

## Applications of Natural Language Processing in Information Science and Knowledge with Emphasis on Digital Libraries

**Mahboubeh Rabiei** 

Phd Candidate of Information Science in Al-Zahra University and Expert of Data Processing at National Library of Iran, Tehran, Iran

**Vahid Reza Mirzaian** 

Assistant Professor, Department of Technology Application in Language Education, Al-Zahra University, Tehran, Iran

### Abstract

Natural language processing as a branch of computational linguistics whose main effort is to use computers in the process of automating the understanding and processing of human natural language and focusing on human-computer interaction has found an important place in various fields of science including information science and knowledge. The main purpose of this study is to identify the sub-branches and sub-fields of information science and knowledge in which natural language processing has been effective, and has been done through library and documentary analysis. The study deals with the role of digital libraries in the field of information science and application of natural languages processing in them. The result of this study shows that natural language processing in many sub-fields related to information science such as information retrieval, bibliometric, document management, automatic information extraction, automatic indexing, automatic text summarization, automatic text classification, question and answer systems and using spell checker technology, debugging user query phrase and predicting their preferred words, translating speech in to text and vice versa and helping users with physical disabilities such as the visually impaired and the blind, surveying and analyzing the sense of libraries and information centers is Traceable.


**Keywords:** Natural Language, Processing Information Science, Digital Libraries.

Corresponding Author: m.rabiei@alzahra.ac.ir


**How to Cite:** Rabiei, M., Mirzaian, V. R. (2023). Applications of Natural Language Processing in Information Science and Knowledge with Emphasis on Digital Libraries, *Journal of Knowledge Retrieval and Semantic Systems*, 9(33), 197-262.

## کاربردهای پردازش زبان طبیعی در علم اطلاعات و دانش‌شناسی با تأکید بر کتابخانه‌های دیجیتال

دانشجوی دکتری علم اطلاعات و دانش‌شناسی گرایش بازیابی اطلاعات  
دانشگاه الزهراء، تهران، ایران

\*  محبوبه ربیعی

استادیار گروه کاربرد فناوری در آموزش زبان، دانشگاه الزهراء، تهران، ایران

 وحیدرضا میرزاییان

### چکیده

پردازش زبان طبیعی به‌عنوان شاخه‌ای از زبان‌شناسی محاسباتی که تلاش عمده آن در جهت به‌کارگیری رایانه در فرآیند خودکارسازی درک و پردازش زبان طبیعی انسان و تمرکز بر تعامل میان انسان و رایانه است، در حوزه‌های مختلف علوم از جمله علم اطلاعات و دانش‌شناسی، جایگاه مهمی یافته است. مطالعه حاضر باهدف اصلی شناسایی زیرشاخه‌ها و حوزه‌های فرعی علم اطلاعات و دانش‌شناسی که پردازش زبان طبیعی در آن‌ها مؤثر واقع شده است و به شیوه کتابخانه‌ای یا تحلیل اسنادی انجام پذیرفته است، این مطالعه ضمن اشاره به حوزه‌هایی از علم اطلاعات که پردازش زبان طبیعی در آن‌ها مفید واقع شده، به نقش کتابخانه‌های دیجیتال در عرصه علم اطلاعات و کاربرد پردازش زبان طبیعی در آن‌ها می‌پردازد. نتایج حاصل از این مطالعه نشان می‌دهد که پردازش زبان طبیعی در بسیاری از حوزه‌های فرعی و مرتبط با علم اطلاعات مانند بازیابی اطلاعات، کتاب‌سنجی، مدیریت اسناد و مدارک، استخراج خودکار اطلاعات، نمایه‌سازی خودکار، خلاصه‌سازی خودکار متون، طبقه‌بندی خودکار متون، نظام‌های پرسش و پاسخ و به‌کارگیری فناوری خطایاب املائی، ابهام‌زدایی از عبارات پرسش کاربران و پیش‌بینی واژه‌های موردنظر آن‌ها، تبدیل گفتار به متن و بالعکس و یاری‌رساندن به کاربران دارای معلولیت‌های جسمی مانند کم‌بینایان و نابینایان، نظر کاوی و تحلیل احساس واژگان مورد استفاده کاربران کتابخانه‌ها و مراکز اطلاع‌رسانی، قابل ردیابی است.

کلیدواژه‌ها: پردازش زبان طبیعی<sup>۱</sup>، علم اطلاعات و دانش‌شناسی<sup>۲</sup>، کتابخانه‌های دیجیتال<sup>۳</sup>.

\* نویسنده مسئول: m.rabiei@alzahra.ac.ir

1. Natural language processing
2. Knowledge and Information science
3. Digital libraries

## مقدمه

با پیشرفت‌های اخیر در فن‌آوری‌های اطلاعات، پردازش زبان طبیعی در بسیاری از زمینه‌ها، کارها را آسان‌تر و عملی‌تر ساخته است. امروزه، به‌ویژه هنگامی که از داده‌های بزرگ در بیشتر تحقیقات استفاده می‌شود، پردازش زبان طبیعی روش‌های آسان و سریع برای پردازش این داده‌ها را فراهم می‌آورد. گرچه پردازش زبان طبیعی در ابتدا برای جلوگیری از انقراض برخی زبان‌های در معرض نابودی بوجد آمد، اما امروزه این رویکردها و فنون برای مطالعه در حوزه سازمان‌دهی و معنی‌بخشیدن به داده‌های بزرگ به کار می‌رود. بدون استفاده از روش‌های پردازش زبان طبیعی بسیاری کارها از قبیل اعتبارسنجی، بازیابی و مصورسازی اطلاعات بسیار وقت‌گیر و زمان‌بر خواهد بود (تاسکین و آل<sup>۱</sup>، ۲۰۱۹). ما در دوره‌ای زندگی می‌کنیم که کاربردهای پردازش زبان طبیعی به یک جریان اصلی بدل شده است. این حوزه به مرحله بلوغ رسیده و استحکام آن را می‌توان در پذیرش و به‌کارگیری فناوری‌های پردازش زبان طبیعی مشاهده کرد. در هسته اصلی هر فعالیت مبتنی بر پردازش زبان طبیعی مسئله مهم شناخت زبان طبیعی نهفته است، مطابق تعریف فرهنگ واژگان آکسفورد پردازش زبان طبیعی استفاده از فنون محاسباتی جهت تجزیه و تحلیل و پردازش زبان و گفتار طبیعی است. پردازش زبان طبیعی رویکردی رایانه‌ای جهت تجزیه و تحلیل متون بر اساس مجموعه‌ای از نظریه‌ها و فناوری‌هاست. گرچه پردازش زبان طبیعی حوزه تحقیقاتی و کاربردی نسبتاً جدید است، اما در مقایسه با سایر رویکردهای فن‌آوری اطلاعات، موفقیت‌هایی که تاکنون به دست آورده، نشان می‌دهد فناوری‌های دسترسی به اطلاعات مبتنی بر پردازش زبان طبیعی همچنان به‌عنوان یکی از حوزه اصلی تحقیق و توسعه در نظام‌های اطلاعاتی در زمان فعلی و در آینده خواهد بود (لیدی<sup>۲</sup>، ۲۰۱۸).

افزایش مجموعه‌های بزرگ دیجیتالی و کتابخانه‌های دیجیتال و ظهور ارزش اقتصادی اطلاعات، نیازمند مدیریت انبوه اطلاعات و مدارک در دسترس است و این امر

---

1.Taskin, Zehraand Al, Umut.

2. Liddy,Elizabeth D.

نیز مستلزم راه‌حل‌های کارآمد است؛ مهم‌ترین آن‌ها فناوری پردازش زبان طبیعی است که در تسهیل مدیریت مدارک بسیار مؤثر واقع شده است. محققان حوزه پردازش زبان طبیعی، باهدف جمع‌آوری دانش در چگونگی درک و استفاده بشر از زبان به‌نحوی که ابزارها و فنون مناسب، قادر به گسترش نظام‌های رایانه‌ای به‌منظور شناخت و دست‌کاری زبان طبیعی باهدف انجام کارهای مدنظر باشند، نقش مهمی ایفا را می‌کنند.

پردازش زبان طبیعی در بسیاری رشته‌ها از جمله علوم کامپیوتر و علم اطلاعات، زبان‌شناسی، ریاضیات، مهندسی الکترونیک، هوش مصنوعی و رباتیک وارد شده است. کاربردهای پردازش زبان طبیعی شامل حوزه‌های مطالعاتی‌ای مانند ترجمه ماشینی، خلاصه‌سازی و پردازش متون زبان طبیعی، رابط کاربری، بازیابی اطلاعات چندزبانه و بین‌زبانی<sup>۱</sup>، تشخیص گفتار، هوش مصنوعی و نظام‌های خبره و مانند آن‌ها است؛ و اما یکی از حوزه‌های کاربردی نسبتاً جدید در پردازش زبان طبیعی که بسیار مورد توجه واقع شده است، شبکه جهانی وب و افزایش کتابخانه‌های دیجیتال است (چودوری<sup>۲</sup>، ۲۰۰۳). محققان بسیاری به نیاز تحقیق در حوزه‌های کاربردی پردازش زبان طبیعی اشاره کرده‌اند، از جمله تحقیق در بازیابی اطلاعات چند زبانی و بین‌زبانی شامل پردازش متون چندزبانه و نظام‌های رابط کاربر چندزبانه به‌منظور بهره‌برداری کامل از کتابخانه‌های دیجیتال (برای مثال بورگمن<sup>۳</sup>، ۱۹۹۷؛ پترز و پیچی<sup>۴</sup>، ۱۹۹۷).

مجموعه‌های روبه رشد از منابع باارزش فرهنگی شامل کتابخانه‌های دیجیتال نوظهور، آرشیوها و موزه‌ها به‌طور فزاینده‌ای در حال انتقال مواد و منابع اولیه و اصلی خود به مواد دیجیتالی جدید هستند، اعم از موادی که در محیط‌های شبکه یافته و یا موادی که قابل جابجایی و انتقال (برای مثال لوح فشرده، دیسک سخت، حافظه جانبی) در محیط‌های نگهداری پایدارتر هستند. متخصصان اطلاعات باید برای استخراج مواد دیجیتالی از رسانه‌های جدا پذیر به شیوه‌ای که منعکس‌کننده فراداده‌های غنی باشد و از یکپارچگی

1. multilingual and cross language information retrieval (CLIR)
2. Chowdhury, G.
3. Borgman, Christine.L.
4. Peters, C. and Picchi, E.

مواد اطمینان حاصل کنند، آمادگی کسب کنند. همچنین آن‌ها می‌بایست به‌عنوان واسط در دسترسی مناسب کاربران را برای رسیدن به مواد و محتوای موردنظرشان درحالی که از افشای غیرعمدی و دسترسی به اطلاعات حساس جلوگیری می‌کنند حمایت کنند. استفاده از منابع اصلی و اولیه اغلب بر شناسایی و ردیابی موجودیت‌های آن منبع مانند (افراد، مکان‌ها، سازمان‌ها و وقایع) و سایر موجودیت‌ها (برای مثال تاریخ و زمان) در سراسر مدارک متمرکز است. تعداد زیادی ابزارهای پردازش زبان طبیعی منبع باز وجود دارد که قابلیت تشخیص و گزارش موجودیت‌های نام‌برده را دارند و پروژه‌های بسیاری در حوزه علوم انسانی وجود دارد که بر ارزش رویکردهای پردازش زبان طبیعی هنگام کار با منابع دیجیتالی تمرکز داشته‌اند. تا به امروز ابزارهای نسبتاً اندکی جهت پردازش منابع دیجیتالی نوظهور توسط کتابخانه‌ها، آرشیوها و موزه‌ها وجود داشته است. چالش‌های مختلفی در رابطه با استفاده از ابزارهای پردازش زبان طبیعی در مجموعه منابع اولیه دیجیتالی نوظهور، شامل آن دسته از منابعی که به‌صورت قانونی از رسانه‌های جدا پذیر به وجود آمدند، وجود دارد. بسیاری از این چالش‌ها به دلیل تنوع مواد و منابع و موارد استفاده بالقوه آن‌هاست (لی و وودز<sup>۱</sup>، ۲۰۱۷).

امروزه پژوهش‌ها در حوزه پردازش زبان طبیعی از ترجمه ماشینی به سمت بازیابی اطلاعات در حرکت است. افزایش دسترسی به مجموعه‌های بزرگی از مدارک دیجیتالی، علاقه به طراحی و ساخت ابزارها و فناوری‌های پردازش زبان طبیعی جهت مدیریت و رسیدگی به این مجموعه‌ها را برانگیخته است؛ به‌ویژه مفهوم کتابخانه‌های دیجیتال با عملکرد و معماری خاص ظهور کرده است. کتابخانه دیجیتال محیطی است که بسیاری از کاربردهای پردازش زبان طبیعی می‌توانند در آن نقش ایفا کنند. فضایی که در آن بازیابی اطلاعات به شیوه سنتی و به میزان اندکی از حوزه پردازش زبان طبیعی استفاده می‌کرد، اکنون نیازمند روش‌های پیچیده‌تر و پیشرفته‌تر مبتنی بر پردازش زبان طبیعی است، چراکه کارکردهای کتابخانه دیجیتال فراتر از بازیابی ساده اطلاعات است (کارو<sup>۲</sup>، ۲۰۱۷).

