بندي موضوعي، ‌وب‌سايتهاي مناسبي را به شكل سلسله مراتبي نظم ببخشد. اين فهرست راهنما از ويراستاران[5] براي رده‌بندي صفحات وب استفاده مي‌كند.**

**وقتي اطلاعات به طور موضوعي رده‌بندي شوند، هر متخصص با روي آوردن به مطالب مربوط به تخصص خود، راحت‌تر به مطلب مورد نياز خود مي‌رسد و ديگر با مطالب نامربوط و انبوه زيادي از ركوردهاي بازيابي شده، روبرو نمي‌شود؛ گويي مشغول تورّق در مجله مورد نظر خود است.**

**در اينجا، مسئله رده‌بندي براي منابع ديجيتالي، حياتي به نظر مي‌رسد؛ به طوري كه اگر اين امر صورت نپذيرد يعني منابع ديجيتالي رده‌بندي نشوند، به علت عدم حضور فيزيكي مثل اين است كه گم شده باشند. (Review o attempts to apply Classification).**

**پايگاه اطلاعاتي نيز مثل يك كتابخانه است كه انبوهي از مدارك دارد. زماني كه كتابها رده‌بندي مي شوند، منابع مربوط كنار هم قرار مي‌گيرند، به عنوان مثال كتابهاي مربوط به تاريخ ايران در دوره قاجار، يك جا در يك قفسه كنار هم قرار مي‌گيرند و مي‌توان از ميان آنها كتاب مورد نظر را پيدا كرد. براي يافتن مطلب مورد نظر در يك پايگاه اطلاعاتي، يا مي‌توان با كليدواژه به جستجو پرداخت و يا زماني كه مدارك در اين پايگاه رده‌بندي موضوعي شده باشند، مي‌توان با ديدن موضوعات كلي چون كشاورزي، شيمي، زيست شناسي و علوم اجتماعي، يكي را انتخاب كرد و سپس با كليك كردن بر روي آن، موضوعات خاص‌تر آن را ديد. براي مثال، با كليك كردن بر روي «شيمي» ساختار ديگري براي شما نمايش داده مي‌شود: شيمي آلي، ‌شيمي معدني،‌ زيست شيمي و ... اين موضوع، تا جاي ممكن خاص مي‌شود، سپس با كليك بر روي موضوع خاص، مي‌توان مدارك مورد نظر خود ديد. (شكل 1).**

**Agriculture**

**Chemistry**

**Analytical Chemistry**

**Biological Chemistry**

**Geophysical Chemistry**

**Biology**

**Social Science**

**شكل1**

**با رده‌بندي منابع ديجيتالي، مي‌توان منابع كتابخانه ها را توسعه داد. وقتي كتابخانه ها اين نوع مواد را در فهرستشان مي‌آورند، مجموعة آنها بدون پرداخت هزينه اضافي گسترش مي‌يابد. براي استفاده‌كنندگان دسترسي از راه دور به آن منابع، از دسترسي به مواد كتابخانه‌اي سنتي آسان‌تر خواهد بود. هنگامي كه جستجوگر، فايلي الكترونيكي را در فهرست مي يابد، ‌تنها با يك كليك ساده بر روي آدرس وب[6]، به اطلاعات مورد نظر خود مي‌رسد. شماره هاي بازيابي به عنوان ابزاري براي بازيابي اطلاعات به كار مي روند. با استفاده از شماره هاي بازيابي، مي‌توان در بيشتر «فهرستهاي عمومي پيوسته»[7] جستجو كرد. وقتي فايلهاي اينترنتي در اين فهرستها قرار مي‌گيرند، استفاده‌كنندگان مي‌توانند منابع را همراه با مواد فيزيكي بر روي قفسه هاي مجازي جستجو كنند. رده‌بندي باعث ايجاد امكان تورّق مي‌شود. تورّق از طريق يك سيستم منطقي سلسله مراتبي، احتمال بيشتري وجود دارد كه سريع‌تر به مدارك مورد نظر رسيد.**

