

## رویکردها و چالشهای رده‌بندی خودکار منابع اطلاعاتی در محیط جدید

رضیه اسماعیل‌پور<sup>۱</sup>

### چکیده

سازماندهی و مدیریت منابع اینترنتی، همواره یکی از دغدغه‌های دانشمندان علوم رایانه و اطلاع‌رسانی است. امروزه همین متخصصان به این نتیجه رسیده‌اند که علم کتابداری در زمینه رویارویی با وضعیت آشفته اطلاعات می‌تواند نقش برجسته‌ای ایفا کند. از آنجا که رده‌بندی، فعالیتی وقت‌گیر و پرهزینه بوده و رده‌بندی منابع اینترنتی برای فهرست‌نویسان دشوار می‌باشد، تحقیقات زیادی برای انجام رده‌بندی به صورت خودکار صورت گرفته است. کاربرد شبکه‌های عصبی مصنوعی با توجه به ضعف روشهای آماری و زبانشناسی رایانه‌ای در عملی کردن فعالیت رده‌بندی خودکار، رو به افزایش است.

کلیدواژه‌ها: رده‌بندی خودکار، شبکه عصبی مصنوعی، اصطلاحنامه، سازماندهی منابع اینترنتی.

### مسائل و مشکلات بازیابی اطلاعات در عصر جدید

عدم کنترل ورود اطلاعات به محیط وب و سرعت افزایش منابع اطلاعاتی از یک سو، و نبود راهکارهای مناسب برای سازماندهی آنها، سبب شده بازیابی این منابع با مشکلاتی از جمله ربط ضعیف، ریزش کاذب فراوان و صرف وقت زیاد از کاربران، همراه باشد. فراوانی تولید اطلاعات، آنقدر تأثیرگذار بوده که باعث ورود اصطلاحات و لغات

---

۱. دانشجوی کارشناسی ارشد دانشگاه شیراز: [negaresma@yahoo.com](mailto:negaresma@yahoo.com)

جدیدی به زبان شده است: سیل اطلاعات، طوفان اطلاعات و اخیراً «Infoglut»<sup>۱</sup> یک بیماری شناخته شده است که جهانیان از آن رنج می‌برند.

وقتی با انبوهی از کتابها روبرو بودیم، برای سازماندهی آنها به یک سیستم رده‌بندی نیاز داشتیم تا بتواند با منظم کردن کتابها در هر رده موضوعی، به بازیابی آنها سرعت بدهد. اکنون نیز با شکل‌های دیگر کتاب روبرو هستیم که همان «فرمت الکترونیکی» است. این گونه اطلاعات به علت قابلیت‌های خاص خود سرعت نشر بیشتری نسبت به مواد چاپی دارند. همچنین، اکثر مواد چاپی، قبل از انتشار، به شکل الکترونیکی در دسترس هستند. زمانی که برای سازماندهی انبوهی از کتابها و بازیابی سریع آنها به یک سیستم رده‌بندی نیاز داریم، ضرورت یک سیستم رده‌بندی مناسب برای اطلاعات روبه رشد دیجیتالی، بیشتر مشخص می‌شود.

با انتشار رو به رشد انواع منابع اطلاعاتی به شکل الکترونیکی (مجله‌ها و روزنامه‌های الکترونیکی و پایگاه‌های اطلاعاتی) و همچنین گرایش رو به رشد میان رشته‌ای شدن علوم، همواره جستجوگران اطلاعات، با انبوه وسیعی از اطلاعات روبرو می‌شوند. برای مثال وقتی کاربر، کلیدواژه‌ای را در موتور جستجوی گوگل جستجو می‌کند، حداقل بین ۳۰۰ تا ۴۰۰ هزار مدرک بازیابی می‌شود که جستجوی وی با خاص کردن آن کلیدواژه، به یافتن ۳۰۰ تا ۴۰۰ مدرک می‌انجامد. کاربر با وقت اندک خود، محال است که بیشتر از ۳۰ تا ۴۰ مورد آن را ببیند؛ چون از نظر زمانی مقرون به صرفه نیست که وقت خود را صرف دیدن اطلاعات بسیار زیادی کند که ممکن است به کار وی نیاید. در این زمان، کاربر مطمئن هم نیست که جستجوی وی کامل بوده یا نه، زیرا موفق به دیدن همه اطلاعات بازیابی شده کلیدواژه خود نبوده است. آمارها نشان می‌دهند کاربران از میان صفحات بازیابی شده در موتورهای جستجو همیشه صفحات کمتری را می‌بینند (Jansen & Spink, 2006). همچنین در یک مجموعه از صفحات بازیابی شده، مدخل‌های تکراری فراوانی دیده می‌شوند.