1. Lee, Christopher A. and Woods, Kam
2. Karoo, Krishna

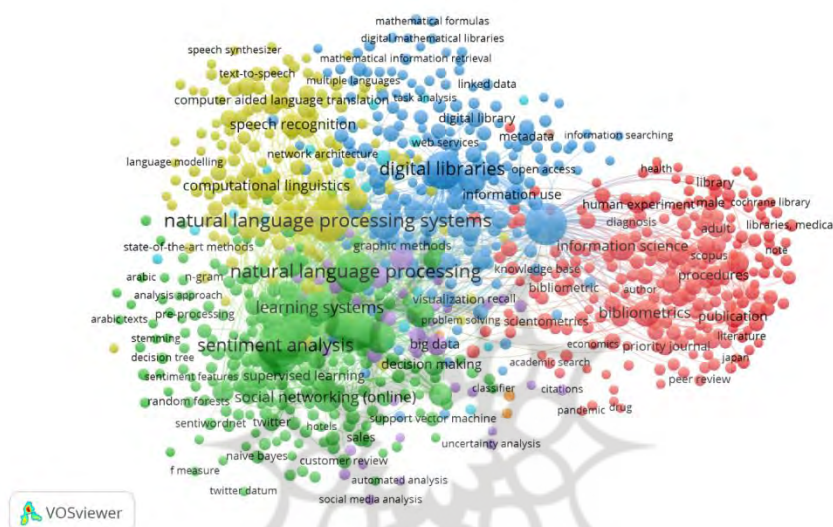
گرچه کاربردهای مختلف پردازش زبان طبیعی مانند ترجمه ماشینی بسیار مهم است، اما هدف ما در این مطالعه صرفاً بررسی کاربردهای پردازش زبان طبیعی در علم اطلاعات و به‌ویژه کاربرد مؤثر آن در کتابخانه‌های دیجیتال است. مطالعه حاضر باهدف اصلی شناسایی زیرشاخه‌ها و حوزه‌های فرعی علم اطلاعات و دانش‌شناسی که پردازش زبان طبیعی در آن‌ها مؤثر واقع شده است.

روش‌شناسی: در مطالعه حاضر که بر اساس تحلیل اسنادی انجام پذیرفته است، برای تحلیل میزان ارتباط علم اطلاعات با پردازش زبان طبیعی و شناسایی زیرشاخه‌ها و حوزه‌های فرعی علم اطلاعات و دانش‌شناسی که پردازش زبان طبیعی در آن‌ها مؤثر واقع شده است، در ابتدا کلیدواژه‌های پرکاربرد در پردازش زبان طبیعی با بررسی منابع مختلف و پرسش از متخصصان استخراج گردید و سپس عبارت پرسش ایجاد شده در پایگاه اسکوپوس (به جهت اینکه پوشش پایگاه اسکوپوس بیشتر از وب آو ساینس است) جستجو گردید. عبارت پرسش مورد جستجو در عنوان مدارک به شرح زیر است:

Title ("natural language processing"OR" NLP "OR "text categorization"OR "information retrieval" OR"information extraction"OR "machine translation" OR "automatic indexing"OR "machine indexing "OR "text summarization "automatic summarization"OR "question and ask systems"OR " bibliometric"OR"word prediction "OR" spell checking"OR"text to speech""OR "speech to text "opinion mining "OR "ambiguity of words "OR"digital libraries") AND"informationscience"

مطابق جستجوی انجام گرفته تعداد ۲۱۰۹ پیشینه بر اساس کلیدواژه‌های جستجو شده در پایگاه اسکوپوس مرتبط با حوزه‌های پردازش زبان طبیعی و علم اطلاعات و دانش‌شناسی بازیابی شد و مورد بررسی قرار گرفت. این پیشینه‌ها شامل ۱۱۶۹ مقاله، ۷۶۳ مقاله کنفرانسی، ۱۱۷ نقد و بررسی، ۲۲ نقد و بررسی کنفرانس، ۱۳ مورد فصلی از کتاب، ۱۰ کتاب، ۵ سرمقاله روزنامه، ۵ مورد غلط‌نامه، ۲ نامه و دو یادداشت می‌باشد. از میان کل پیشینه‌های بازیابی شده بیشترین تعداد از لحاظ نوع پیشینه مربوط به مقالات منتشر شده (۱۱۶۹ مورد) و به لحاظ نوع دسترسی بیشترین آن‌ها (۳۸۵ مقاله) مربوط به مقاله‌های دسترسی آزاد و زبان انگلیسی (۱۹۴۰ مورد) و بالاترین تعداد آن‌ها در کشور ایالات متحده

(۴۳۷ مورد) بوده است. در نهايت نتايج حاصله پس از انتقال از پایگاه اسکوپوس به نرم افزار اکسل، به وسيله نرم افزار وی آ اس ویور<sup>۱</sup> بصري سازی شده و به شکل نمودار به نمایش درآمد (تصویر ۱). سپس منابع بازيابی شده مورد مطالعه قرار گرفت و برجسته ترين کاربردهای پردازش زبان طبيعي مرتبط با علم اطلاعات و دانش شناسی استخراج گردید.



شکل ۱: ساختار شبکه‌ای مجموع پرکاربردترین کلیدواژه‌های به کاررفته در مقالات پایگاه اسکوپوس در حوزه پردازش زبان طبيعي و علم اطلاعات.

### سطوح و ظايف و کارکردهای پردازش زبان طبيعي

پردازش زبان طبيعي به هفت سطح اساسی، از ساده ترين تا مشکل ترين تقسیم می شود. (فلدمن<sup>۲</sup>، ۱۹۹۹؛ لیدی<sup>۳</sup>، ۲۰۱۰). به این ترتیب، وظايف پردازش زبان طبيعي برای بسیاری مطالعات تحقیقاتی و متون حجیم در هر زبانی به راحتی قابل تجزیه و تحلیل است. سطح اول «آواشناسی» است که واحد مورد مطالعه در این سطح آوا می باشد که مطالعه اصوات گفتار انسان را تشکیل می دهد و با خواص فیزیکی آواها یا اصوات سروکار دارد. در

1. VOSviewer
2. Feldman, S.
3. Liddy, Elizabeth D.

پژوهش‌های بسیاری، در مورد زبان‌ها برای جلوگیری از انقراض آن‌ها مطالعات مختلفی انجام شده است. مطالعات انجام شده در «سطح آوایی» برای درک زبان‌های گفتاری یا تلفظ‌ها طراحی شده است. در سطح دوم «واج‌شناسی»، واحد مورد مطالعه واج می‌باشد، سطح بعدی، «تک‌واژشناسی»، با کوچک‌ترین قسمت‌های کلمات به نام «تک‌واژ» سروکار دارد. در به‌کارگیری تک‌واژشناسی ویژگی‌های اصلی زبان‌ها را می‌توان آشکار کرد و ریشه و پسوند کلمات را تشخیص داد (برای مثال کاربردهای مشابه برای زبان ترکی که یک‌زبان ترکیبی در دنیا است). هدف اصلی سطح واژگان، درک معانی کلمات است. سطح چهارم یا «سطح نحوی» یا جمله‌شناسی، ساختارهای دستوری جملات و چندین جمله را نمایان می‌سازد و مطالعاتی در مورد آن انجام شده است، درحالی‌که هدف «سطح معنایی» آشکار ساختن معنای واژگان است و به بررسی معانی در زبان‌های انسانی می‌پردازد. واحد مورد مطالعه در سطح نحوی عبارت یا جمله می‌باشد، این دو سطح (معنایی و نحوی) برای مطالعات سطح زبان گفتاری ترکیب می‌شوند. سطح ششم، سطح عملی یا «کاربردشناسی» است که هدف آن درک هدفمند کاربرد زبان‌ها در موقعیت‌ها و شرایط گوناگون است و سطح آخر «گفتارشناسی» یا تحلیل گفتمان، پیچیده‌ترین سطح است. برای دستیابی موفقیت‌آمیز به وظایف پردازش زبان طبیعی، لازم است همه یا تعدادی از این سطوح مورد بررسی قرار گیرند، هر سطح ادامه سطح قبلی است. به‌عبارت‌دیگر در حالی‌که ساده‌ترین کار پردازش زبان طبیعی در سطح آوایی انجام می‌شود، پیچیده‌ترین آن در سطح کاربردی انجام می‌شود (تاسکین و آل، ۲۰۱۹). در تحلیل گفتمان، برخلاف تحلیل‌های سنتی زبان‌شناسی، صرفاً با عناصر نحوی و لغوی تشکیل‌دهنده جمله سروکار نداریم، بلکه فراتر از آن به عوامل بیرونی متن، مانند بافت موقعیتی، فرهنگی، اجتماعی، سیاسی و ... مرتبط است. تحلیل گفتمان روشی نوین در پژوهش متن‌های ارتباطی است.

### کاربردهای پردازش زبان طبیعی در علم اطلاعات

برای درک قدرت پردازش زبان طبیعی و تأثیر آن بر زندگی ما، باید نگاهی به کاربردهای آن بی‌اندازیم. بسیاری از انواع کاربردهای پردازش زبان طبیعی بر اساس روش‌های



داده‌محورمانند شبکه‌های عصبی و مدل پنهان مارکوف<sup>۱</sup> بنا شده است. از آنجایی که پیشرفت در اکثر این روش‌ها با بهره‌گیری از داده صورت می‌پذیرد، داده‌های بزرگ و با کیفیت بالا به منابع بسیار باارزشی تبدیل شده‌اند. به عنوان مثال، برخی از اثرات مفید آن‌ها در ترجمه ماشینی، ابهام‌زدایی از معنی کلمه، خلاصه‌سازی، حاشیه‌نویسی نحوی، شناسایی موجودیت‌های برچسب‌گذاری شده در میان سایر کاربردهای پردازش زبان طبیعی قابل مشاهده است. داده‌های اولیه که با گذشت قرن‌ها همچنان به صورت دستی در قالب مخازنی از متون بدون ساختار به نام بایگانی گردآوری می‌شدند، در اوایل دهه ۱۹۸۰، انقلاب در صنعت رایانه منجر به روش جدیدی برای پردازش داده‌ها در درجه اول از نظر ظرفیت ذخیره‌سازی شده است. شکل جدیدی از کل داده‌های الکترونیکی که پایگاه داده نامیده می‌شود، برای سهولت ورود و بازیابی داده‌ها طراحی شد و چند سال بعد دانشمندان پردازش زبان طبیعی به زیرمجموعه‌هایی از پایگاه داده نام «پیکره زبانی» دادند، به مفهوم مجموعه‌ای از نمونه‌های متنی که به‌طور طبیعی وجود دارند و بر اساس برخی معیارهای صریح جمع‌آوری می‌شوند تا زبان را نشان دهند (زروالا و لاخوآجا،<sup>۲</sup> ۲۰۱۸).

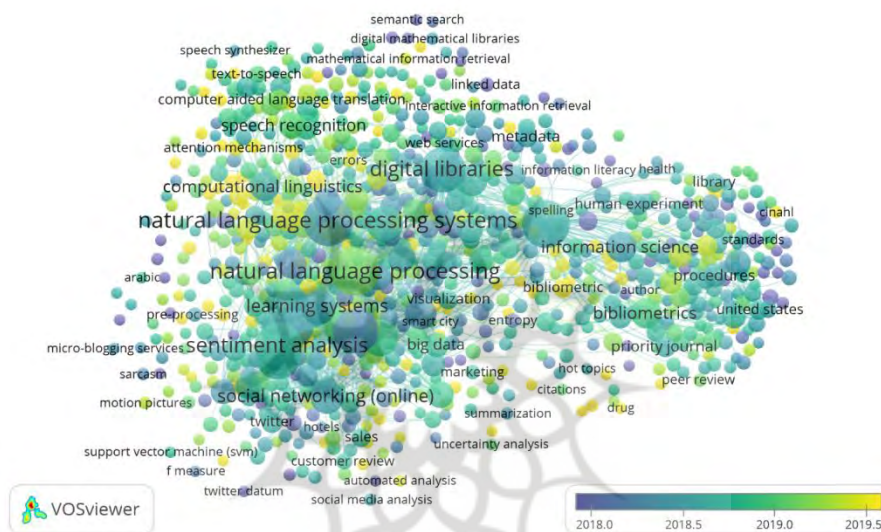
## یافته‌ها

بر اساس مطالعه حاضر کاربردهای پردازش زبان طبیعی در علم اطلاعات شامل بازیابی اطلاعات، متن کاوی و متن کاوی بصری، تحلیل احساس واژه و عقیده کاوی، استخراج خودکار اطلاعات، کتابخانه‌های دیجیتال، نمایه‌سازی خودکار، خلاصه‌سازی خودکار اطلاعات، طبقه‌بندی خودکار متون، اصطلاح‌نامه‌ها، نظام‌های پرسش و پاسخ، کتاب‌سنجی، تبدیل گفتار به نوشتار و بالعکس می‌باشد. همچنین از آنجا که اهداف مختلفی برای پردازش متون زبان طبیعی وجود دارد، بیشترین کاربردهای پردازش زبان طبیعی در گذشته مربوط به حوزه‌های زیر بیان شده است: بازیابی اطلاعات، استخراج خودکار اطلاعات، ترجمه ماشینی، خلاصه‌سازی و طبقه‌بندی خودکار متون. بر اساس تحلیل نموداری پیشینه‌های

1. Markov model

2. Zerual, Imad and Lakhouaja, Abdelhak

حاصل‌شده، مشخص شد علاوه بر اینکه که پردازش زبان طبیعی با علم اطلاعات رابطه نزدیکی دارد، همچنین مطابق بررسی کلیدواژه‌ها می‌توان نتیجه گرفت که حوزه موضوعی تحلیل احساس واژه‌ها<sup>۱</sup> و عقیده کاوی<sup>۲</sup> و نیز حوزه شبکه‌های اجتماعی<sup>۳</sup> و کتابخانه‌های دیجیتال و مطالعات کتاب‌سنجی در سال‌های اخیر در پژوهش‌های حوزه پردازش زبان طبیعی و علم اطلاعات مورد استقبال فراوانی قرار گرفتند (تصویر ۲).