**رده‌بندي خودكار، چاره عصر جديد اطلاعات**

**با افزايش حجم روز افزون اطلاعات و رشد سريع وب، نياز به سازماندهي منابع اينترنتي و بالا بردن قابليت جستجو و دستيابي به منابع تحت وب، روز به روز بيشتر نمايان مي‌شود. بنابراين مسئله سازماندهي منابع اينترنتي براي بهينه‌سازي بازيابي اطلاعات، امروزه فكر اكثر دانشمندان و متخصصان را به خود معطوف نموده است. متخصصان علوم اطلاع‌رساني، علوم رايانه و فناوريهاي اطلاعاتي به فكر طرح سيستمي مناسب براي سازماندهي اين منابع افتاده‌اند. همين متخصصان با تلاش بر روي انواع برنامه‌ها مانند برنامه‌هاي رايانه‌اي مديريت اسناد يا برنامه هاي آرشيوي، به اين نتيجه رسيده‌اند كه علم كتابداري، بيشتر مي‌تواند در اين زمينه (رويارويي با وضعيت آشفتة دنياي جديد اطلاعات) آنها را ياري دهد. بنابراين، به فكر استفاده از رده‌بندي‌هاي كتابخانه‌اي مانند رده‌بندي دهدهي ديويي[8]، رده‌بندي دهدهي جهاني[9]، رده‌بندي كتابخانه كنگره[10] براي سازماندهي منابع اينترنتي افتادند (Jenkins et.al.). «كيت شافر»[11] معتقد است از لحاظ تاريخي، اين كتابداران بوده اند كه دنياي اطلاعات را سازماندهي كرده اند. قرنهاست آنها با موفقيت از عهده سازماندهي و مديريت اطلاعات به هر شكل آن، برآمده اند. آنان با توليد فهرست براي كتابخانه‌ها، مديريت خود را توسعه داده اند. حتي امروزه با توليد ابرداده براي اشياي ديجيتالي، سازماندهي و بازيابي آنها كارآمدتر شده است. «شافر» مي‌گويد كه دنياي كاملاً آشفته و سازماندهي نشدة وب و دنياي سازمان يافتة كتابخانه‌ها امروزه مي‌توانند تأثيرهاي مثبتي بر روي يكديگر بگذارند. از طرفي، مسئوليتهاي كتابداران، روز به روز در حال افزايش است، به طوري كه آنها به سختي از پس وظايف سنتي محوله خود برمي‌آيند و مسلماً فهرست‌نويسي و سازماندهي حجم وسيع اطلاعات الكترونيكي، از توان كاري آنها خارج است (Shafer, 1996). از آنجا كه رده‌بندي، يك فعاليت بسيار وقت‌گير و پرهزينه است و همة منابع الكترونيكي هم هرگز نمي‌توانند توسط انسان فهرست‌نويسي شوند تحقيقات زيادي براي انجام رده‌بندي به صورت خودكار صورت گرفته است.**

**چالشهاي رده‌بندي خودكار**

**رده‌بندي، فعاليتي است كه به مهارت تفكر انتزاعي نياز دارد و اين مهارت را نمي‌توان به سادگي با هوش مصنوعي يا سيستمهاي خبره جايگزين كرد. (Review of attempts to apply Classification, 1997) هنوز نرم‌افزارهاي رايانه‌اي نتوانسته اند هيچ يك از رده‌بنديهاي سنتي كتابخانه‌اي را به طور كامل به صورت خودكار درآورند. تلاش براي به وجود آوردن چنين سامانه‌هاي هوشمندي رو به افزايش است (Toth, 2002).**

**اكثر موتورهاي جستجوي خودكار از راهبردهاي نمايه سازي و مكانيزمهاي ‌بازيابي اطلاعات استفاده مي كنند، اما تعداد بسيار اندكي از آنها از سامانة رده‌بندي خودكار استفاده كرده اند. در رده‌بندي منابع اينترنتي بايد حجم زياد مدارك و رشد صعودي آنها را نيز در نظر گرفت.**

**امروزه وب، در حال تغيير به «وب معنايي»[12] است. وب معنايي در صورتي موفق است كه با مجموعه اي از اطلاعات ساختار يافته روبرو باشد؛ مثل صفحات وبي كه داراي كد موضوعي باشند.**