---

1. Too Much Information.

### اهمیت و ضرورت رده‌بندی اطلاعات در عصر جدید

با تولید روزافزون اطلاعات نیاز مبرم به رده‌بندی کردن آنها برای بهینه‌سازی بازیابی اطلاعات، پررنگ‌تر می‌شود. یافتن اطلاعات مورد نیاز، تنها با جستجوی کلید واژه‌ای از طریق موتورهای کاوش امکان‌پذیر نیست. دانشمندان معمولاً با تورق نشریات معتبر در زمینه علمی خود، بیشتر و راحت‌تر به اطلاعات مورد نیاز خود دست می‌یابند. زیرا اکثر مواقع، فردی که به دنبال اطلاعات است، تعریف مشخصی از آنچه می‌خواهد، ندارد و نمی‌تواند کلید واژه‌ای معین و مشخص را انتخاب و سپس بر اساس آن جستجو کند. بنابراین، افراد با تورق مطالب بیشتر، امکان یافتن اطلاعات مورد نیاز خود را دارند. امروزه فهرستهای راهنما<sup>۱</sup> به خوبی امکان تورق را فراهم می‌آورند؛ مثلاً فهرست راهنمای ODP<sup>۲</sup> که در نوع خود، بزرگترین فهرست جهان به شمار می‌رود، توانسته با یک رده‌بندی موضوعی، وب‌سایتهای مناسبی را به شکل سلسله مراتبی نظم ببخشد. این فهرست راهنما از ویراستاران<sup>۳</sup> برای رده‌بندی صفحات وب استفاده می‌کند.

وقتی اطلاعات به طور موضوعی رده‌بندی شوند، هر متخصص با روی آوردن به مطالب مربوط به تخصص خود، راحت‌تر به مطلب مورد نیاز خود می‌رسد و دیگر با مطالب نامربوط و انبوه زیادی از رکوردهای بازیابی شده، روبرو نمی‌شود؛ گویی مشغول تورق در مجله مورد نظر خود است.

در اینجا، مسئله رده‌بندی برای منابع دیجیتالی، حیاتی به نظر می‌رسد؛ به طوری که اگر این امر صورت نپذیرد یعنی منابع دیجیتالی رده‌بندی نشوند، به علت عدم حضور فیزیکی مثل این است که گم شده باشند. (Review o attempts to apply Classification).

پایگاه اطلاعاتی نیز مثل یک کتابخانه است که انبوهی از مدارک دارد. زمانی که کتابها رده‌بندی می‌شوند، منابع مربوط کنار هم قرار می‌گیرند، به عنوان مثال کتابهای

- 
1. Directory.
  2. Open Directory Project.
  3. Human Editors.

مربوط به تاریخ ایران در دوره قاجار، یک جا در یک قفسه کنار هم قرار می‌گیرند و می‌توان از میان آنها کتاب مورد نظر را پیدا کرد. برای یافتن مطلب مورد نظر در یک پایگاه اطلاعاتی، یا می‌توان با کلیدواژه به جستجو پرداخت و یا زمانی که مدارک در این پایگاه رده‌بندی موضوعی شده باشند، می‌توان با دیدن موضوعات کلی چون کشاورزی، شیمی، زیست‌شناسی و علوم اجتماعی، یکی را انتخاب کرد و سپس با کلیک کردن بر روی آن، موضوعات خاص‌تر آن را دید. برای مثال، با کلیک کردن بر روی «شیمی» ساختار دیگری برای شما نمایش داده می‌شود: شیمی آلی، شیمی معدنی، زیست‌شیمی و ... این موضوع، تا جای ممکن خاص می‌شود، سپس با کلیک بر روی موضوع خاص، می‌توان مدارک مورد نظر خود دید. (شکل ۱).

Agriculture

Chemistry

Analytical Chemistry  
Biological Chemistry  
Geophysical Chemistry

Biology

Social Science

شکل ۱

با رده‌بندی منابع دیجیتالی، می‌توان منابع کتابخانه‌ها را توسعه داد. وقتی کتابخانه‌ها این نوع مواد را در فهرستشان می‌آورند، مجموعه آنها بدون پرداخت هزینه اضافی گسترش می‌یابد. برای استفاده کنندگان دسترسی از راه دور به آن منابع، از دسترسی به مواد کتابخانه‌ای سنتی آسان‌تر خواهد بود. هنگامی که جستجوگر، فایل الکترونیکی را در فهرست می‌یابد، تنها با یک کلیک ساده بر روی آدرس وب<sup>۱</sup>، به اطلاعات مورد نظر خود می‌رسد. شماره‌های بازیابی به عنوان ابزاری برای بازیابی اطلاعات به کار می‌روند. با

---

1. URL (Uniform Resource Locator).

استفاده از شماره‌های بازیابی، می‌توان در بیشتر «فهرستهای عمومی پیوسته»<sup>۱</sup> جستجو کرد. وقتی فایل‌های اینترنتی در این فهرستها قرار می‌گیرند، استفاده‌کنندگان می‌توانند منابع را همراه با مواد فیزیکی بر روی قفسه‌های مجازی جستجو کنند. رده‌بندی باعث ایجاد امکان تورق می‌شود. تورق از طریق یک سیستم منطقی سلسله‌مراتبی، احتمال بیشتری وجود دارد که سریع‌تر به مدارک مورد نظر رسید.