شکل ۲. ساختار شبکه‌ای پرکاربردترین کلیدواژه‌های به‌کاررفته در مقالات پایگاه اسکوپوس در

حوزه پردازش زبان طبیعی و علم اطلاعات بین سال‌های ۲۰۱۸-۲۰۲۰

## بازیابی اطلاعات

بازیابی اطلاعات یکی از بدیهی‌ترین فرایندها در چرخه زیستی اطلاعات است، از این رو بازیابی کارآمد اطلاعات، حتی فرایندهای پیش از خود نظیر سازمان‌دهی و ذخیره‌سازی اطلاعات را نیز تحت تأثیر قرار می‌دهد. به بیان دیگر اطلاعات هرچند دارای بار ارزشی بالایی باشد، در صورت عدم بازیابی چیزی جز کالای منسوخ‌شده نخواهند بود

1. Sentiment analysis
2. Opinion mining
3. Social networking

(فرهادپور و مطلبی، ۱۳۹۰). رشد فزاینده اطلاعات از یک سو و ارزش تصمیم گذاری مبتنی بر اطلاعات جهت استفاده مناسب از فرصت‌ها در محیط رقابتی پیچیده از سوی دیگر، بازیابی اطلاعات را به یکی از مهم‌ترین دغدغه‌های حوزه علم اطلاعات و کتابخانه‌ها و مراکز اطلاع‌رسانی تبدیل کرده است.

هدف اصلی تبدیل بازیابی اطلاعات به یکی از نقاط کانونی پژوهش‌های حوزه اطلاعات را باید در تلاش برای مطالعه و درک فرایندهای بازیابی اطلاعات باهدف طراحی، ساخت و ارزیابی نظام بازیابی دانست تا بتواند بین اطلاعات تولیدشده توسط عوامل انسانی و نیاز کاربران ارتباط برقرار کند (اینگورسن<sup>۱</sup>، ۲۰۰۲). مفهوم اصلی نظام‌های بازیابی اطلاعات در ارائه مدارک مرتبط با نیازهای کاربران نهفته است (بورلاند<sup>۲</sup>، ۲۰۰۳)؛ و این نقطه چالش اصلی نظام‌های بازیابی اطلاعات محسوب می‌شود، چراکه پژوهش‌ها باگذشت زمان از تمرکز صرف بر ساختار نظام‌های بازیابی اطلاعات فراتر رفته و به مسائل مربوط به کاربرد از قبیل طراحی رابط کاربری، معیارهای داوری ربط کاربر و... سوق داده شده‌اند.

در سال‌های اخیر با گسترش روزافزون اطلاعات، کاربرد پردازش زبان طبیعی برای دستیابی به اطلاعات معنادار اهمیت بیشتری یافته است که هدف اصلی آن در مطالعات مختلف تحت عنوان بازیابی مدرک یا بازیابی متن تعریف شده است و امکان دسترسی افراد به هر بخش از پاراگراف، کتاب یا متون حجیم را فراهم می‌سازد (لوئیس و جونز<sup>۳</sup>، ۱۹۹۶). بازیابی اطلاعات را می‌توان به‌عنوان یک موفقیت بزرگ در پردازش زبان طبیعی در نظر گرفت. نظام‌های بازیابی اطلاعات مجموعه‌ای از مدارک به زبان طبیعی را باهدف بازیابی دقیق مدارک مرتبط با نیاز کاربر عمل جستجو را انجام می‌دهند. برخلاف نظام پایگاه‌های داده که نیازمند داده‌هایی با ساختار بالا و دارای معنانشناسی صوری هستند، نظام‌های بازیابی با متون بدون ساختار زبان طبیعی سروکار دارند و برخلاف نظام‌های بازیابی خبره تلاشی

- 
1. Ingwersen, Peter
  2. Borlund
  3. Lewis, David and Jones Karen Sparck

در جهت استنباط پاسخ‌ها نمی‌کنند و صرفاً مجموعه‌ای از مدارک را که مرتبط با سؤال کاربر است در پاسخ به او ارائه می‌دهند. رشد بی‌رویه و چشمگیر تعداد منابع تمام متن به زبان طبیعی که به شکل الکترونیکی در دسترس است، ابزارهایی را می‌طلبد که به کاربران دریافتن مدارک موردنیازشان یاری رسانند (وورهیز، ۲۰۰۰).

بازیابی اطلاعات را می‌توان ابزاری در نظر گرفت که نیازهای اطلاعاتی کاربران را برطرف می‌سازد.<sup>۱</sup> بهترین روش شناخته‌شده برای رسیدن به این هدف بازیابی مدرک است که در آن موتورهای جستجو، مجموعه‌ای از مدارک را جستجو می‌کنند تا مدارک موردنیاز کاربر را شناسایی کنند (راسل رز و استیونسون<sup>۲</sup>، ۲۰۰۹). فرض کنید که می‌خواهید در مورد مسئله‌ای تحقیق کنید، در این صورت پرسشی را در قالب عبارت جستجو در نظام‌های رایانه‌ای مطرح می‌کنید و مدارکی مطابق درخواست‌تان، به شما برگردانده می‌شود، این نظام‌ها را «نظام‌های بازیابی اطلاعات» نامیده‌اند؛ گرچه اخیراً بیشتر آن‌ها را بازیابی مدارک یا نظام‌های بازیابی متن می‌نامند (وورهیز، ۲۰۰۰)، این روش دارای محدودیت‌هایی است که کاربران را ملزم به انجام کارهای اضافی برای شناسایی اطلاعات خود می‌کند، به این مفهوم که کاربران مجبورند تمام مدارک را بخوانند تا اطلاعات موردنظر خود را بیابند. کاربران اغلب به دنبال اطلاعات کاملاً مشخصی هستند، برای مثال در پاسخ به سؤال کاربر با عبارت «موزارت در هنگام مرگ چندساله بوده؟» ممکن است مدرک یا مدارکی بازیابی شوند که زندگی‌نامه موزارت را توصیف می‌کنند و حاوی مقادیر زیادی اطلاعات غیر مرتبط باشند که نیاز به بررسی داشته باشند. این کار می‌تواند به‌سادگی با شناسایی بخش‌هایی از مدرک که حاوی اطلاعات ساختاریافته است انجام شود، اما در مورد تفسیر بخش‌هایی از مدرک که حاوی اطلاعات بدون ساختار است، با چالش مواجه خواهیم شد.

به‌طور کلی دو رویکرد اصلی در بازیابی اطلاعات وجود دارد: رویکرد آماری و رویکرد معنایی؛ هدف رویکرد معنایی تحلیل معنایی و نحوی جهت درک نسبی از زبان

1. Voorhees, Ellen M.

2. Russel-Ross, Tony, Stevenson, Mark

طبیعی است و رویکرد آماری مدارکی را که بر اساس برخی قواعد آماری شباهت بیشتری با پرسش دارند را در رتبه بالاتری قرار می‌دهد. امروزه همچنین روش‌های آماری و معنایی در ترکیب باهم به کار می‌روند.

رویکردهای استاندارد آماری در بازیابی متن مانند مدل بولی، بولی گسترش‌یافته، فضا-برداری و مدل‌های احتمالاتی برای توصیف محتوای مدارک به عبارات نمایه‌ای متکی هستند. در این روش‌ها واژه‌ها پیش‌پردازش می‌شوند که این مرحله شامل ریشه‌یابی کلمات، حذف کلمات زائد مانند حروف اضافه است که به آن‌ها واژه‌های «بازدارنده»<sup>۱</sup> گفته می‌شود. عبارات نمایه‌ای از فهرست کلمات موجود در مدرک تهیه شده‌اند. این رویکرد گاهی با عنوان «کوله‌ای از کلمات»<sup>۲</sup> شناخته می‌شود. فنونی مانند حذف واژه‌های «بازدارنده» یا غیرمجاز و یا وزن دهی به عبارات معمولاً برای استفاده بهینه از عبارات نمایه‌ای به کار می‌روند (رابرتسون و اسپارک جونز،<sup>۳</sup> ۱۹۹۷). به‌هرحال استفاده از مدل «کوله‌ای از کلمات» برای پردازش مدارک و متون زبان طبیعی ایدئال نیست، این مدل در پردازش مدارک به زبان طبیعی چندین مورد را نمی‌تواند تحت پوشش قرار دهد:

۱. برخی واژه‌ها بیش از یک معنی دارند، بنابراین اگر اطلاعات اضافی‌ای در مورد آن‌ها نداشته باشیم، نمی‌توانیم اطمینان حاصل کنیم که مدرک بازیابی شده مرتبط با آن واژه است یا خیر. (برای مثال کلمه «شیر» هم نام یک حیوان است و هم به معنای نوعی لبنیات یا ماده خوراکی و نیز به معنای شیر آب است).

۲. مفاهیم یکسان را می‌توان را با جمله‌های متفاوت به شیوه‌های مختلف بیان کرد. زبان طبیعی این امکان را فراهم می‌سازد که عبارات مشابه به شیوه‌های مختلف با استفاده از مترادف‌ها، بازنویسی و استعاره‌ها ایجاد شوند. برای مثال این جمله را در نظر بگیرید «قهوه استارباکس بهترین است» و با جمله «هرزمان که احساس کنم نیاز به کافئین دارم به شرکتی که دارای شعبات مختلف در سیاتل است مراجعه می‌کنم» این دو جمله هیچ کلمه یا

---

1. Stop words  
 2. Bag of Words  
 3. Rabertson, S. and Spark, Jones, K.

عبارت مشابهی ندارند ولی تقریباً یک مفهوم را بیان می‌کنند.  
۳. موضوع یک مدرک به راحتی به وسیله اصطلاحات به کاررفته در آن تعیین نمی‌شود.

فنون پردازش زبان طبیعی باهدف تحلیل هوشمندانه مدارک و دریافت معنای آنها انجام می‌شود و همین امر موجب شده که برای حل معضلات بازیابی اطلاعات این فنون، مفید و مؤثر واقع شوند. این فنون به شرح زیر است:

شناسایی و تشخیص موجودیت‌ها: در این فرایند مفاهیم و موجودیت‌های کلیدی مانند نام افراد، مکان‌ها و سازمان‌ها در متن شناسایی می‌شود. مزیت تشخیص درست این موجودیت‌ها این است که امکان نمایه‌سازی دقیق‌تر مدرک را فراهم می‌آورند و نهایتاً موجب می‌شوند که جستجو کنندگان به مدرک مرتبط‌تری دست یابند و نیز با شناسایی این موجودیت‌ها می‌توان پیوندهایی به سایر مدارک مرتبط ایجاد کرد. در حال حاضر از این روش برای برجسته کردن نام برخی افراد و پیوند نام آنها به مدارک حاوی زندگی‌نامه‌شان استفاده می‌شود. برای به کارگیری چنین نظام‌هایی از تمام یا برخی از فنون زیر استفاده می‌کنند:

۱. مدل مارکوف<sup>۱</sup>،

۲. ماشین‌های برداری پشتیبان<sup>۲</sup>،

۳. قوانینی که به صورت دستی ایجاد شده‌اند. مطالعات فریبگی

دو روش اول مبتنی بر رویکرد یادگیری ماشینی هستند. برخی از این فنون تشخیص موجودیت، به ویژگی‌های درون‌متنی و برخی دیگر بر شواهد خارج از متن تمرکز دارند (راسل رز و استیونسون، ۲۰۰۹).

ابهام‌زدایی از مفهوم واژه‌ها<sup>۳</sup>: رفع ابهام از واژه، یافتن درست کلمات هم‌نویس در یک

- 
1. Markov model
  2. Support Vector Machine
  3. Word sense disambiguation

متن است، به‌عنوان مثال «مرد» در معنای اسمی به انسان ذکور بالغ یا در معنای فعلی بن ماضی سوم شخص از مصدر «مردن» اطلاق می‌شود؛ بنابراین یکی از وظایف رفع ابهام انتخاب یکی از این دو حالت برای کاربرد در جمله ما است. این واقعیت که واژه‌ها می‌توانند دارای چندمعنا و مفهوم باشند «چندمعنایی» نامیده می‌شود. این مسئله علاوه بر اینکه این مسئله بازیابی اطلاعات را با چالش مواجه می‌کند و به‌عنوان یک معضل برای پردازش زبان طبیعی نیز مطرح است. پردازش زبان طبیعی فوننی را برای شناسایی خودکار معانی واژه‌های موجود در متن با عنوان ابهام‌زدایی از معانی واژه‌ها در نظر گرفته است. یکی از علل ابهام مربوط به نحو واژه‌هاست، به‌عبارت‌دیگر اینکه واژه‌ها در چه جایگاه نحوی (اسم، صفت، فعل و...) قرار گیرند بر معنای آن تأثیرگذار است. پردازش زبان طبیعی برای شناسایی نقش دستوری واژه‌ها از برچسب‌گذاری واژه‌ها در متن استفاده می‌کند، برچسب‌های اختصاص‌یافته به هر یک از واژه‌ها نقش دستوری آن‌ها را به شکل اسم، فعل، صفت، قید و ... مشخص می‌کند. کروتز<sup>۱</sup> (۱۹۹۷) از برچسب‌گذاری واژه‌ها برای بازیابی بهتر استفاده کرد و نشان داد این کار تا چه اندازه بر بهبود بازیابی مؤثر است. منابع واژگانی ماشین‌خوان بزرگ مانند وردنت<sup>۲</sup> که در سال ۱۹۸۰ به وجود آمد، به‌سرعت جهت ابهام‌زدایی واژه‌ها به کار رفت. مزیت آن‌ها این است که هم فهرستی از معانی واژه‌ها و هم اطلاعاتی که برای تشخیص معنای درست هر واژه لازم است را فراهم می‌سازند (لسک، ۱۹۶۸). لسک<sup>۳</sup> (۱۹۶۸) تخمین زد که با این شیوه ۵۰-۷۰٪ واژه‌ها ابهام‌زدایی می‌شوند. این رویکرد مشابه مدل «کوله کلمات» در بازیابی اطلاعات فراگیر است که معنای یک مدرک را با واژگان به‌کاررفته در آن نمایان می‌سازد (راسل رز و استیونسون، ۲۰۰۹). رفع ابهام از واژگان در بسیاری موارد از جمله بازیابی اطلاعات به زبان طبیعی، نمایه‌سازی مدارک در حوزه بازیابی اطلاعات و نیز در طبقه‌بندی متون کاربرد دارد (جرارد، لوئیز و جرمن، ۲۰۰۰).