**از آنجا كه تخصص شبكه‌هاي عصبي مصنوعي در رده‌بندي است، كاربرد شبكه‌هاي عصبي مصنوعي مانند [13]SOM با ضعف روشهاي آماري و زبانشناسي رايانه‌اي در عملي كردن فعاليت رده‌بندي خودكار، رو به افزايش است.**

**تلاش براي انجام رده‌بندي به طور خودكار**

**پروژه‌هاي مختلفي در مورد اينكه چطور اصطلاحات موضوعي بازيابي شده از يك سايت را مي‌توان به كدهاي رده‌بندي تبديل كرد. از شاخص‌ترين پروژه‌هايي كه در اين زمينه صورت گرفته، مي‌توان اين موارد را ذكر كرد:**

**«Noradic WAIS/World Wide Web»: اولين فعاليت انجام شده در اين زمينه محسوب مي‌شود. اين پروژه را در سال 1993 كتابخانه دانشگاه «لاند»[14] سوئد با همكاري «كتابخانه ملي تكنولوژي دانمارك» و با استفاده از رده‌بندي دهدهي جهاني انجام داد. اين پروژه كه يك سال طول كشيد، سه مرحله داشت: بازيابي خودكار مدارك جديد، رده‌بندي خودكار بر اساس رده‌بندي دهدهي جهاني و ساخت يك ساختار درختي موضوعي بر اساس موضوعات اصلي رده‌بندي دهدهي جهاني. از سال 1996 اين برنامه روزآمد نشد.**

**پروژة [15]«GERHARD» را كتابخانه دانشگاه «اُلدنبرگ»[16] آلمان در سال 1996 با استفاده از رده‌بندي دهدهي جهاني انجام داد.**

**پروژه [17]«DESIRE» را كتابخانه دانشگاه لاند سوئد در دو مرحله در سالهاي 1998-1996 و 2000-1998 انجام داد. تمركز اين پروژه بر علوم مهندسي بود و با استفاده از اصطلاحنامه مهندسي [18]«EI» كه خود، شماره رده‌بندي دهدهي جهاني را به همراه اصطلاحات به همراه داشت، نرم‌افزار موفقي ارائه داد.**

**كتابخانه‌هاي بسياري نيز در سراسر جهان برنامه هايي براي رده‌بندي خودكار منابع ديجيتالي فراهم كرده اند. مثلاً گروهي از اساتيد بخش رايانه و فناوري اطلاعات دانشگاه «والور هامپتون»[19] در انگلستان براي برنامه كتابخانه‌اي اين دانشگاه به نام [20]«WWLib» يك «رده بند خودكار»[21] طراحي كرده اند. مهمترين پروژه در زمينه رده‌بندي خودكار پروژه «Scorpion» است كه او.سي.ال.سي[22] از سال 1998 كار آن را آغاز كرد. هدف اين پروژه، ايجاد ابزار لازم براي نمايه سازي، رده‌بندي و فهرست‌نويسي منابع الكترونيكي به صورت خودكار با تركيب علم كتابداري و تكنيكهاي بازيابي اطلاعات است. برنامه «Scorpion» از نظر روشهاي آماري و زبانشناسي تقريباً مشابه «GERHARD» عمل مي‌نمايد، با اين تفاوت كه از نسخه الكترونيكي ديويي[23] استفاده مي‌كند. رده‌بندي ديويي به علت كامل بودن رده ها و مشخص و آشكار بودن مفاهيم و تعاريف آن، مورد استفاده قرار گرفت. در سايت «او.سي.ال.سي» قسمتي به معرفي پروژه‌ها و فعاليتهايي كه در زمينه خودكار انجام شده‌اند، اختصاص دارد.[24] در تحقيقي كه «شافر»، مشاور تحقيقاتي «او.سي.ال.سي» در ارزيابي اين پروژه انجام داد، نتايج رده‌بندي خودكار با برنامه Scorpion با شمارة رده‌بنديهايي كه فهرستنويسان به مداركي خاص داده بودند، مقايسه شد. ميزان مطابقت كار اين برنامه 95% نشان داده شد، اما اين نتيجه جاي بحث نيز دارد. اين روش ارزيابي، تنها ميزان مطابقت كدهاي موضوعي داده شدة اين برنامه را با كدهاي موضوعي ارائه شدة انسان مقايسه مي‌كند و اينكه آيا اين برنامه (Scorpion) كد موضوعي درستي را به مدرك مي دهد يا نه، مشخص نمي‌كند. بنابراين براي ارزيابي بهتر آن بايد روشهاي ديگري را سنجيد (Shafer, 1997).**