### رده‌بندی خودکار، چاره‌ عصر جدید اطلاعات

با افزایش حجم روز افزون اطلاعات و رشد سریع وب، نیاز به سازماندهی منابع اینترنتی و بالا بردن قابلیت جستجو و دستیابی به منابع تحت وب، روز به روز بیشتر نمایان می‌شود. بنابراین مسئله سازماندهی منابع اینترنتی برای بهینه‌سازی بازیابی اطلاعات، امروزه فکر اکثر دانشمندان و متخصصان را به خود معطوف نموده است. متخصصان علوم اطلاع‌رسانی، علوم رایانه و فناوریهای اطلاعاتی به فکر طرح سیستمی مناسب برای سازماندهی این منابع افتاده‌اند. همین متخصصان با تلاش بر روی انواع برنامه‌ها مانند برنامه‌های رایانه‌ای مدیریت اسناد یا برنامه‌های آرشیوی، به این نتیجه رسیده‌اند که علم کتابداری، بیشتر می‌تواند در این زمینه (رویاریبی با وضعیت آشفته دنیای جدید اطلاعات) آنها را یاری دهد. بنابراین، به فکر استفاده از رده‌بندی‌های کتابخانه‌ای مانند رده‌بندی دهدهی دیویی<sup>۲</sup>، رده‌بندی دهدهی جهانی<sup>۳</sup>، رده‌بندی کتابخانه کنگره<sup>۴</sup> برای سازماندهی منابع اینترنتی افتادند (Jenkins et.al.). «کیت شافر»<sup>۵</sup> معتقد است از لحاظ تاریخی، این کتابداران بوده‌اند که دنیای اطلاعات را سازماندهی کرده‌اند. قرن‌هاست آنها با موفقیت از عهده سازماندهی و مدیریت اطلاعات به هر شکل آن، برآمده‌اند. آنان با تولید فهرست برای کتابخانه‌ها، مدیریت خود را توسعه داده‌اند. حتی امروزه با تولید ابر داده برای اشیای

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.

دیجیتالی، سازماندهی و بازیابی آنها کارآمدتر شده است. «شافر» می‌گوید که دنیای کاملاً آشفته و سازماندهی نشده وب و دنیای سازمان یافته کتابخانه‌ها امروزه می‌توانند تأثیرهای مثبتی بر روی یکدیگر بگذارند. از طرفی، مسئولیتهای کتابداران، روز به روز در حال افزایش است، به طوری که آنها به سختی از پس وظایف سنتی محوله خود برمی‌آیند و مسلماً فهرست‌نویسی و سازماندهی حجم وسیع اطلاعات الکترونیکی، از توان کاری آنها خارج است (Shafer, 1996). از آنجا که رده‌بندی، یک فعالیت بسیار وقت‌گیر و پرهزینه است و همه منابع الکترونیکی هم هرگز نمی‌توانند توسط انسان فهرست‌نویسی شوند تحقیقات زیادی برای انجام رده‌بندی به صورت خودکار صورت گرفته است.

### چالشهای رده‌بندی خودکار

رده‌بندی، فعالیتی است که به مهارت تفکر انتزاعی نیاز دارد و این مهارت را نمی‌توان به سادگی با هوش مصنوعی یا سیستمهای خبره جایگزین کرد. (Review of attempts to apply Classification, 1997) هنوز نرم‌افزارهای رایانه‌ای نتوانسته‌اند هیچ یک از رده‌بندیهای سنتی کتابخانه‌ای را به طور کامل به صورت خودکار درآورند. تلاش برای به وجود آوردن چنین سامانه‌های هوشمندی رو به افزایش است (Toth, 2002).

اکثر موتورهای جستجوی خودکار از راهبردهای نمایه‌سازی و مکانیزمهای بازیابی اطلاعات استفاده می‌کنند، اما تعداد بسیار اندکی از آنها از سامانه رده‌بندی خودکار استفاده کرده‌اند. در رده‌بندی منابع اینترنتی باید حجم زیاد مدارک و رشد صعودی آنها را نیز در نظر گرفت.

امروزه وب، در حال تغییر به «وب معنایی»<sup>۱</sup> است. وب معنایی در صورتی موفق است که با مجموعه‌ای از اطلاعات ساختار یافته روبرو باشد؛ مثل صفحات وبی که دارای کد موضوعی باشند.