---

1. Krovetz, R.

2. WordNet

3. Lesk, Michael

متن کاوی<sup>۱</sup>: بیشتر افراد با اصطلاح داده کاوی آشنایی دارند که هدف آن شناسایی روابط جدید و جالب میان مفاهیم در یک پایگاه داده رابطه‌ای است. داده کاوی به عنوان شاخه‌ای از هوش مصنوعی در پی کشف اطلاعات و دانش از متون ساختاریافته است، اما در بسیاری محیط‌ها داده‌ها به جای جداول پایگاه داده‌ای ساختاریافته، به شکل متون زبان طبیعی و بدون ساختار یا نیمه ساختاریافته هستند، کشف دانش و اطلاعات مفید از چنین منابعی از طریق تشخیص و نمایش الگوهای معنادار، متن کاوی نامیده می‌شود. همچنین متن کاوی به عنوان ابزار تحلیل هوشمند متن از مجموعه داده‌های بزرگ نیز شناخته می‌شود.

امروزه راه کارهای بسیاری جهت سازمان‌دهی و مدیریت حجم بالای اسناد و مدارک و کشف اطلاعات مفید از داده‌ها در حال پیشرفت و تحقیق است. جستجوی مدرک مشابه به عنوان بخشی از متن کاوی در نظر گرفته شده و روش‌های هوش مصنوعی در این مراحل به طور گسترده مورد استفاده واقع شده، از این رو متن کاوی یکی از عملیات اساسی در حوزه مدیریت و سازمان‌دهی مدارک محسوب می‌شود (ساراجوگلو، توتونکو واله وردی<sup>۲</sup>، ۲۰۰۸). پردازش زبان طبیعی به عنوان مکمل داده کاوی در ترسیم بهتر الگوها و روش‌های متن کاوی نقش مهمی ایفا می‌کند. فرایند متن کاوی شامل گردآوری مدارک یا آماده‌سازی متن، پردازش متون، الگوسازی، تجزیه و تحلیل الگو به وسیله خوشه‌بندی و استخراج دانش است. در اکثر زمینه‌های متن کاوی نیاز به پیش پردازش متن به وسیله ابزارهای پردازش زبان طبیعی و تبدیل داده‌های متنی به بردارهای عددی است.

از آنجایی که بیشتر منابع در کتابخانه‌ها و مراکز اطلاع‌رسانی در قالب متنی نگهداری می‌شوند، متن کاوی ارزش بالقوه بالایی در ارائه خدمات کارآمدتر به کاربران این مراکز را دارد و از آنجایی که پردازش دستی متون و مدارک کاری طاقت فرساست، روش‌های خودکار پردازش جایگزین شیوه‌های سنتی شده است. یکی از کاربردهای متن کاوی در علم اطلاعات در حوزه مدیریت و سازمان‌دهی اطلاعات است، از آنجایی هدف نظام‌های سازمان‌دهی دانش نظم دهی به محتوای اسناد و مدارک و متون است، فنون متن کاوی مانند

1. Text mining

2. Saracoglu, Ridvan, Tutuncu, Kemal and Allahverdi, Novruz



استخراج خودکار اطلاعات، رهگیری عنوان، خلاصه‌سازی خودکار، طبقه‌بندی، خوشه‌بندی، پیوند مفهومی و مصورسازی اطلاعات در رسیدن به اهدافی همچون طبقه‌بندی موضوعی مقالات، کتاب‌ها و ...، مشابهت‌یابی بین مستندات مختلف، جستجو بین انبوه منابع یاری می‌رساند. متن کاوی به سرعت به یکی از حوزه‌های مهم و پرکاربرد در حوزه‌های تحت پوشش پردازش زبان طبیعی تبدیل شد. متن کاوی ارتباط مستقیمی با استخراج اطلاعات و کشف دانش به‌عنوان حوزه‌های فرعی و مرتبط با علم اطلاعات دارد.

متن کاوی بصری (مصورسازی اطلاعات): استفاده بشر از نمادهای تصویری جهت انتقال مفاهیم و اندیشه‌ها، تاریخی به قدمت تاریخ بشر دارد. تصویرنگاری به ماندگاری اندیشه بشری در قالب تصاویر کمک شایانی کرده است. انسان عصر حاضر نیز بر دریافت اطلاعات با استفاده از حس بصری برای دسترسی و درک اطلاعات بیش‌ازپیش تأکید می‌نماید.

در فرایند مصورسازی اطلاعات منابع عظیم متنی در سلسله‌مراتب تصویری یا نقشه قرار می‌گیرند و علاوه بر قابلیت جستجوی ساده، قابلیت مرور نیز فراهم می‌شود. مصورسازی زمانی مفید است که کاربر به محدود کردن طیف گسترده‌ای از مدارک و اسناد و کشف موضوعات مرتبط نیاز دارد (پرثی و حمیدی، ۱۳۹۶)؛ به عبارت دیگر گاهی نتایج خروجی متن کاوی پس از ارزیابی برای ارائه به مدیران بصری سازی می‌شوند. ظهور فنون جدید مصورسازی و نظام‌های مبتنی بر آن مانند ترسیم مقیاس جزء<sup>۱</sup>، کوگار<sup>۲</sup>، گوئیدو<sup>۳</sup>، وب وایب<sup>۴</sup>، اینفو کریستال<sup>۵</sup> و نظایر آن نویدبخش چیرگی بر مشکلات حوزه بازیابی اطلاعات است. در این شیوه‌ها با هدایت محتوا و ربط آن به سوی فضاهای چندبعدی، کاربر قادر به مرور مجموعه‌ای از مدارک و بازیابی منابع موردنیاز خود می‌شود

- 
1. Component scale drawing
  2. Cougar
  3. GUIDO
  4. WebVibe
  5. InfoCrystal

(مورس، لوئیس و اولسن<sup>۱</sup>، ۲۰۰۱). اینفوکرستال یکی از روش‌هایی است که روابط بین عنصرهای مختلف مورد جستجو در نظام‌های بازیابی اطلاعات را به ساده‌ترین شکل ممکن به تصویر می‌کشد (فرهادپور و مطلبی، ۱۳۹۰).

ارتباط فعلی میان زبان بصری و پردازش زبان طبیعی به ترجمه زبان‌های گرافیکی به زبان طبیعی یا ارائه تصویری زبان‌های پردازش متن محدود شده است، اما ما برای باوریم که توان بالقوه زیادی در اجرا و پیاده‌سازی تصویری نظام‌های پردازش زبان طبیعی وجود دارد. یک دلیل برای این مدعا ویژگی مدولار بودن الگوریتم‌های پردازش زبان طبیعی است، به این معنا که زبان بصری جریان داده یک روش طبیعی برای ارائه برنامه‌های پردازش زبان طبیعی است. همچنین نیاز مبرمی به ابزارهای تولیدی وجود دارد که امکان مصورسازی داده‌ها را به همراه مدارک متنی را پس از آنکه مورد تحلیل فنون پردازش زبان طبیعی قرار گرفت، فراهم آورد. هدف چنین ابزارهایی مصورسازی داده‌ها و کمک به محققان و توسعه‌دهندگان نظام‌های پردازش زبان طبیعی به وسیله فراهم ساختن امکانات جهت استفاده مجدد برنامه‌های پردازش زبان طبیعی، مدیریت مجموعه‌های بزرگ متنی و مصورسازی نتایج پردازش است (راجرز و همکاران، ۱۹۹۷).

### تحلیل احساس<sup>۲</sup> و واژه و عقیده‌کاوی<sup>۳</sup>

عقاید تقریباً در تمام فعالیت‌های انسانی بسیار مهم و از تاثیرگذارترین رفتارهای ما محسوب می‌شوند. برداشت‌های ما از واقعیت و انتخاب‌هایی که با توجه به آن‌ها انجام می‌دهیم تا حد قابل توجهی به نحوه مشاهده و ارزیابی دیگران از جهان پیرامون مان وابسته است، به عبارت دیگر اعتقادات و برداشت ما از واقعیت مشروط به این است که دیگران چگونه دنیا را می‌بینند و ارزیابی می‌کنند، بنابراین هنگام نیاز به تصمیم‌گیری، غالباً نظرات دیگران را جویا می‌شویم، این مسئله در مورد سازمان‌ها نیز صادق است (لئو<sup>۴</sup>، ۲۰۱۲) و کتابخانه‌ها

1. Morse, Emile, Lewis, Michael and Olsen, Kai A.

2. Sentiment analysis

3. Opinion minding

4. Liu, Bing

نیز به عنوان سازمان‌ها و مراکز اطلاع‌رسانی از این امر مستثنی نمی‌باشند. با افزایش حجم اطلاعات متنی تولیدشده توسط کاربران در وبسایت کتابخانه‌های دیجیتال و مراکز اطلاع‌رسانی، تجزیه و تحلیل احساسات در متون و تعیین قطبیت متون باهدف تشخیص خودکار نظر نویسنده متن، به یکی از موضوعات جذاب برای محققان حوزه داده‌کاوی و پردازش زبان طبیعی تبدیل شده است. عقیده کاوی یا نظر کاوی امروزه به عنوان یکی از کاربردهای پراهمیت پردازش زبان طبیعی مطرح است. حوزه نظر کاوی یا تحلیل نظرات، سعی در تشخیص خودکار مثبت یا منفی بودن نظرات و یا تشخیص احساس این نظرات را دارد تا بتواند خلاصه‌ای از نظرات واردشده را در مورد یک خدمتی که به کاربران ارائه شده را در اختیار قرار دهد تا بدین گونه زمینه را جهت تصمیم‌گیری درست ارائه‌دهندگان خدمات فراهم سازد (شاکری و اسمعیلی تفت، ۱۳۹۴).

تحلیل احساس واژه حوزه مطالعاتی است که احساسات و عواطف واژگان کاربران را ارزیابی می‌کند و یکی از گسترده‌ترین زمینه‌های تحقیقاتی پردازش زبان طبیعی در سال‌های اخیر است. در گذشته (تا قبل از سال ۲۰۰۰) به جهت آنکه پیش‌تر نظرات کاربران کمتر به شکل فرم‌های دیجیتالی ارائه می‌شد، موضوع تحلیل احساس واژگان نیز کمتر مورد توجه بوده است، اما در سال‌های اخیر تجزیه و تحلیل احساس واژگان مورد استفاده کاربران کتابخانه‌ها، به یکی از موضوعات چالش‌برانگیز و مورد علاقه محققان در حوزه پردازش زبان طبیعی تبدیل شده است (تصویر ۳). رشد سریع تحلیل احساسات با رسانه‌های اجتماعی هم‌زمان شده است و در مرکز پژوهش‌های حوزه رسانه‌های اجتماعی قرار گرفته است. به کارگیری روش‌های سنتی (تحلیل توسط انسان) در تحلیل احساس واژه کاربران، با مشکلات فراوانی روبه‌رو است، بنابراین در سال‌های اخیر خودکارسازی نظام‌های تحلیل احساس مورد توجه قرار گرفته است. تحلیل احساسات عمدتاً در سه سطح انجام می‌گیرد:

سطح مدرک: وظیفه این سطح طبقه‌بندی مدارک و اسناد بر اساس دارا بودن احساس مثبت یا منفی می‌باشد؛ به عبارت دیگر کل یک مدرک را در نظر گرفته می‌شود و احساس مثبت، منفی یا خنثی به آن نسبت داده می‌شود. برای مثال با بررسی نظرات در رابطه با یک

محصول یا خدمت، نظام تحلیل‌کننده تعیین می‌کند که در مجموع سطح احساس واژگان مثبت یا منفی است که با عنوان طبقه‌بندی احساسات سطح مدرک شناخته می‌شود. از آنجایی که در این سطح احساسات در مورد یک موجودیت واحد بررسی می‌شوند، لذا در مورد مدارکی که ماهیت چندگانه دارند قابل اجرا نیستند.

سطح جمله: وظیفه این سطح بررسی احساس جملات از نظر مثبت، منفی یا خنثی بودن است و رابطه نزدیکی با طبقه‌بندی ذهنی دارد.

سطح موجودیت و مؤلفه: دو سطح قبلی (مدرک و جمله) احساسات واژگان را به طور کامل پوشش نمی‌دهند، اما سطح جنبه و موجودیت که قبلاً سطح ویژگی‌ها نامیده می‌شد، تحلیل کامل‌تری ارائه می‌دهد. در این سطح ویژگی‌های خاصی از یک موجودیت مورد توجه قرار می‌گیرد؛ به عبارت دیگر در این سطح برچسب احساسی هر موجودیت تعیین می‌شود (لئو، ۲۰۱۲).