**استفاده از رده‌بندي كتابخانة كنگره به عنوان سيستم پايگاه دانش براي انجام رده‌بندي خودكار پروژه جديد «او.سي.ال.سي» است.**

**روند رده‌بندي خودكار**

**روند دادن كد رده‌بندي در اكثر پروژه‌ها شباهتهايي با هم دارند. وقتي مدركي وارد برنامة رده‌بندي خودكار مي‌شود، مراحلي را به ترتيب پشت سر مي گذارد. روشهاي نمايه سازي خودكار براي تجزيه و تحليل متن مورد استفاده قرار مي گيرد. اين روشها كه بيشتر موتورهاي جستجو نيز از آن استفاده مي‌كنند، دو نوع هستند:**

**· «فايلهاي امضايي»:[25] يك نوع كد دهي به مدارك هستند كه الگوريتمهاي خاص خود را دارند و اين كار را به وسيله كوتاه‌سازي لغات با استفاده از فهرست كلمات غيرمجاز[26] و ريشه يابي[27] انجام مي دهند.**

**· «نمايه معكوس»:[28] در يك پايگاه اطلاعاتي، يك «فايل اصلي»[29] وجود دارد كه در آن، همه ركوردها با كد خود ذخيره شده‌اند. جستجو در اين فايل اصلي، بخصوص وقتي تعداد ركورد هاي آن زياد مي‌شود، بسيار وقت‌گير است؛ بنابراين برنامه نويسان، فايلي را تدارك ديده اند كه به صورت معكوس، فيلدها را با شمارة ركورد هاي آنها در بر دارد. مثلاً در برابركلمه Lancaster شماره ركورد هايي كه اين كلمه را دارند، قرار مي گيرد و وقتي كلمه Lancaster جستجو مي‌شود، برنامه به جاي چك كردن تك تك فايلها، ركوردهايي را كه شماره آنها در برابر اين كلمه در فايل معكوس قرار گرفته اند، بازيابي مي‌كند. فايلهاي معكوس نسبت به فايلهاي امضايي براي ذخيره‌سازي به حافظة بالاتري نياز دارند.**

**شايان ذكر است، در زمان تجزيه و تحليل منابع اينترنتي برچسب‌هاي[30] مربوط به ابرداده، از متن اصلي جدا شده و براي تهيه يك نمايه استفاده مي شوند.(Gietz, 2001)**

**مدرك بعد از ورود به سيستم، تجزيه و تحليل مي‌شود، بدين صورت كه براي مشخص شدن موضوع مدرك، ابتدا كلمات غير مجاز متن (مثل حروف اضافه يا ربط) از آن جدا مي‌شود. كلماتي كه داراي بسامد[31] بالا و پايين باشند نيز از متن جدا مي شوند. كلماتي كه داراي بسامد بالا هستند، معمولاً كلمات غير مجازند و كلماتي كه داراي بسامد پايين هستند، درجه اهميت پاييني دارند. بنابراين، كلماتي كه داراي بسامد متوسط هستند، اهميت بالايي دارند و به عنوان كليدواژه‌ها انتخاب مي شوند. در اين مرحله، از روشهاي آماري استفاده مي‌شود (Gietz, 2001).**

**اين مورد، به قانون «زيفت»[32] نيز اشاره دارد كه در تحليل استنادي از آن استفاده مي‌شود. در بعضي از روشها ريشه‌يابي هم صورت مي‌گيرد؛ بدين معنا كه ريشه كلمات جدا مي‌شوند، مثلاً:**