۱. Semantic Web یا وب معنایی نسبت به وب، یک انقلاب محسوب می‌شود که در آن اطلاعات، قابل خواندن و تجزیه و تحلیل توسط ماشین است. در حالی که صفحات وب کنونی را فقط انسان می‌تواند بخواند. وب معنایی این اجازه را به مرورگرها و دیگر نرم‌افزارها می‌دهد تا اطلاعات را خوانده، به راحتی تجزیه و تحلیل کنند (Wikipedia).

رویکردها و چالشهای رده‌بندی خودکار ... / ۹۷

از آنجا که تخصص شبکه‌های عصبی مصنوعی در رده‌بندی است، کاربرد شبکه‌های عصبی مصنوعی مانند<sup>۱</sup> SOM با ضعف روشهای آماری و زبانشناسی رایانه‌ای در عملی کردن فعالیت رده‌بندی خودکار، رو به افزایش است.

### تلاش برای انجام رده‌بندی به طور خودکار

پروژه‌های مختلفی در مورد اینکه چطور اصطلاحات موضوعی بازیابی شده از یک سایت را می‌توان به کدهای رده‌بندی تبدیل کرد. از شاخص‌ترین پروژه‌هایی که در این زمینه صورت گرفته، می‌توان این موارد را ذکر کرد:

«Noradic WAIS/World Wide Web»: اولین فعالیت انجام شده در این زمینه محسوب می‌شود. این پروژه را در سال ۱۹۹۳ کتابخانه دانشگاه «لاند»<sup>۲</sup> سوئد با همکاری «کتابخانه ملی تکنولوژی دانمارک» و با استفاده از رده‌بندی دهدهی جهانی انجام داد. این پروژه که یک سال طول کشید، سه مرحله داشت: بازیابی خودکار مدارک جدید، رده‌بندی خودکار بر اساس رده‌بندی دهدهی جهانی و ساخت یک ساختار درختی موضوعی بر اساس موضوعات اصلی رده‌بندی دهدهی جهانی. از سال ۱۹۹۶ این برنامه روزآمد نشد.

پروژه<sup>۳</sup> «GERHARD» را کتابخانه دانشگاه «الْدنبرگ»<sup>۴</sup> آلمان در سال ۱۹۹۶ با استفاده از رده‌بندی دهدهی جهانی انجام داد.

پروژه<sup>۵</sup> «DESIRE» را کتابخانه دانشگاه لاند سوئد در دو مرحله در سالهای ۱۹۹۶-۱۹۹۸ و ۲۰۰۰-۱۹۹۸ انجام داد. تمرکز این پروژه بر علوم مهندسی بود و با استفاده از اصطلاحنامه مهندسی<sup>۱</sup> «EI» که خود، شماره رده‌بندی دهدهی جهانی را به همراه اصطلاحات به همراه داشت، نرم‌افزار موفقی ارائه داد.

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.

کتابخانه‌های بسیاری نیز در سراسر جهان برنامه‌هایی برای رده‌بندی خودکار منابع دیجیتالی فراهم کرده‌اند. مثلاً گروهی از اساتید بخش رایانه و فناوری اطلاعات دانشگاه «والور هامپتون»<sup>۱</sup> در انگلستان برای برنامه کتابخانه‌ای این دانشگاه به نام «WWLib»<sup>۲</sup> یک «رده بند خودکار»<sup>۳</sup> طراحی کرده‌اند. مهمترین پروژه در زمینه رده‌بندی خودکار پروژه «Scorpion» است که او.سی.ال.سی.<sup>۴</sup> از سال ۱۹۹۸ کار آن را آغاز کرد. هدف این پروژه، ایجاد ابزار لازم برای نمایه‌سازی، رده‌بندی و فهرست‌نویسی منابع الکترونیکی به صورت خودکار با ترکیب علم کتابداری و تکنیکهای بازیابی اطلاعات است. برنامه «Scorpion» از نظر روشهای آماری و زبانشناسی تقریباً مشابه «GERHARD» عمل می‌نماید، با این تفاوت که از نسخه الکترونیکی دیویی<sup>۵</sup> استفاده می‌کند. رده‌بندی دیویی به علت کامل بودن رده‌ها و مشخص و آشکار بودن مفاهیم و تعاریف آن، مورد استفاده قرار گرفت. در سایت «او.سی.ال.سی.»<sup>۶</sup> قسمتی به معرفی پروژه‌ها و فعالیت‌هایی که در زمینه خودکار انجام شده‌اند، اختصاص دارد.<sup>۶</sup> در تحقیقی که «شافر»، مشاور تحقیقاتی «او.سی.ال.سی.» در ارزیابی این پروژه انجام داد، نتایج رده‌بندی خودکار با برنامه Scorpion با شماره رده‌بندی‌هایی که فهرست‌نویسان به مدارکی خاص داده بودند، مقایسه شد. میزان مطابقت کار این برنامه ۹۵٪ نشان داده شد، اما این نتیجه جای بحث نیز دارد. این روش ارزیابی، تنها میزان مطابقت کدهای موضوعی داده شده این برنامه را با کدهای موضوعی ارائه شده انسان مقایسه می‌کند و اینکه آیا این برنامه (Scorpion) کد موضوعی درستی را به مدرک می‌دهد یا نه، مشخص نمی‌کند. بنابراین برای ارزیابی بهتر آن باید روشهای دیگری را سنجید (Shafer, 1997).