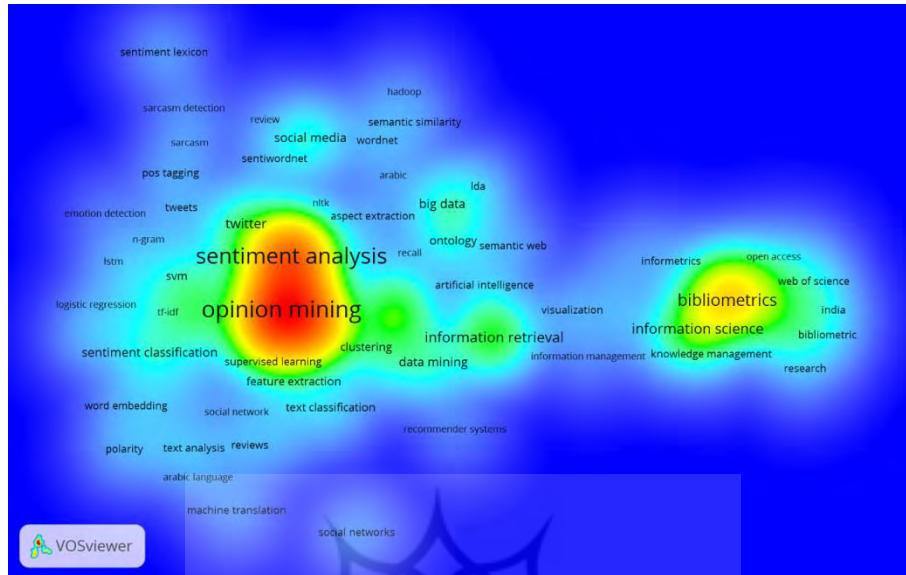
تحلیل احساس به شیوه‌های یادگیری ماشینی، مبتنی بر فرهنگ واژگان و شیوه ترکیبی طبقه‌بندی می‌شود؛ به عبارت دیگر تعیین قطبیت می‌تواند به وسیله روش‌های یادگیری ماشینی و یا روش‌های مبتنی بر فرهنگ واژگان انجام گیرد. روش‌های یادگیری ماشینی به سه شکل نظارتی، نیمه نظارتی و بدون نظارت انجام می‌گیرد و روش‌های مبتنی بر فرهنگ واژه‌ها، به دو شکل روش‌های واژه‌نامه‌ای و روش‌های مبتنی بر پیکره انجام می‌گیرند. واژه‌نامه‌ها هم با استفاده از هستان‌شناسی و هم بدون آن ایجاد می‌شوند (نوربهبانی، ۱۳۹۷).

در شیوه مبتنی بر یادگیری ماشینی از الگوریتم‌های یادگیری ماشینی مانند ماشین بردار پشتیبان و ویژگی‌های زبان‌شناختی استفاده می‌شود. شیوه فرهنگ واژگان مبتنی بر مجموعه‌ای از عبارات احساسی به همراه رتبه احساسی آن‌ها است و شیوه ترکیبی، ترکیبی از هر دو روش است. شیوه یادگیری ماشینی در حوزه پردازش زبان طبیعی به‌طور کلی و در حوزه تحلیل احساس به‌طور خاص از ویژگی‌هایی استفاده می‌کنند که شامل واژگان یا ترکیبی از واژگان پیکره متنی هستند و مدل‌های آماری را نیز به این پدیده‌های زبانی

نگاشت می‌دهند (بلیتزر<sup>۱</sup>، ۲۰۰۸). در شیوه‌های یادگیری ماشینی که به ایجاد مدلی برای نگاشت ویژگی‌ها به خروجی مطلوب مبادرت دارد، انتخاب ویژگی‌های مناسب در تولید مدل نقش مهمی دارد. یکی از چالش‌های این شیوه وجود ویژگی‌های بالاست که نه تنها باعث مشکلات محاسباتی و زمانی می‌شود، بلکه در صحت مدل نیز تأثیر نامطلوب دارد. بدین منظور از شیوه کاهش بعد برای استخراج و فشردگی ویژگی‌ها استفاده می‌کنند (برادران و گلپررابوکی، ۱۳۹۸). پردازش زبان طبیعی به شناخت احساس عبارات و جملات و در نتیجه در تحلیل بهتر نظرسنجی‌ها بسیار مفید واقع شده است. یکی از کاربردهای تحلیل احساسات در حوزه علم اطلاعات، در بخش فراهم آوری و خرید کتاب برای کتابخانه‌هاست. برای مثال در کتابخانه‌های دانشگاهی، شواهد نشان داده است که میزان مراجعه و استفاده از کتاب‌های کتابخانه‌های دانشگاهی نسبت به هزینه صورت گرفته برای خرید، به ویژه کتاب‌های لاتین بسیار پایین است، بنابراین به علت پایین بودن میزان مراجعه دانشجویان، اصلاح فرایند فراهم آوری و مجموعه‌سازی در کتابخانه‌ها ضروری می‌نماید. براین اساس تحلیل احساسات کاربران می‌تواند نظر و عقاید واقعی آن‌ها را استخراج کند و روند مجموعه‌سازی در کتابخانه‌ها را بهبود بخشد (عباسی و همکاران، ۱۳۹۶). همچنین یکی از راه‌های استخراج اطلاعات از داده‌های بدون ساختار متنی تحلیل احساسات است (ژان و فنگک<sup>۲</sup>، ۲۰۱۵).

پژوهشگاه علوم انسانی و مطالعات فرهنگی  
پرتال جامع علوم انسانی

- 
1. Blitzer, John.
  2. Fang, X. Zhan, J.



شکل ۳. تحلیل واژگانی پرکاربردترین و به‌روزترین کلیدواژه‌های مشترک حوزه پردازش زبان طبیعی و علم اطلاعات در مقالات سال‌های اخیر در پایگاه اسکوپوس

### استخراج خودکار اطلاعات<sup>۱</sup>

علی‌رغم تشابه دو عبارت بازیابی اطلاعات و استخراج اطلاعات، این دو عبارت متفاوت‌اند. یک نظام بازیابی اطلاعات مجموعه‌ای از مدارک که احتمالاً مورد نیاز کاربر بوده است را به او بازمی‌گرداند، اما یک نظام استخراج اطلاعات، متون را تجزیه و تحلیل کرده و تنها بخش‌هایی را که مورد نیاز کاربر است به او بازمی‌گرداند (طالبیان کوچکسرایبی، ۱۳۸۶).

نقطه شروع استخراج اطلاعات، تحلیل متون بدون ساختار است. نرم‌افزار استخراج اطلاعات عبارات کلیدی را استخراج و ارتباطات متون را تعیین هویت می‌کند (کاردان، کیهانی‌نژاد، ۱۳۹۱). استخراج اطلاعات یکی از فناوری‌های پردازش زبان طبیعی است که هدف آن استخراج اطلاعات کاملاً مشخص از مدارک است. زمانی که اطلاعات مرتبط

1. Automatic information extraction

شناسایی شد، در یک قالب کاملاً ساختاریافته به‌عنوان الگو ذخیره می‌شوند. کاربرد استخراج اطلاعات بر اساس شناسایی، برچسب زدن و استخراج عناصر کلیدی، مانند نام افراد، مؤسسات، مکان‌ها و کشورها از متونی با مقیاس بزرگ است (لیدی، ۲۰۱۸). استخراج اطلاعات ممکن است همراه با سایر کاربردهای پردازش زبان طبیعی مورد استفاده قرار گیرد، به این دلیل که داده‌های استخراج شده ممکن است بر اساس وظایف پیچیده پردازش زبان طبیعی باشد (بلیک<sup>۱</sup>، ۲۰۱۱). این روش به‌ویژه زمانی که با حجم بالایی از متون سروکار داریم، مفید است.

### پردازش زبان طبیعی در کتابخانه‌های دیجیتال

کتابخانه‌های دیجیتال از دیدگاه حوزه‌های مختلف مفهوم متفاوتی دارد برای مثال از دیدگاه علوم کامپیوتر، کتابخانه‌های دیجیتال گسترش فن‌آوری‌های شبکه، پایگاه داده‌ها و موتورهای جستجو هستند، اما از نظر علم اطلاعات، کتابخانه‌های دیجیتال موسسه هستند نه ماشین؛ به عبارت دیگر کتابخانه‌های دیجیتال توسعه منطقی کتابخانه‌های سنتی هستند که مأموریت آن‌ها دسترسی، سازمان‌دهی و اشاعه اطلاعات است. یا از دیدگاه گروهی دیگر کتابخانه دیجیتال به منزله مرکزی جهت استفاده ارائه‌دهندگان محتوا، ناشران، موزه‌ها و فروشندگان تجاری است؛ ابزاری جهت دموکراتیک سازی برای دولت‌ها؛ و یا مجرای برای ارائه خدمات جدید برای مریبان است؛ اما از نقطه نظر پردازش زبان طبیعی، کتابخانه دیجیتال ارائه فرصتی جدید به منظور حوزه‌های کاربردی است که در آن می‌توان ضمن به‌کارگیری فناوری‌های موجود، آن‌ها را کامل کرد و ابداعات بیشتری انجام داد. کتابخانه‌های دیجیتال چالش‌های منحصر به فردی جهت بازنمایی و بازیابی اطلاعات دارند، این امر موجب شده که به محیط ایدئالی برای کاربرد روش‌های پردازش زبان طبیعی به‌منظور پشتیبانی از بازیابی و کشف دانش تبدیل شوند. در ادامه به برخی از کارکردهای پردازش زبان طبیعی در کتابخانه‌های دیجیتال پرداخته‌ایم.

---

1. Black, Catherine.

### مدیریت مدارک و اسناد با استفاده از پردازش زبان طبیعی

تقریباً همه کاربردهای پردازش زبان طبیعی در محیط کتابخانه دیجیتال مرتبط و بالقوه مفید هستند. به‌طور خاص، روش‌های بازیابی اطلاعات جزئی جدایی‌ناپذیر از موتورهای جستجو هستند و به همین ترتیب تقریباً در هر کتابخانه دیجیتالی همراه با همه فن‌آوری‌های پشتیبانی‌کننده مانند ابهام‌زدایی از مفهوم و معنای کلمات و... نیز وجود دارند. در اینجا مشخصات اساسی کتابخانه‌های دیجیتال و چگونگی تأثیر آن‌ها بر روش‌های به‌کارگیری رویکردهای پردازش زبان طبیعی ذکر شده است:

- مجموعه‌های بزرگی از متون: پردازش مبتنی بر جمله اگر با انواع راه‌کارهای دیگر همراه نشود از کاربرد محدودی برخوردار است. پردازش معنایی واژگانی همانند پردازش متنی از اهمیت بالایی برخوردار است؛ تجزیه و تحلیل گفتمان نیز خصوصاً برای خلاصه‌سازی مفید است.

منابع دیجیتالی متشکل از متن، صدا، تصویر و منابع چندرسانه‌ای: منابع در قالب‌های مختلفی وجود دارند، اما فراداده‌های مرتبط با آن‌ها همگن و مبتنی بر متن هستند. انواع مختلف منابع: مدارک متنی موجود در کتابخانه‌های دیجیتال فقط محدود به شبکه‌های اطلاع‌رسانی و مقالات علمی که درساهای اخیر بیشتر مورد توجه پردازش زبان طبیعی بوده است، نمی‌شود. فناوری به‌کاررفته در کتابخانه دیجیتال باید متناسب با زمینه‌های مختلف، انواع مختلف مدارک و با ویژگی‌های ساختاری متفاوت سازگاری و تطابق داشته باشد.

وجود مجموعه‌های چندزبانه: منابع و فن‌آوری‌ها برای زبان‌های مختلف برای دسترسی به این موارد لازم است.

پیوندهای اساسی میان مدارک: مجموعه کتابخانه‌های دیجیتال معمولاً مواد و منابعی که به لحاظ موضوعی باهم مرتبط‌اند یا به طرق خاص دیگر به هم پیوند خورده‌اند را در خود جای می‌دهند این کار به‌عنوان یک امر پذیرفته‌شده درآمده و امکان طبقه‌بندی منابع را فراهم می‌کند. در مقالات تحقیقاتی منتشرشده این پیوندها، با ایجاد ارتباط میان



انواع مدارک امکان مطالعه مجموعه‌ای از مدارک و نیز امکان مطالعات کتاب‌سنجی و تحلیل‌های آماری یا مطالعه تحلیل‌های استنادی را فراهم می‌آورد.

اتکا بر فراداده برای توصیف منابع: گرایش به پذیرش طرح‌های استاندارد شده و رمزگذاری به این دلیل که فراداده‌ها با انواع مختلف فراداده‌های تعریف شده و ارائه شده توسط کاربر، همراه شوند. خلق فراداده‌ها زمانی که به عملیات طبقه‌بندی بیانجامد، بر پردازش زبان طبیعی متمرکز می‌شود.