**در بعضي از روشها ريشه يابي[33] هم صورت مي گيرد بدين معني كه ريشه كلمات جدا مي شوند مثلاً:**

**Measurement measure**

**Interaction Interact**

**Experimental experiment**

**بعد از اين مرحله نوبت مي‌رسد به ابن‌كه دريابيم كلمات انتخاب شده از كدام قسمت متن يافت شده اند. كلماتي كه در عنوان، مقدمه، چكيده، عناوين فصول و بحث و نتيجه گيري يافت شوند، نسبت به كلمات يافت شده از ديگر قسمتهاي متن درجه اعتبار بالايي دارند.**

**پس از مشخص شدن موضوع و كليدواژه‌هاي متن، با استفاده از روشهاي آماري و بر حسب اولويت، بايد به مدرك، كد رده‌بندي[34] تعلق بگيرد. نرم‌افزار «Scorpion» بر اساس يك پايگاه اطلاعاتي قابل جستجو در فايل رده‌بندي ديويي ساخته شده است. بنابراين، مدركي كه مي‌خواهد كد رده‌بندي بگيرد، مثل يك سؤال در برابر سيستم رده‌بندي ديويي قرار مي گيرد و كد رده‌بندي به عنوان نتيجه جستجو به مدرك تعلق مي‌گيرد. نرم‌افزار بازيابي آن «SMART» نام دارد (Marcella & Maltby, 2000).**

**بر اساس طرح رده‌بنديي كه اتخاذ مي‌شود، يك پايگاه اطلاعاتي از سيستم رده‌بندي تهيه شده و به مدركي كه با كليدواژه‌هاي مشخص خود وارد اين پايگاه مي‌شود، يك كد رده‌بندي اختصاص مي يابد. پروژة «DESIRE II» كه فقط در زمينه علوم مهندسي برنامه خود را اجرا كرد، در برنامه رده‌بندي خودكار خود از اصطلاحنامة «EI» كه مختص علوم مهندسي است، استفاده كرده است. بعد از ساختن يك نمايه از كلمات بيرون آمده از متن، آن را با واژه‌هاي يك اصطلاحنامه (اصطلاحنامه اي كه در كنار واژه‌هاي آن، كد هاي رده‌بندي نيز وجود دارند) مطابقت مي دهد و پس از جدا كردن واژه‌هاي غير مجاز، هر جا كه مطابقتي بين اصطلاح موجود در نمايه و اصطلاح موجود در اصطلاحنامه يافت، شماره مربوط را به آن مدرك اختصاص مي دهد. دقت اين برنامه رده‌بندي بين 57 تا 66% نشان داده شده است (Toth, 2002).**

**اصطلاحنامه[35]**

**به كار گيري اصطلاحنامه در رده‌بندي خودكار منابع اطلاعاتي بسيار مفيد است. اصطلاحنامه‌ها[36] مجموعه اي از اصطلاحات مرتبط هستند كه با يك نظم سلسله مراتبي، مفاهيم اعم و اخص و كلماتي را كه داراي شكل يكسان، ولي معاني گوناگون هستند،[37] آورده‌اند. پشتيباني اصطلاحنامه‌ها بسيار گران است و به دانش خاص نياز دارد. (Gietz, 2001) اصطلاحنامه‌ها نقش مهمي در نظامهاي ذخيره و بازيابي اطلاعات دارند. ظهور وب، همراه با توسعه و پيشرفتهاي اخير در كاربرد اصطلاحنامه‌ها به عنوان ابزارهاي بازيابي اطلاعات، باعث تولد نسل جديدي از اصطلاحنامه‌ها شده است. اصطلاحنامه‌هاي وب محور، راه خود را به محيطهاي بازيابي و سازماندهي اطلاعات وب محور باز نموده و در تهيه ابر داده ها، نمايه سازي صفحات، سايتهاي وب، پايگاههاي داده و موتورهاي جستجو استفاده مي‌شوند. ساختار هاي معنايي موجود در اصطلاحنامه‌ها مي‌توانند هم در سازماندهي و هم در بازيابي اطلاعات وب و منابع دانش نقش داشته باشند (رنجبر، 1385).**