استفاده از رده‌بندی کتابخانه کنگره به عنوان سیستم پایگاه دانش برای انجام رده‌بندی خودکار پروژه جدید «او.سی.ال.سی.» است.

1. Wolverhampton.
2. Wolverhampton Web Library.
3. Automatic Classifier.
4. OCLC (Online Computer Library Center).
5. Electronic Dewey for Windows.
6. Automatic Classification Research.



## روند رده‌بندی خودکار

روند دادن کد رده‌بندی در اکثر پروژه‌ها شباهتهایی با هم دارند. وقتی مدرکی وارد برنامه رده‌بندی خودکار می‌شود، مراحل را به ترتیب پشت سر می‌گذارد. روشهای نمایه سازی خودکار برای تجزیه و تحلیل متن مورد استفاده قرار می‌گیرد. این روشها که بیشتر موتورهای جستجو نیز از آن استفاده می‌کنند، دو نوع هستند:

- «فایلهای امضایی»: یک نوع کد دهی به مدارک هستند که الگوریتمهای خاص خود را دارند و این کار را به وسیله کوتاه‌سازی لغات با استفاده از فهرست کلمات غیرمجاز<sup>۲</sup> و ریشه‌یابی<sup>۳</sup> انجام می‌دهند.

- «نمایه معکوس»: در یک پایگاه اطلاعاتی، یک «فایل اصلی»<sup>۴</sup> وجود دارد که در آن، همه رکوردها با کد خود ذخیره شده‌اند. جستجو در این فایل اصلی، بخصوص وقتی تعداد رکورد های آن زیاد می‌شود، بسیار وقت گیر است؛ بنابراین برنامه نویسان، فایلی را تدارک دیده اند که به صورت معکوس، فیلدها را با شماره رکورد های آنها در بر دارد. مثلاً در برابر کلمه Lancaster شماره رکورد هایی که این کلمه را دارند، قرار می‌گیرد و وقتی کلمه Lancaster جستجو می‌شود، برنامه به جای چک کردن تک تک فایلهای رکوردهایی را که شماره آنها در برابر این کلمه در فایل معکوس قرار گرفته اند، بازیابی می‌کند. فایلهای معکوس نسبت به فایلهای امضایی برای ذخیره‌سازی به حافظه بالاتری نیاز دارند.

شایان ذکر است، در زمان تجزیه و تحلیل منابع اینترنتی برچسب‌های<sup>۶</sup> مربوط به ابر داده، از متن اصلی جدا شده و برای تهیه یک نمایه استفاده می‌شوند. (Gietz, 2001)

مدرک بعد از ورود به سیستم، تجزیه و تحلیل می‌شود، بدین صورت که برای مشخص شدن موضوع مدرک، ابتدا کلمات غیر مجاز متن (مثل حروف اضافه یا ربط) از

- 
1. Signature Files.
  2. Stop Word List.
  3. Stemming.
  4. Inverted File.
  5. Master File.
  6. Tag.

آن جدا می‌شود. کلماتی که دارای بسامد<sup>۱</sup> بالا و پایین باشند نیز از متن جدا می‌شوند. کلماتی که دارای بسامد بالا هستند، معمولاً کلمات غیر مجازند و کلماتی که دارای بسامد پایین هستند، درجه اهمیت پایینی دارند. بنابراین، کلماتی که دارای بسامد متوسط هستند، اهمیت بالایی دارند و به عنوان کلیدواژه‌ها انتخاب می‌شوند. در این مرحله، از روشهای آماری استفاده می‌شود (Gietz, 2001).

این مورد، به قانون «زیفت»<sup>۲</sup> نیز اشاره دارد که در تحلیل استنادی از آن استفاده می‌شود. در بعضی از روشها ریشه‌یابی هم صورت می‌گیرد؛ بدین معنا که ریشه کلمات جدا می‌شوند، مثلاً:

در بعضی از روشها ریشه‌یابی<sup>۳</sup> هم صورت می‌گیرد بدین معنی که ریشه کلمات جدا می‌شوند مثلاً:

Measurement —————> measure  
Interaction —————> Interact  
Experimental —————> experiment

بعد از این مرحله نوبت می‌رسد به این که دریابیم کلمات انتخاب شده از کدام قسمت متن یافت شده‌اند. کلماتی که در عنوان، مقدمه، چکیده، عناوین فصول و بحث و نتیجه‌گیری یافت شوند، نسبت به کلمات یافت شده از دیگر قسمتهای متن درجه اعتبار بالایی دارند.