یک وظیفه و یا مأموریت برای جامعه کاربران با نیازهای نسبتاً خاص، عادات و رفتار جستجو (برخلاف محیط ناهمگون وب): این جمله نیازهای یک کتابخانه دیجیتال را تعریف می‌کند. توجه داشته باشید که این مسئله بازتولید مجموعه‌های شخصی موسیقی یا تصاویر دیجیتال که جامعه هدف آن یک فرد است را مستثنی نمی‌کند؛ اما این مأموریت می‌تواند بر پردازش با محدود کردن آن تمرکز کند. یک تفاوت مهم بین وب به عنوان یک کل و کتابخانه دیجیتال، «جامعه» مخاطب آن است: کتابخانه دیجیتال برای جامعه‌ای طراحی شده که محتوای مجموعه (از طریق معیارهای جداکننده برای گنجاندن مدارک) و خدماتی که ارائه شده است را تشخیص می‌دهند. مفهوم اخیر یعنی ارائه خدمات در تعریف کتابخانه دیجیتال در جامعه علم اطلاعات، بسیار مهم است. یک کتابخانه دیجیتال نمی‌تواند صرفاً گردآورنده مدارک با حداقل کارایی و عملکرد باشد. این خصوصیات، به کارگیری پردازش زبان طبیعی را که مورد نیاز و مفید برای کتابخانه‌های دیجیتال است را ضروری می‌سازد.

مروری بر ابزارهای پردازش زبان طبیعی در مدیریت اسناد و مدارک  
این بخش طیفی از کاربردهای پردازش زبان طبیعی را در مدیریت مدارک در چهار جنبه ترسیم می‌کند:

فراهم آوری منابع (شامل ایجاد، ارائه و ذخیره)،

پردازش محتوا،

دسترسی کاربران به مدارک،

ابزارهای سازمان‌دهی دانش.

فراهم آوری منابع (شامل ایجاد، ارائه و ذخیره): این جنبه با دریافت و گردآوری منابع و سؤالات مربوط به ارائه مدارکی که به زبان حساس هستند، سروکار دارد. مجموعه کتابخانه هرگز نهایی نیست. به‌طور مداوم توسط مواد و منابع جدیدالورود افزایش می‌یابد. اینکه چه ماده‌ای اضافه می‌شود، بر اساس خط‌مشی‌های کتابخانه و بر اساس تعدادی از معیارها تعیین می‌شود. اگر مسائل اقتصادی را کنار بگذاریم، معیارها ممکن است شامل موارد زیر باشد: موضوع، نوع، مخاطب، نویسنده.

مدارک با شیوه‌های دانلود کردن، ایجاد، رقومی‌سازی و تبدیل از فرمت‌های دیگر به کتابخانه دیجیتال افزوده می‌شوند. در برخی موارد گردآوری و افزودن منابع جدید به کتابخانه می‌تواند با استفاده از ابزارهای پردازش زبان طبیعی به‌طور خودکار انجام شود. این امر به‌ویژه هنگامی صادق است که معیار انتخاب ما شامل موضوع باشد؛ وضعیتی که معیارهای انتخاب کتابخانه دیجیتال را به‌عنوان ویژگی مدارک بیان می‌کند؛ محتوای مدارک جدید را می‌توان با مشخصات مقایسه کرد و توسط الگوریتم طبقه‌بندی خودکار پردازش کرد. یک کار کاملاً مشابه نیز به نام «تصفیه و پالایش اطلاعات»<sup>۱</sup> انجام می‌شود که بین سیستم بازیابی خودکار و یک کار برقرار می‌گیرد و تعداد مدارک بازیابی شده را محدود می‌کند. علاوه بر آن، گاهی لازم است مدارک غیر متنی با استفاده از ابزارهای پردازش زبان طبیعی مانند تشخیص نویسه نوری، تشخیص و شناسایی دست خط برای آرشیوهای تاریخی، رونویسی از مواد صوتی، ترجمه ماشینی، استخراج متن از زبان نشانه‌گذاری فرا متن (اچ تی ام آل)<sup>۲</sup> یا قالب‌های پی‌دی‌اف<sup>۳</sup> و... به مدارک متنی تبدیل شوند؛ به این دلیل که این منابع توسط انواع تبدیل‌ها به‌دست آمده‌اند و بنابراین کاملاً مستعد خطا هستند، یک مرحله اضافی برای بررسی هجایی، دستور زبان و سبک اسناد می‌تواند انجام شود.

- 
1. Informationfilterng
  2. HTML(Hyper text markup language)
  3. PDF

### تعیین ابزارهای پردازش مناسب

ابزاری که برای پردازش مدارک استفاده می‌شود، به‌عنوان مثال استخراج‌کننده‌های عبارات و اصطلاحات، برچسب‌زن‌های نقش دستوری کلمات، خلاصه‌کننده‌ها و... که به نوع زبان حساس هستند: به‌عنوان مثال متون فارسی به ابزارهای متفاوتی از زبان چینی و یا آلمانی احتیاج دارند. در شناخت کتابخانه‌های دیجیتال امروزی منطقی است که بپذیریم کتابخانه‌های دیجیتال، چندزبانه در نظر گرفته شده‌اند. برای بهینه‌سازی عملکرد کلی سیستم مدیریت کتابخانه، مطلوب است که در عملکرد آن، شناسایی و رمزگذاری خودکار زبان نیز گنجانده شود، مانند سیستم‌هایی که در سال‌های اخیر بر اساس ویژگی‌ها و مشخصات آن-گرام<sup>۱</sup> گسترش یافته‌اند.

### شرح و توصیف مدرک

برای ارائه نمایش و ذخیره مدارک در کتابخانه دیجیتال، تهیه نوعی پیشینه که قابلیت دسترسی داشته باشند ضروری است. این مربوط به مدخل‌ها یا شناسه‌های کتابشناختی در کتابخانه‌های سنتی و یا یک پیشینه فراداده است (برای مثال فراداده توصیفی). این پیشینه که معمولاً به صورت صریح یا به صورت دستی رمزگذاری شده و تولید می‌شود و یا به‌طور خودکار با استخراج فراداده از منبع تولید می‌شود. در این حالت هیچ معناشناسی در فرایند مطرح نیست و از فناوری پردازش زبان طبیعی بسیار اندک استفاده می‌شود. با این حال، مستندسازی نام‌نویسندگان و عنوان منطقی است و به ابزارهای پردازش زبان طبیعی مشابه ابزارهای مستندسازی موجودیت‌های نام‌گذاری شده نیاز دارد. برای نمونه ابزارهای خودکار تشخیص مهر تاریخ برای مدارکی که فاقد مهر تاریخ هستند و یا شناسایی خودکار زبان مدارک و رمزگذاری یا قالب‌های تاریخ مدارک از این موارد هستند. فراداده‌های توصیفی یا فیزیکی که در بالا شرح داده‌ایم، اغلب برای بازیابی توسط کاربران کتابخانه کافی نبوده و ایدئال نمی‌باشند. فراداده‌های افزوده دیگری نیز می‌توانند به‌وسیله پردازش

1. N-gram

خودکار محتوا تولید شوند.

## پردازش محتوا

پردازش محتوا بخش عمده‌ای از تلاش مدیریت مدارک است که شامل تولید فراداده‌های توصیفی پیشرفته، به‌منظور سهولت در بازیابی مدارک توسط کاربران است، علاوه بر آن شامل قابلیت‌های بازیابی ارائه‌شده توسط جستجوی متن کامل است. فراداده‌های به‌دست‌آمده باید در سیستم سازمان‌دهی دانش کتابخانه دیجیتال گنجانده شوند. پردازش محتوا دلالت بر تجزیه و تحلیل زبانی یا محتوای مفهومی مدارک متنی دارد و نمایش خروجی مناسب برای این مدارک تولید می‌کند (مانند اصطلاحات نمایه شده، خلاصه‌ها، کدهای رده‌بندی و...); بنابراین پردازش محتوا وظایف سنتی رده‌بندی، نمایه‌سازی و خلاصه کردن مدارک را پوشش می‌دهد. رده‌بندی بر قرار دادن مدارک و منابع با موضوع‌های مشابه در کنار یکدیگر دلالت دارد و معمولاً از یک طرح رده‌بندی استفاده می‌کنند (مانند طرح رده‌بندی دیویی یا رده‌بندی دهدهی جهانی و...). همچنین در ادامه دنیای دیجیتال ارائه سلسله‌مراتبی از راهنماها را دربر خواهد داشت. نمایه‌سازی (که ممکن است توسط جوامع مختلف به اشکال گوناگون تفسیر شود) در اینجا نمایه‌سازی شامل توصیف مدارک توسط فهرست کوتاهی از عبارات یا ارائه اصطلاحات و کلیدواژه‌هایی است که نمایانگر موضوعات اصلی مورد بحث در یک مدرک هستند. خلاصه‌سازی شکل کوتاه شده مدرک است که معمولاً به سبک روایی می‌باشد.

این وظایف پردازش محتوا توسط سه فناوری اصلی پردازش زبان طبیعی انجام می‌شود. اولین مورد این سه‌گانه با رده‌بندی، طبقه‌بندی و خوشه‌بندی خودکار آغاز می‌شود. دومین وظیفه پردازش محتوا، نمایه‌سازی است که با روش‌های بازیابی و جستجو اجرا می‌شود. تمام اشکال بازیابی اطلاعات و موضوعات مربوط به حاشیه‌نویسی خودکار (به‌ویژه مترادف برای نمایه‌سازی خودکار) و استخراج فراداده به کتابخانه‌های دیجیتال بسیار مرتبط هستند. نمونه‌هایی از کاربرد نمایه‌سازی شامل این موارد است: افزودن

پردازش زبان طبیعی به یادگیری ماشین (ماشین بردار پشتیبان، الگوریتم جنگل تصادفی<sup>۱</sup>) به منظور بهبود استخراج عبارات کلیدی از مدارک علمی، تمایز قائل شدن برای مفهوم واژه در مجموعه مدارک و اسناد تاریخی به منظور بهبود شناخت و دسترسی به بایگانی‌های رقومی خاص، برای جستجو و بازیابی: به کارگیری فرهنگ لغت ترجمه برای امکان بازیابی اسناد تاریخی قدیمی (مثلاً این قابلیت که بتوان برای اسناد تاریخی به زبان‌های باستانی فارسی پرسش را با زبان فارسی مدرن مطرح کرد و پاسخ را دریافت کرد) و یا به کارگیری ترکیبی روش‌های اندازه‌گیری فراوانی وزنی عبارات و اصطلاحات و فراوانی معکوس عبارات و شبکه‌های اجتماعی (جامعه کاربران) به منظور بهبود بخشیدن الگوریتم‌های رتبه‌بندی بازیابی اطلاعات. سومین کار پردازش محتوا، خلاصه‌سازی خودکار است که سعی در بهبود و جایگزینی خلاصه‌سازی انسانی دارد، به این شکل که با تمرکز بر محتوا، جملات از مدرک ذکر شده بر اساس عناصر موجود در متن اسناد استخراج می‌شوند.

در چارچوب کتابخانه دیجیتال، پردازش محتوا (به‌طور خاص: نمایه‌سازی، خلاصه کردن و رده‌بندی مدارک) سیستم را قادر می‌سازد تا اطلاعات را به مدخل‌های اصلی کتابشناختی شامل فراداده مانند عنوان مدرک، نویسنده، تاریخ ایجاد، یو آر ال<sup>۲</sup>، قالب و ... اضافه کند. نتیجه پردازش محتوا افزودن کلیدواژه‌های نمایه‌سازی شده، کدهای رده‌بندی، به معنای نقاط دسترسی افزوده است که می‌بایست بازیابی را سهولت بخشند، یا خلاصه‌سازی که ربط مدرک با نیاز کاربر را تسهیل می‌سازد.

### نمایه‌سازی خودکار<sup>۳</sup> (ماشینی)

نمایه‌سازی یکی از شیوه‌های سازمان‌دهی اطلاعات جهت بازیابی کارآمدتر اطلاعات و از شاخه‌های فرعی علم اطلاعات است و کاربرد وسیعی در کتابخانه‌ها و مراکز اطلاع‌رسانی دارد. هدف نمایه‌سازی، راهنمایی کاربر به محتوا و مکان فیزیکی مدارک، به منظور تسهیل

---

1. Random forest algorithm  
 2. URL  
 3. Automatic indexing

بازیابی است؛ به عبارت دیگر نمایه‌سازی و بازیابی پیوندی دوسویه دارند. نمایه‌سازی فرایند تحلیل محتوای اطلاعاتی پیشینه‌ای از دانش و بیان کردن محتوای اطلاعاتی مذکور به زبان نمایه‌سازی از طریق اصطلاحات نمایه‌ای است که به صورت دستی یا ماشینی انجام می‌شود، به این مفهوم که اگر مراحل نمایه‌ای توسط انسان انجام شود، نمایه‌سازی دستی و چنانچه توسط رایانه انجام پذیرد، خودکار یا ماشینی نامیده می‌شود. اکثر روش‌های نمایه‌سازی خودکار اصطلاحات نمایه‌ای خود را از زبان طبیعی انتخاب می‌کنند؛ به این صورت که تک‌واژها و عبارات چند واژه‌ای، به طور مستقیم از عنوان، چکیده و متن یک مدرک استخراج می‌شوند (موئنز، ۲۰۰۳).

<sup>۱</sup> نمایه‌سازی ماشینی عمدتاً به دو شیوه کوئیک و کواک صورت می‌گیرد، در شیوه اول که موسوم به نمایه‌سازی کوئیک<sup>۲</sup> یا درون بافتی است، تطبیق واژه‌های عنوان مدرک با حذف واژگان غیرمجاز به وسیله فهرست واژگان غیرمجاز که به رایانه داده شده است، انجام می‌پذیرد و در مقابل شیوه کواک<sup>۳</sup> یا برون بافتی که برای حل مسائل ناشی از ضرورت کوتاه کردن عنوان و ساده‌سازی خوانایی مدخل‌ها به وجود آمد

تاریخچه نمایه‌سازی خودکار بر مبنای بسامد تکرار واژه به دهه ۱۹۵۰ و تحقیقات لون<sup>۴</sup> (۱۹۷۵) و باکسن دال<sup>۵</sup> (۱۹۵۸) بازمی‌گردد. در واقع اندیشه نمایه‌سازی ماشینی هم‌زمان با پیدایش مدل‌های برداری و احتمالاتی در دهه ۱۹۶۰ شکل گرفت. مدل‌های برداری که زیرمجموعه‌ای از فنون بازیابی اطلاعات هستند، بر این فرض استوارند که معنای یک مدرک از اصطلاحات مستخرج شده از آن به دست می‌آید و مدل‌های احتمالاتی نیز از نظریه احتمالات در بازیابی اطلاعات بهره می‌جویند.