**مشكلات روشهاي آماري و زبانشناسي رايانه‌اي باعث شده است كه اين پروژه‌ها فكر استفاده از ابزارهاي ديگري را در سر بپرورانند. امروزه توجه رده‌بندي خودكار از منابع متني به منابع چند رسانه اي نيز جلب شده است.**

**شبكة عصبي مصنوعي[38]**

**در پروژه‌هاي ياد شده، از روشهاي اكتشافي، وزن‌دهي اصطلاح و روشهاي زبانشناسي رايانه‌اي استفاده شده بود. اگرچه ساختار كلي مراحل رده‌بندي خودكار در پروژه‌ها مشابه هم هستند، اما از فناوريهاي متفاوتي استفاده كرده اند. به نظر مي‌رسد بعضي از فناوريها با بعضي از موضوعات سازگاري بيشتري دارند و سيستمهايي كه تنها بر روي يك زمينة موضوعي كار كرده بودند، موفق تر بوده اند. همكاريهاي بين محققان اين طرحها باعث گرديده روشهاي پژوهش آنها تصحيح و بهبود يابد (Gietz, 2001).**

**نياز روز افزوني به تغيير و سازگاري سيستمهاي رده‌بندي كتابخانه‌اي براي تبديل شدن به يك ابزار تورّق براي منابع اينترنتي به وجود آمده است. براي به وجود آوردن اين امكان، به روشهاي «خوشه‌بندي»[39] و تكنيكهاي مبتني بر محتوا يا مبتني بر استناد نياز است. اين ويژگيها را در پتانسيل شبكه‌هاي عصبي مصنوعي مي‌توان يافت.**

**بر اساس تجزيه و تحليلهاي به عمل آمده، مشخص شده است كه شبكه‌هاي عصبي مصنوعي به خوبي اصطلاحنامه‌هاي سلسله مراتبي مي‌توانند در اين زمينه موفق باشند.**

**با افزايش حجم اطلاعات و پيشرفت فناوريهاي اطلاعاتي، استفاده از الگوريتمهاي سنتي براي بازيابي سريع داده ها كافي نيست. استفاده از شبكه‌هاي عصبي مصنوعي به عنوان يكي از تكنيكهاي هوش مصنوعي، راهكار مناسبي براي افزايش سرعت بازيابي اطلاعات در حجم انبوه است. شبكه‌هاي عصبي مصنوعي بازنمون مناسبي از دانش براي كاربردهاي بازيابي اطلاعات ارائه مي دهند. گره هاي شبكه عصبي نمايانگر عناصر بازيابي اطلاعات از قبيل كليد واژه، نويسنده و ... هستند و پيوندهاي موجود در شبكه براي انتقال ورودي از لايه به لاية ديگر استفاده شده و در نتيجه به دستيابي به خروجي شبكه، كه همان بازيابي مدرك است، مي انجامد. شبكه‌هاي عصبي چند لايه، از تعدادي گره و پيوند تشكيل شده است . اطلاعات از طريق گره هاي ورودي به شبكه عصبي وارد شده، سپس با استفاده از پيوندها به لايه‌هاي بعدي (پنهان) منتقل گرديده و در نهايت، خروجي شبكه از گره هاي لايه خروجي به دست مي‌آيد (شكل2).**

**با در نظر گرفتن خصوصيات شبكه‌هاي عصبي، به نظر مي‌رسد اين تكنيكِ هوش مصنوعي براي بازيابي اطلاعات، مؤثر باشد . در حال حاضر، در مدلهاي كاربردي شبكه‌هاي عصبي در بازيابي اطلاعات، تحقيقاتي صورت گرفته است و در آينده با پيشرفت بيشتر سخت افزار و نرم‌افزار، به نظر مي‌رسد به سرعت بتوان از شبكه‌هاي عصبي استفاده هاي مؤثرتري نمود. با ارزانتر شدن سخت افزار مورد نياز مدلهاي شبكه عصبي، امكان استفاده از آنها، با استفادة موازي براي طبقه بندي مدارك، امكان‌پذير شده و بازيابي اطلاعات با استفاده از شبكه‌هاي عصبي مصنوعي، بسيار سريعتر خواهد شد (Mehrad & Koleini, 2007). همچنين، با پيشرفتهاي نرم‌افزاري در ايجاد روشهاي جديد كنترل توابع شبكه، به منظور پياده‌سازي مدلهاي شبكه عصبي، تحولي مهم در سرعت بازيابي اطلاعات فراهم مي‌شود.**