پس از مشخص شدن موضوع و کلیدواژه‌های متن، با استفاده از روشهای آماری و بر حسب اولویت، باید به مدرک، کد رده‌بندی<sup>۴</sup> تعلق بگیری. نرم‌افزار «Scorpion» بر اساس یک پایگاه اطلاعاتی قابل جستجو در فایل رده‌بندی دیویی ساخته شده است.

#### 1. Frequency.

۲. قانون زیف بر بسامد واژه‌ها و مفاهیم موجود در متون تکیه دارد. به موجب این قانون، اگر کلمات موجود در یک مقاله را به ترتیب نزولی بسامد آنها مرتب کنیم، هر یک از کلمات در متن، نسبت معکوس با بسامد رخداد کلمه خواهد داشت (مقصودی، ۱۳۷۸).

#### 3. Stemming.

#### 4. Notation.

رویکردها و چالشهای رده‌بندی خودکار ... / ۱۰۱

بنابراین، مدرکی که می‌خواهد کد رده‌بندی بگیرد، مثل یک سؤال در برابر سیستم رده‌بندی دیویی قرار می‌گیرد و کد رده‌بندی به عنوان نتیجه جستجو به مدرک تعلق می‌گیرد. نرم‌افزار بازیابی آن «SMART» نام دارد (Marcella & Maltby, 2000).

بر اساس طرح رده‌بندی که اتخاذ می‌شود، یک پایگاه اطلاعاتی از سیستم رده‌بندی تهیه شده و به مدرکی که با کلیدواژه‌های مشخص خود وارد این پایگاه می‌شود، یک کد رده‌بندی اختصاص می‌یابد. پروژه «DESIRE II» که فقط در زمینه علوم مهندسی برنامه خود را اجرا کرد، در برنامه رده‌بندی خود کار خود از اصطلاحنامه «EI» که مختص علوم مهندسی است، استفاده کرده است. بعد از ساختن یک نمایه از کلمات بیرون آمده از متن، آن را با واژه‌های یک اصطلاحنامه (اصطلاحنامه ای که در کنار واژه‌های آن، کد های رده‌بندی نیز وجود دارند) مطابقت می‌دهد و پس از جدا کردن واژه‌های غیر مجاز، هر جا که مطابقتی بین اصطلاح موجود در نمایه و اصطلاح موجود در اصطلاحنامه یافت، شماره مربوط را به آن مدرک اختصاص می‌دهد. دقت این برنامه رده‌بندی بین ۵۷ تا ۶۶٪ نشان داده شده است (Toth, 2002).

### اصطلاحنامه<sup>۱</sup>

به کارگیری اصطلاحنامه در رده‌بندی خودکار منابع اطلاعاتی بسیار مفید است. اصطلاحنامه‌ها<sup>۲</sup> مجموعه ای از اصطلاحات مرتبط هستند که با یک نظم سلسله مراتبی، مفاهیم اعم و اخص و کلماتی را که دارای شکل یکسان، ولی معانی گوناگون هستند،<sup>۳</sup> آورده‌اند. پشتیبانی اصطلاحنامه‌ها بسیار گران است و به دانش خاص نیاز دارد. (Gietz, 2001) اصطلاحنامه‌ها نقش مهمی در نظامهای ذخیره و بازیابی اطلاعات دارند. ظهور وب، همراه با توسعه و پیشرفتهای اخیر در کاربرد اصطلاحنامه‌ها به عنوان ابزارهای بازیابی اطلاعات، باعث تولد نسل جدیدی از اصطلاحنامه‌ها شده است. اصطلاحنامه‌های وب محور، راه خود را به محیطهای بازیابی و سازماندهی اطلاعات وب محور باز نموده و در

1. Thesaurus.
2. Thesauri.
3. Homonyms.

تهیه ابر داده‌ها، نمایه‌سازی صفحات، سایت‌های وب، پایگاه‌های داده و موتورهای جستجو استفاده می‌شوند. ساختارهای معنایی موجود در اصطلاحنامه‌ها می‌توانند هم در سازماندهی و هم در بازیابی اطلاعات وب و منابع دانش نقش داشته باشند (رنجبر، ۱۳۸۵). مشکلات روش‌های آماری و زبان‌شناسی رایانه‌ای باعث شده است که این پروژه‌ها فکر استفاده از ابزارهای دیگری را در سرپروارند. امروزه توجه رده‌بندی خودکار از منابع متنی به منابع چند رسانه‌ای نیز جلب شده است.