مشکلات مربوط به کیفیت و هزینه نمایه‌سازی انسانی یا دستی، نمایه‌سازی خودکار را به یک هدف جذاب تبدیل کرده است. یکی دیگر از مشکلات نمایه‌سازی دستی عدم

- 
1. Moens, Marie Francine
  2. KWIC: Keyword in context
  3. KWOC: Keyword out of context
  4. Luhn, H.
  5. Baxendale, P.B.

یکدستی نتیجه نهایی حاصل از کار دو نمایه‌ساز یا دو گروه نمایه‌سازی است که در موضوع یکسانی دست به نمایه‌سازی زده‌اند. به گفته کلوردن تنها ۶۰٪ اصطلاح‌های به‌کاررفته در تزاروس یا اصطلاح‌نامه توسط دو فرد ممکن است یکسان باشد. همچنین اگر مدرک واحدی را به دو نمایه‌ساز باتجربه بسپاریم، در نمایه‌سازی حاصل‌شده تنها ۳۰٪ اصطلاحات نمایه‌ای به‌کاررفته توسط آن‌ها ممکن است مشترک باشد. به همین ترتیب اگر دو شخص واسط (شخصی که جستجوی در پایگاه داده را بجای کاربر انجام دهد مانند کتابدار) پرسش یکسانی را از یک پایگاه انجام دهند تنها ۴۰٪ نتیجه بازیابی ممکن است همسان باشد و نیز اگر از دو دانشمند یا مهندس خواسته شود که در مورد ربط نتایج بازیابی شده با پرسش قضاوت کنند، تنها ممکن است ۶۰٪ اتفاق نظر میان آن‌ها وجود داشته باشد. در ادامه کلوردن پیشنهاد کرد که راه‌حل بسیاری از این مشکلات با حذف اصطلاح‌نامه و جایگزین کردن عنوان یا چکیده مقاله به شکل متن آزاد است؛ رویکردی که موفقیت‌هایی به همراه داشت. به دلیل مشکلات مذکور تنها راه پیش روی محققان در نمایه‌سازی، به‌کارگیری فنون زبان‌شناسی محاسباتی و پردازش زبان طبیعی است. طبق گفته بورکو، در طول دهه گذشته پیشرفت زیادی حاصل شده است.

در طی دهه ۱۹۵۰ و اوایل ۱۹۶۰ نوعی خوش‌بینی وجود داشت که این هدف به‌سادگی محقق خواهد شد. گرچه، بسیاری از ایده‌های ارائه‌شده برای تجزیه و تحلیل متن، نمایه‌سازی و ترجمه ماشینی به اثبات رسیده است، اما پیچیدگی و غنای زبان طبیعی این امر را با مشکل مواجه کرد. نمایه‌سازی مستلزم تجزیه و تحلیل محتوای مدارک و نمایاندن چنین ویژگی‌هایی به‌وسیله برچسب‌های توصیفی است. کتاب‌های نمایه‌سازی شده متداول‌ترین نمونه از نظام‌های بازیابی اطلاعات را نشان می‌دهند. (کوریسینسکی و نیوول، ۱۹۹۰).  
نمایه‌سازی خودکار مراحل زیر را می‌طلبد:

۱. شناسایی واژه‌های انفرادی از متن و که این مرحله تحلیل واژگان نامیده می‌شود،
۲. انتخاب واژه‌های پرکاربرد و بسامد بالا از طریق حذف واژه‌های غیرمجاز (به

روش تطبیق با فهرست با واژگان غیرمجاز)،

۳. تبدیل واژه‌های باقی‌مانده به شکل ریشه آن‌ها (حذف پسوندها و پیشوندها)،

۴. محاسبه رایانه‌ای بسامد رخداد ریشه‌های تحلیل‌شده در متن به منظور تعیین تابع

ارزش‌گذاری هر ریشه،

۵. ریشه‌هایی که نسبت به برخی ارزش‌های قراردادی آستانه‌ای، ارزش بالاتری

دارند، برای متنی که در آن ظاهر شده‌اند، به عنوان کلیدواژه تعیین می‌شوند البته در برخی

نظام‌ها ممکن است کلیدواژه ارزش متناسب با تابع ارزش‌گذاری داشته باشد

(ویکری<sup>۱</sup>، ۲۰۰۵).

به‌هرحال تعیین واحدهای متنی و مشخص کردن حد و حدود واژه برای رایانه یا

ماشین از مسائل اساسی در گزینش اصطلاحات نمایه‌ای در نمایه‌سازی خودکار است

(گیلوری، ۱۳۷۹). به‌علاوه امکان تشخیص واژه‌های مفهومی از واژه‌های غیر مفهومی در

فرایند انتخاب اصطلاحات نمایه‌ای تأثیر بسیاری دارد. آنچه که مسلم است، ماشین امکان

تشخیص را تنها از طریق تطبیق واژه‌های استخراج‌شده از متن یا متناسب شده به متن با

فهرست واژه‌های غیرمجاز به دست می‌آورد، بنابراین در اختیار داشتن چنین فهرستی و

ارائه آن به برنامه‌های رایانه‌ای یکی از اقدامات سودمند در نمایه‌سازی خودکار است

(سنجی و داورپناه، ۱۳۸۸). منظور از فهرست واژگان غیرمجاز گروهی از کلمات (به نه

برای، با آن و...) مانند است که به فراوانی در متن وجود دارند، اما به‌تنهایی بار معنایی

ندارند، بلکه در ارتباط با سایر واژه‌ها مفهوم پیدا می‌کنند که این فرایند در مرحله تحلیل

متن اتفاق می‌افتد. همچنین این دسته از واژه‌ها که به‌عنوان واژگان غیرمجاز می‌شناسیم، در

پرسش کاربر و میزان ربط مدرک بازیابی شده نیز تأثیری ندارند و جداسازی آن‌ها از متن

موجب صرفه‌جویی در زمان و حجم بایگانی‌های نمایه و عدم بازیابی مدارک غیر مرتبط و

جلوگیری از ریزش کاذب خواهد شد (زو و همکاران<sup>۲</sup>، ۲۰۰۶).

نمایه‌سازی ماشینی مستلزم به‌کارگیری نرم‌افزارهایی جهت انجام نمایه‌سازی است،

---

1. Vickery, Alina

2. Zuo, Feng, Wang, Fu Lee, Deng, Xiaotie, Han, Song, and Wang, Lu Sheng



نرم افزارهای بسیاری جهت نمایه سازی ماشینی و با کمک رایانه ابداع شده است که در اینجا به تعدادی از آنها اشاره می کنیم:

نرم افزار سیندکس<sup>۱</sup>: جهت تهیه نمایه های خودکار برای کتاب، روزنامه و نشریات ادواری است. این نرم افزار با کاربردی بسیار آسان که در ساخت اصطلاح نامه ها و مستندات موضوعی نیز مورد استفاده واقع می شود، بنیانی برای نمایه سازی خودکار تلقی می شود.

ماک رکس<sup>۲</sup>: نرم افزاری جهت نمایه سازی ماشینی است که بسیاری از نرم افزارهای مشهور جهان از آن استفاده می کنند. این نرم افزار برای حجم بالای متون بسیار مناسب است. از ویژگی های مهم آن، ساخت مدخل ها، فهرست مستندات، بررسی دقیق مراحل، صرفه جویی در زمان، استفاده از حروف ویژه، بررسی مجدد نمایه ها و اصلاح آنهاست.

اسکای ایندکس<sup>۳</sup>: تعداد سطوح در سر عنوان موضوعی، برچسب های فصل و جلد، امکان تعریف پیشوند و پسوند برای تمامی برچسب ها از ویژگی های این نرم افزار است. دکستر<sup>۴</sup>: برای نمایه سازی در واژه پردازهایی مانند ورد طراحی شده و نیازی به ورود و ویرایش یا توجه به شناسه های افزوده توسط نمایه ساز نمی باشد. همچنین شناسه های نمایه را به صورت جدول ترتیبی و قابل ویرایش در اختیار نمایه ساز قرار می دهد.

ریتراپور<sup>۵</sup>: در اصل ابزار نمایه سازی تصاویر است و دارای محیطی ساده است و دارای قابلیت های بسیاری مانند تولید انواع نمایه های فرعی در قالب فصل ها و بخش های مختلف با توجه به نیاز کاربر و تولید و ذخیره خودکار کاربرگه های دسترسی سریع، مدیریت و سازمان دهی تصاویر، حذف تصاویر مشابه و... است.

همچنین نرم افزار هوم سایت<sup>۶</sup> و فرانت پیج<sup>۷</sup> که نمایه سازی را به شکل ساده و سطحی و سریع انجام می دهند و به دلیل مشکلاتی که دارند در مقایسه با نرم افزارهای پیشرفته،

- 
1. Cindex(<http://www.indexers.com>)
  2. Macrex(<http://www.macrex.com>)
  3. Sky index professional
  4. Dexter
  5. Retriever(<http://www.adjuge.net/retriever>)
  6. Homesite
  7. Front page

کتر مورد استفاده قرار می‌گیرند.

### نمایه‌سازی ماشینی در زبان فارسی

با گسترش مدارک الکترونیکی به زبان فارسی و به تبع آن کاربران فارسی‌زبان و ویژگی‌های خاص زبان و خط فارسی، نیاز به توجه و بهبود در روش‌های نمایه‌سازی را بیش‌ازپیش آشکار ساخته است، با این وجود به نظر می‌رسد تاکنون پژوهش‌های اندکی در زمینه نمایه‌سازی ماشینی برای زبان فارسی انجام گرفت است. یک نمایه‌ساز ماشینی قوی برای متون فارسی، می‌بایست از بلوک‌های ساختاری اصطلاح‌نامه‌های جامع دیجیتال، موتورهای استنتاج نحوی، پایگاه قواعد دستوری، الگوریتم‌های وزن دهی گوناگون و مانند آن تشکیل شده باشد که متناسب با زبان فارسی باشد. برای ساخت نمایه‌ساز ماشینی برای زبان فارسی، به عنوان یک نرم‌افزار پیچیده باید یک مدل مفهومی جامع با روش‌شناسی درخور طراحی شود و جهت پیاده‌سازی آن بستری مناسب فراهم شود (جلالی منش، علیدوستی و خسروجردی، ۱۳۹۱).

نیاکان (۱۳۸۳) از میان ویژگی‌های زبان فارسی، به عدم وجود حرف تعریف، عامل جنسیت و جمع شکسته به عنوان عواملی که به سود نمایه‌سازی است اشاره می‌کند. همچنین وجود کسره اضافه در زبان فارسی را به عنوان مزیت نام‌برده که شکل اضافه را بسیار آسان ساخته است و نیز بیان داشته که در زبان فارسی صفت و موصوف و عدد و معدود باهم مطابقت ندارند و یک صفت می‌تواند به چندین موصوف بازگردد.

دشواری‌های زبان فارسی در نمایه‌سازی بیشتر معطوف به دستور زبان آن است که در سه بخش قرار می‌گیرد:

۱. میان گفتار و نوشتار تفاوت‌های بسیاری وجود دارد؛ به این مفهوم که افراد در سخن گفتن الگوهایی را به کار می‌برند که در نوشتن از آن‌ها استفاده نمی‌کنند.
۲. اضافه‌ها که عبارتند از نسبت میان دو واژه یا گروهی از واژه‌ها باهم که گونه‌ای از پیوند لفظی یا معنوی را مشخص می‌کنند و دارای سه نوع ملکی، تخصیصی و بیانی هستند.
۳. چگونگی نگارش: برای مثال، جدا یا پیوسته‌نویسی مانند «یک‌طرفه» یا

«یک طرفه»، واژه‌های ترکیبی مانند «منابع آب» یا «مکانیک خاک»، شیوه دو گونه املای برخی واژه‌ها مانند «جست‌وجو» و «جستجو»، جایگاه الفبایی واژه‌ها و ناتوانی در نشان دادن صداها.

داوودآبادی (۱۳۸۳) نیز به تعدادی از مشکلات و پیچیدگی‌های درک زبان فارسی اشاره داشته است که بر نمایه‌سازی ماشینی تاثیر گذار است، از جمله اینکه زبان فارسی از ساختار ریخت‌شناسی پیچیده‌ای برخوردار است و یک شکل ظاهری می‌تواند واحدهای معنایی متفاوتی را نشان دهد. علاوه بر آن، بی ترتیب بودن زبان، دشواری در تعیین حدود عبارات به‌ویژه گروه‌های اسمی، چندمعنایی‌ها و چند نقشی بودن واژه‌ها مانند «شیر» و حذف واژه‌ها و عبارات به قرینه لفظی یا معنوی از جمله این مشکلات به شمار می‌رود. فرایندهای نمایه‌سازی ماشینی و بسیاری از زیرسیستم‌ها و الگوریتم‌های آن، از یک الگوی کم‌وبیش ثابت پیروی می‌کنند و تاثیر ویژگی‌های زبانی بیشتر بر جزئیات یا داده‌های ورودی به این الگوریتم‌ها، توالی فرایندها یا طراحی قواعد معطوف است. در بحث انتخاب میان نمایه‌سازی ماشینی یا نمایه‌سازی انسانی به علت رشد فزاینده انتشارات الکترونیکی گرایش به نمایه‌سازی خودکار و به‌وسیله نرم‌افزارهایی که هر کدام با روشی خاص و با رعایت جامعیت و مانعیت مطلوب به نمایه‌سازی انواع قالب‌های منابع اعم از متن، صوت و تصویر افزایش یافته است. البته باید توجه داشته باشیم آنچه که مدنظر و هدف غایی است، تولید نمایه‌ای با کیفیت مطلوب است و به دلایل اقتصادی و هزینه‌بر بودن نمایه‌سازی دستی، سرمایه‌گذاری به سمت نمایه‌سازی دستی محل تردید واقع شده است و گرایش مطالعات و تحقیقات به سمت وسوی طراحی نرم‌افزارهای توانمند گواهی بر این مدعا است.