**گره‌هاي خروجي گره‌هاي پنهان گره‌هاي ورودي**

**روند كار**

**شكل2. نمونه يك شبكه عصبي**

**نتيجه‌گيري**

**دنياي جديد اطلاعات با حجم روزافزون منابع اينترنتي دچار مسئله‌اي شده است كه جز با كشف راهي مناسب براي سازماندهي اين اطلاعات قابل حل نيست. رده‌بندي يكي از شيوه‌هاي سازماندهي اطلاعات براي بازيابي سريع آنهاست كه با استفاده از طرحهاي رده‌بندي مثل رده‌بندي ديويي، رده‌بندي دهدهي جهاني و رده‌بندي كنگره صورت مي‌گيرد. از آنجا كه رده‌بندي، فعاليتي وقت‌گير و پرهزينه بوده و فهرست‌نويسي منابع اينترنتي، خارج از توان فهرستنويسان است، پروژه‌هاي زيادي براي انجام رده‌بندي به صورت خودكار و با ماشين، اجرا شده‌اند. استفاده از تكنيكهاي هوش مصنوعي چون شبكه‌هاي عصبي مصنوعي، در تحقق اين هدف سؤالهايي مؤثر است. اما در كاربرد طرح مناسبي براي انجام رده‌بندي به صورت خودكار، مطرح است كه مؤسساتي چون «او.سي.ال.سي» تحقيقات زيادي براي پاسخگويي به آنها آغاز كرده‌اند.**

**§ آيا طرحهاي رده‌بندي كتابخانه‌اي مانند رده‌بندي دهدهي ديويي و رده‌بندي كتابخانة كنگره مي‌توانند براي رده‌بندي منابع ديجيتالي به كار بروند؟ و آيا نيازي به توليد سيستم رده‌بندي مخصوصي براي رده‌بندي منابع ديجيتالي متناسب با ويژگيهاي خاص آنها ضرورت دارد؟**

**§ آيا مي‌توان از اين رده‌بنديها براي رده‌بندي خودكار منابع اينترنتي و ديگر منابع ديجيتالي بهره برد؟**

**§ چه پيشرفتهايي لازم است تا سيستمهاي رده‌بندي خودكار، اين كار را به دقت يك متخصص رده‌بندي انجام دهند؟**

**§ آيا ايجاد مركزي واحد براي دادن كد رده‌بندي به صفحات وب به صورت اجباري ضرورت دارد؟**

**منابع**

**ـ رنجبر، جهانگير (1385). «آيا در محيط اطلاعاتي جديد جايي براي اصطلاحنامه‌ها باقي مانده است؟»،كتابداري و اطلاع‌رساني، جلد9، شماره اول، صص114-115.**

**ـ مقصودي، رؤيا (1378). «بررسي استنادي و تطبيقي پايان‌نامه‌هاي كارشناسي ارشد روانشناسي تربيتي دانشگاههاي شهيد چمران اهواز و شيراز در سالهاي 1372 الي 1376». پايان‌نامه كارشناسي ارشد علوم كتابداري و اطلاع‌رساني، دانشگاه شهيد چمران اهواز، ص36.**

**-Automatic Classification Research at OCLC. Available at: http://www.oclc.org/research/projects/auto\_class/default.htm**

**-DESIRE: Project Deliverable. (1997), Available at: http://www.ub.lu.se/desire/radar/reports/D3.2.3/class\_v10.html**

**-Jenkins, Charlotte, et al. Automatic Classification of Web Resources using Java and Dewey Decimal classification, Available at: http://www.scit.wlv.ac.uk/seed/www7paper/**