### شبکه عصبی مصنوعی<sup>۱</sup>

در پروژه‌های یاد شده، از روش‌های اکتشافی، وزن‌دهی اصطلاح و روش‌های زبان‌شناسی رایانه‌ای استفاده شده بود. اگرچه ساختار کلی مراحل رده‌بندی خودکار در پروژه‌ها مشابه هم هستند، اما از فناوریهای متفاوتی استفاده کرده‌اند. به نظر می‌رسد بعضی از فناوریها با بعضی از موضوعات سازگاری بیشتری دارند و سیستمهایی که تنها بر روی یک زمینه موضوعی کار کرده بودند، موفق تر بوده‌اند. همکاریهای بین محققان این طرحها باعث گردیده روش‌های پژوهش آنها تصحیح و بهبود یابد (Gietz, 2001).

نیاز روز افزونی به تغییر و سازگاری سیستمهای رده‌بندی کتابخانه‌ای برای تبدیل شدن به یک ابزار تورق برای منابع اینترنتی به وجود آمده است. برای به وجود آوردن این امکان، به روش‌های «خوشه‌بندی»<sup>۲</sup> و تکنیکهای مبتنی بر محتوا یا مبتنی بر استناد نیاز است. این ویژگیها را در پتانسیل شبکه‌های عصبی مصنوعی می‌توان یافت.

بر اساس تجزیه و تحلیلهای به عمل آمده، مشخص شده است که شبکه‌های عصبی مصنوعی به خوبی اصطلاحنامه‌های سلسله مراتبی می‌توانند در این زمینه موفق باشند.

با افزایش حجم اطلاعات و پیشرفت فناوریهای اطلاعاتی، استفاده از الگوریتمهای سنتی برای بازیابی سریع داده‌ها کافی نیست. استفاده از شبکه‌های عصبی مصنوعی به عنوان یکی از تکنیکهای هوش مصنوعی، راهکار مناسبی برای افزایش سرعت بازیابی اطلاعات در حجم انبوه است. شبکه‌های عصبی مصنوعی بازنمون مناسبی از دانش برای

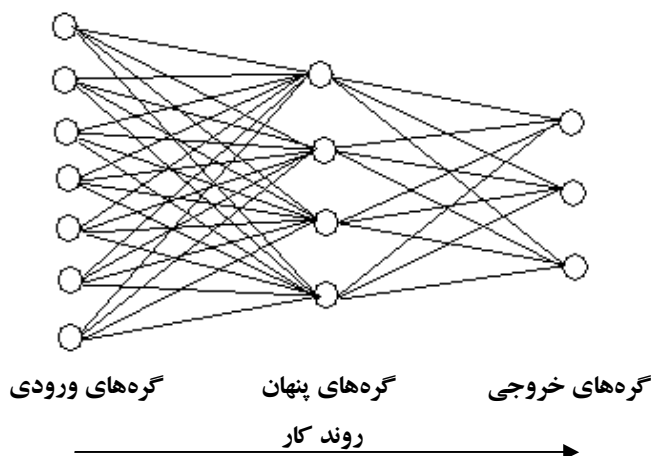
۱. Artificial Neural Network شبکه عصبی مصنوعی نمونه شبیه سازی شده از شبکه عصبی انسان است که از مدل‌های آن در پردازش داده‌ها، رده‌بندی و پیش‌بینی نتایج کار استفاده می‌شود (wikipedia).

2. Clustering.

رویکردها و چالشهای رده‌بندی خودکار ... / ۱۰۳

کاربردهای بازیابی اطلاعات ارائه می‌دهند. گره‌های شبکه عصبی نمایانگر عناصر بازیابی اطلاعات از قبیل کلید واژه، نویسنده و ... هستند و پیوندهای موجود در شبکه برای انتقال ورودی از لایه به لایه دیگر استفاده شده و در نتیجه به دستیابی به خروجی شبکه، که همان بازیابی مدرک است، می‌انجامد. شبکه‌های عصبی چند لایه، از تعدادی گره و پیوند تشکیل شده است. اطلاعات از طریق گره‌های ورودی به شبکه عصبی وارد شده، سپس با استفاده از پیوندها به لایه‌های بعدی (پنهان) منتقل گردیده و در نهایت، خروجی شبکه از گره‌های لایه خروجی به دست می‌آید (شکل ۲).

با در نظر گرفتن خصوصیات شبکه‌های عصبی، به نظر می‌رسد این تکنیک هوش مصنوعی برای بازیابی اطلاعات، مؤثر باشد. در حال حاضر، در مدل‌های کاربردی شبکه‌های عصبی در بازیابی اطلاعات، تحقیقاتی صورت گرفته است و در آینده با پیشرفت بیشتر سخت افزار و نرم افزار، به نظر می‌رسد به سرعت بتوان از شبکه‌های عصبی استفاده‌های مؤثرتری نمود. با ارزانتر شدن سخت افزار مورد نیاز مدل‌های شبکه عصبی، امکان استفاده از آنها، با استفاده موازی برای طبقه بندی مدارک، امکان پذیر شده و بازیابی اطلاعات با استفاده از شبکه‌های عصبی مصنوعی، بسیار سریعتر خواهد شد (Mehrad & Koleini, 2007). همچنین، با پیشرفتهای نرم‌افزاری در ایجاد روشهای جدید کنترل توابع شبکه، به منظور پیاده‌سازی مدل‌های شبکه عصبی، تحولی مهم در سرعت بازیابی اطلاعات فراهم می‌شود.



شکل ۲. نمونه یک شبکه عصبی

### نتیجه‌گیری

دنیای جدید اطلاعات با حجم روزافزون منابع اینترنتی دچار مسئله‌ای شده است که جز با کشف راهی مناسب برای سازماندهی این اطلاعات قابل حل نیست. رده‌بندی یکی از شیوه‌های سازماندهی اطلاعات برای بازیابی سریع آنهاست که با استفاده از طرح‌های رده‌بندی مثل رده‌بندی دیویی، رده‌بندی دهدهی جهانی و رده‌بندی کنگره صورت می‌گیرد. از آنجا که رده‌بندی، فعالیتی وقت‌گیر و پرهزینه بوده و فهرست‌نویسی منابع اینترنتی، خارج از توان فهرست‌نویسان است، پروژه‌های زیادی برای انجام رده‌بندی به صورت خودکار و با ماشین، اجرا شده‌اند. استفاده از تکنیک‌های هوش مصنوعی چون شبکه‌های عصبی مصنوعی، در تحقق این هدف سؤال‌هایی مؤثر است. اما در کاربرد طرح مناسبی برای انجام رده‌بندی به صورت خودکار، مطرح است که مؤسساتی چون «او.سی.ال.سی» تحقیقات زیادی برای پاسخگویی به آنها آغاز کرده‌اند.

§ آیا طرح‌های رده‌بندی کتابخانه‌ای مانند رده‌بندی دهدهی دیویی و رده‌بندی کتابخانه کنگره می‌توانند برای رده‌بندی منابع دیجیتالی به کار بروند؟ و آیا نیازی به تولید سیستم رده‌بندی مخصوصی برای رده‌بندی منابع دیجیتالی متناسب با ویژگی‌های خاص آنها ضرورت دارد؟

§ آیا می‌توان از این رده‌بندیها برای رده‌بندی خودکار منابع اینترنتی و دیگر منابع دیجیتالی بهره برد؟

§ چه پیشرفتهایی لازم است تا سیستم‌های رده‌بندی خودکار، این کار را به دقت یک متخصص رده‌بندی انجام دهند؟

§ آیا ایجاد مرکزی واحد برای دادن کد رده‌بندی به صفحات وب به صورت اجباری ضرورت دارد؟

### منابع

- رنجبر، جهانگیر (۱۳۸۵). «آیا در محیط اطلاعاتی جدید جایی برای اصطلاحنامه‌ها باقی مانده است؟»، کتابداری و اطلاع‌رسانی، جلد ۹، شماره اول، صص ۱۱۴-۱۱۵.
- مقصودی، رؤیا (۱۳۷۸). «بررسی استنادی و تطبیقی پایان‌نامه‌های کارشناسی ارشد روانشناسی تربیتی دانشگاه‌های شهید چمران اهواز و شیراز در سالهای ۱۳۷۲ الی ۱۳۷۶».
- پایان‌نامه کارشناسی ارشد علوم کتابداری و اطلاع‌رسانی، دانشگاه شهید چمران اهواز، ص ۳۶.

۱۰۵ / ... رویکردها و چالشهای رده‌بندی خودکار

---

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

ضمیمه

دو نمونه از وب‌سایت‌هایی که کاربران برای سازماندهی صفحات وب خود از طرح‌های رده‌بندی استفاده کرده‌اند:

۱. رابط کاربر wplib در دانشگاه والور هامپتون با استفاده از رده‌بندی دهدهی دیویی



۲. وب‌سایت Kids Click (<http://www.kidsclick.org/dewey.html>)

سایت Beyond Bookmark: Schemes for organizing the web  
وب‌سایت‌هایی را که از طرح‌های رده‌بندی در سازماندهی اطلاعات خود استفاده کرده‌اند، معرفی می‌کند. آدرس این وب‌سایت:

<http://www.public.iastate.edu/CYBERSTACKS/CTW.htm#Numeric>