**-Jansen, Bernard J., Spink, Amanda. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. Information Processing and Management. Vol.42, No.1. p. 248.**

**-Toth, Erzsebet. (2002) Innovative Solutions in Automatic Classification: A Brief Summary, Libri,. page 48-53.**

**-Marcella, Rita & Maltby, Arthur. (2000). The Future of Classification. Hampshire: Gower Publishing Company.**

**-Shafer, Keith. (1996) Scorpion Project explores using Dewey to organize the Web. OCLC Newsletter. No. 222. Available at: http://digitalarchive.oclc.org/da/ViewObjectMain.jsp?fileid=0000001701:000000042840&reqid=85#scor**

**- Gietz, Peter. Report on automatic classification systems. (2001). Available at: http://www.daasi.de/reports/Report-automatic-classification.html**

**- Shafer, Keith. (1997) Evaluating Scorpion Results. Available at: http://digitalarchive.oclc.org/da/ViewObjectMain.jsp?objid= 0000003411&frame=true**

**-Review of attempts to apply classification in automated services. Available at: http://www.ukoln.ac.uk/metadata/desire/ classification/ class\_7.htm**

**- Mehrad, J.Koleini, S. (2007) USING SO NEURAL NETWORK IN TEXT INFORMATION RETRIEVAL. Iranian Journal of Information Science and Technology. Vol 5, No.1. P.53-64.**

**- http://en.wikipedia.org/wiki/Semantic-web**

**- http://en.wikipedia.org/wiki/Artificial-neural-network**

**--------------------------------------------------------------------------------**

**1. دانشجوي كارشناسي ارشد دانشگاه شيراز: negaresma@yahoo.com**

**1. Too Much Information.**

**1. Directory.**

**2. Open Directory Project.**

**3. Human Editors.**

**1. URL (Uniform Resource Locator).**

**1. OPAC (Online Public Access Catalogues).**

**2. DDC (Dewey Decimal Classification).**

**3. UDC (Universal Decimal Classification).**

**4. LCC (Library of Congress Classification).**

**5. Keith Shafer.**

**1. Semantic Web يا وب معنايي نسبت به وب، يك انقلاب محسوب مي‌شود كه در آن اطلاعات، قابل خواندن و تجزيه و تحليل توسط ماشين است. در حالي‌كه صفحات وب كنوني را فقط انسان مي‌تواند بخواند. وب معنايي اين اجازه را به مرورگرها و ديگر نرم‌افزارها مي‌دهد تا اطلاعات را خوانده، به راحتي تجزيه و تحليل كنند (Wikipedia).**

**1. Self Organizing Map.**

**2. Lund.**

**3. German Harvest Automated Retrieval and Directory.**

**4. Oldenburg.**

**5. Development of European Service for Information on research and Education.**

**6. Engineering Index.**

**1. Wolverhampton.**

**2. Wolverhampton Web Library.**

**3. Automatic Classifier.**

**4. OCLC (Online Computer Library Center).**

**5. Electronic Dewey for Windows.**

**6. Automatic Classification Research.**

**1. Signature Files.**

**2. Stop Word List.**

**3. Stemming.**

**4. Inverted File.**

**5. Master File.**

**6. Tag.**

**1. Frequency.**

**2. قانون زيف بر بسامد واژه‌ها و مفاهيم موجود در متون تكيه دارد. به موجب اين قانون، اگر كلمات موجود در يك مقاله را به ترتيب نزولي بسامد آنها مرتب كنيم، هر يك از كلمات در متن، نسبت معكوس با بسامد رخداد كلمه خواهد داشت (مقصودي، 1378).**

**3. Stemming.**

**4. Notation.**

**1. Thesaurus.**

**2. Thesauri.**

**3. Homonyms.**

**1. Artificial Neural Networkشبكه عصبيِ مصنوعيِ نمونه شبيه سازي شده از شبكه عصبي انسان است كه از مدلهاي آن در پردازش داده‌ها، رده‌بندي و پيش‌بيني نتايج كار استفاده مي شود (wikipedia) .**

**2. Clustering.**