### خلاصه‌سازی خودکار اطلاعات<sup>۱</sup>

از آنجایی که دسترسی به اطلاعات مفید، مرتبط و خلاصه یکی از نیازهای کاربران کتابخانه‌ها و مراکز اطلاع‌رسانی به‌ویژه در کتابخانه‌های دیجیتال در قالب منابع الکترونیکی

است، بنابراین وجود ابزارهایی که قادر باشند در کمترین زمان ممکن اطلاعات را به صورت خلاصه از متون و منابع استخراج کنند و در اختیار کاربران قرار دهند، بسیار ضروری و مؤثر است. خلاصه‌سازی خودکار به مفهوم استفاده از ابزارهای ماشینی و مبتنی بر رایانه جهت ایجاد خلاصه‌ای مفید و معتبر و از آن جهت که کیفیت خلاصه‌های تولیدشده هنوز به اندازه خلاصه‌های انسانی نیست، یکی از مسائل چالش برانگیز در پردازش زبان طبیعی محسوب می‌شود (ننکو و آومک کوئن، ۲۰۱۲). کارکرد نظام‌های خلاصه‌مبتهی بر استفاده از برخی زبان‌ها یا روش‌های آماری برای انتخاب مهم‌ترین کلمات یا عبارات از جمله‌ها یا پاراگراف‌ها در متون بزرگ و ایجاد خلاصه‌ای معنی‌دار جهت بازنمایی و ارائه متن است (چودوری، ۲۰۰۳).

ابتدایی‌ترین خلاصه‌سازی‌ها توسط لون (۱۹۵۷) و دایننگ (۱۹۹۳) با استفاده از معیار فراوانی برای واژه‌های کلیدی و حذف واژه‌های غیرمجاز انجام پذیرفته است. خلاصه‌سازها را می‌توان از جنبه‌های متفاوتی (مانند تعداد مدارک یا اسناد ورودی، هدف، روش تهیه و محتوا) تقسیم‌بندی نمود (ننکو و آومک کوئن، ۲۰۱۲)، نظام‌های خلاصه‌ساز از منظر منابع به خلاصه‌سازهای تک‌سندی و خلاصه‌سازهای چندسندی، و از منظر نوع خلاصه به خلاصه‌سازهای استخراجی و چکیده‌ای تقسیم‌بندی می‌شوند. در خلاصه‌سازی استخراجی، جملات مهم متن، عیناً در خلاصه ذکر می‌شود، این روش در مقایسه با خلاصه‌سازی چکیده‌ای از پیچیدگی کمتری برخوردار است، چراکه خلاصه‌سازی چکیده‌ای با تغییر در ساختار جملات، تلاش می‌کند که جملات جدیدی ایجاد کند (هووی، ۲۰۰۳). روش‌های خلاصه‌سازی خودکار به ساختار زبانی متون نگارش شده وابسته هستند و روش‌های پردازش زبان طبیعی نقش اساسی در ارائه خلاصه خودکار ایفا می‌کنند. گرچه در روش‌های خلاصه‌سازی استخراجی می‌توان از فنون موجود برای یک‌زبان، به اندکی تغییر در زبان دیگر نیز استفاده کرد، اما در شیوه چکیده‌ای لازم است به طور کامل از اصول نگارش آن زبان تبعیت شود. بدین منظور تاکنون ابزارها و فنون متعددی برای تولید خلاصه در زبان انگلیسی برای به وجود آمده است و پژوهش‌های قابل ملاحظه‌ای در این

زمینه انجام پذیرفته است که عمدتاً در زمینه خلاصه‌سازی استخراجی می‌باشند. یکی از انواع خلاصه‌سازی‌های خودکار، خلاصه‌سازی مبتنی بر عبارت پرس و جوی ارائه شده توسط کاربر است. در این روش برخلاف نظام‌های موجود برخلاف خلاصه‌سازی‌های عمومی، خلاصه‌ای با توجه به نیاز کاربر تولید می‌کنند. در این روش پس از طی مراحل پیش‌پردازش و جداسازی جملات و واژه‌های مربوط به آن‌ها، ابهام‌زدایی معنایی واژگان انجام می‌گیرد و سپس به محاسبه شباهت بین عبارت پرسش و جملات موجود در متن پرداخته می‌شود و در نهایت خلاصه‌سازی انجام می‌گیرد (سپهریان، سدید پور و شیرازی، ۱۳۹۳).

در زبان فارسی نیز نظام‌های خلاصه‌ساز استخراجی متنوعی به وجود آمده که نظام «فارسی فام» جزء نخستین آن‌هاست. در هر روش خلاصه‌سازی استخراجی سه مرحله اصلی پیش‌پردازش، پردازش و انتخاب جملات وجود دارد. یکی از مهم‌ترین چالش‌ها برای چنین نظام‌هایی، مرحله پیش‌پردازش زبان طبیعی نظیر ریشه‌یابی، حذف کلمات غیرمجاز، برچسب‌زنی نقش واژه‌ها و تعیین واژه‌های کلیدی است که پس از این مراحل به هر جمله در متن امتیازی تعلق می‌گیرد و در نهایت جمله با امتیاز بالاتر انتخاب می‌شود (ننکو و آومک کوئن<sup>۱</sup>، ۲۰۱۲).

ویژگی‌های نگارشی و رسم‌الخط زبان فارسی موجب چالش‌هایی در مرحله پیش‌پردازش شده است. از جمله این چالش‌ها می‌توان موارد زیر را نام برد: صورت‌های مختلف نوشتاری برخی حروف، عدم رعایت فاصله‌گذاری‌ها، تنوع نوشتاری برخی کلمات مانند اتاق و اطاق، صورت‌های مختلف نوشتاری برای پیشوندهای افعالی و اسمی مانند آنان و آن‌ها و نیز چالش‌های معنایی مانند ابهام معنایی در برخی کلمات مانند «شیر» که به سه مفهوم کاملاً متفاوت «شیر خوراکی»، «شیر جنگل» و «شیر آب» به کار می‌رود. برخی دیگر از این ابهامات معنایی به دلیل فقدان اطلاعات آوایی است، مانند «مرد» و «مُرد» که هر دو به یک شکل «مرد» نوشته می‌شوند. همچنین در زبان

1. Nenkova, Ani and Mackeown, Kathleen

فارسی به علت آنکه حروف در کلمات به هم چسبیده می‌باشند، موجب پیچیدگی‌هایی در پردازش زبان فارسی شده است و نیز ویژگی‌های دستور زبان فارسی مانند ابهام در قواعد تولید بن مضارع مانند «گفت» از بن «گو»، وجود افعال مرکب مانند «برداشتن» و متصل شدن برخی افعال با اسم مانند «بیدارند» به جای «بیدار هستند» (حسینی خواه احمدی و محبی، ۱۳۹۶).

### طبقه‌بندی خودکار متون<sup>۱</sup>

طبقه‌بندی متون در تقاطع حوزه یادگیری ماشینی و بازیابی اطلاعات مطرح است (نایت، ۱۹۹۹)، و به‌طور عمیقی بر اصول حاکم بر بازیابی اطلاعات متکی است (تساتارونیس، وارلامیس و وازیرجیانیس<sup>۲</sup>، ۲۰۱۰)، همچنین برخی مشخصات آن با استخراج داده از متون و داده کاوی مشترک است. باین‌حال مرز دقیق آن کماکان مورد بحث است. فناوری‌های اطلاعاتی مدرن و خدمات مبتنی بر وب با چالش انتخاب، پالایش و مدیریت مقادیر رو به رشد اطلاعات متنی که دسترسی به آن‌ها حیاتی است، مواجه هستند. طبقه‌بندی متون فرایندی است که در آن متن‌ها را به یک یا چند طبقه از قبل تعریف شده بر اساس محتوا یا زبان نگارش متن نسبت می‌دهیم (برگر و مرکل<sup>۳</sup>، ۲۰۰۴). طبقه‌بندی متون از زیرشاخه‌های بازیابی اطلاعات است که به کاربران امکان می‌دهد، مجموعه متون مورد علاقه خود را با پیمایش در سلسله مراتب طبقه‌بندی، با سهولت بیشتری مرور کنند، این الگو نه تنها برای بازیابی و پالایش اطلاعات، بلکه برای توسعه خدمات کاربر مدار برخط نیز مؤثر است. استخراج خودکار اطلاعات، نظام‌های پرسش و پاسخ و خلاصه‌سازی خودکار اطلاعات نیز که از شاخه‌های فرعی بازیابی اطلاعات و از کاربردهای مهم پردازش زبان طبیعی هستند، از مزیت‌های طبقه‌بندی متون در جهت کمک به انتخاب حوزه دانشی که معمولاً برنامه‌های کاربردی زبانی استفاده می‌کنند، بهره می‌جویند. فنون

1. Automatic text categorization

2. Tsatsaronis, George, Varlamis, Iraklis and Vazirgiannis, Michalis

3. Berger, Helmut and Merkl, Dieter

یادگیری ماشینی معمولاً در طبقه‌بندی خودکار متون به کار می‌روند. طبقه‌بندی خودکار متون به‌ویژه زمانی که پیاده‌سازی الگوریتم به شکل دقیق و صحیح انجام پذیرفته باشد، کاربرد وسیعی در طراحی نظام‌های بازیابی اطلاعات دارد. طبقه‌بندی کننده‌های متون، متن را به دسته‌هایی طبقه‌بندی می‌کنند که نشانگر موضوعات مهم مدرک است. در دسترس قرار دادن اطلاعات مربوط به طبقه‌ها، استفاده از حوزه‌های خاص فنون پردازش زبان طبیعی را امکان‌پذیر می‌سازد. به‌عنوان مثال، طبقه‌بندی کننده‌های متون، متن را به قسمت‌های کوچک‌تر مانند بخش یا پاراگراف تقسیم می‌کنند. این دانش می‌تواند به‌وسیله نظام‌های بازیابی اطلاعات، نظام‌های پرسش و پاسخ و خلاصه‌سازی خودکار متون، به ترتیب دریافتن حقایق مرتبط، انتخاب پاسخ مناسب و انتخاب بخش‌های مربوط به دامنه هدف یاری رساند. طبقه‌بندی متون به‌عنوان حوزه تحقیقاتی فعالی در بازیابی اطلاعات و یادگیری ماشینی از طیف گسترده‌ای از الگوریتم‌های یادگیری نظارت‌شده استفاده می‌کند (موسچیتی<sup>۱</sup>، ۲۰۰۳).

در طبقه‌بندی متون مراحل آماده‌سازی متون به‌منظور استخراج ویژگی‌ها از متن (حذف تگ‌های اچ تی ام ال<sup>۲</sup> یا ایکس ام ال<sup>۳</sup>، کدگذاری متون، حذف واژه‌های غیرمجاز، به دست آوردن ریشه کلمات و...)، شاخص‌گذاری مستندات، و زدن‌دهی به خصوصیات استخراجی، یادگیری (تعریف معیار شباهت، الگوریتم یادگیری و ارزیابی کارایی) پیاده‌سازی می‌شود. واژه‌هایی که وزن بیشتری دریافت کردند دارای اهمیت بیشتری در متن هستند و به‌عنوان کلیدواژه نمایه‌ای انتخاب می‌شوند و سپس متون در دسته‌های از پیش تعیین شده قرار می‌گیرند. گرچه طبقه‌بندی پست الکترونیک و تشخیص نامه‌های الکترونیکی بی‌ارزش، طبقه‌بندی رخدادهای خبری، تشخیص موضوع داده‌ها، پالایش متن، طبقه‌بندی سلسله مراتبی صفحات وب از جمله موارد کاربرد طبقه‌بندی متون است، اما کاربرد مدنظر ما در

---

1. Moschitti, Alessandro

2. HTML

3. XML

این مطالعه، استفاده از شاخص‌های کلیدی در بازیابی اطلاعات است. از موارد دیگر کاربردهای طبقه‌بندی متون، می‌توان سامانه‌های خودکار پاسخ به سؤالات، طبقه‌بندی گفتاری که ترکیبی از طبقه‌بندی متون و تشخیص گفتار است، طبقه‌بندی متون چندرسانه‌ای از طریق عنوان‌های متنی، تشخیص نویسنده برای متون ادبیاتی نامشخص یا مورد بحث، تشخیص خودکار سبک ادبی متن و رتبه‌بندی خودکار کیفیت متن نام برد (رضایی و همکاران، ۱۳۹۶). برخی از روش‌ها و رویکردهایی که بر اساس خواص آماری متون الگوریتم‌های یادگیری ماشینی هستند، عبارتند از: ماشین‌های بردار پشتیبان (از اصل به حداقل رساندن ریسک ساختاری در اختصاص یا عدم اختصاص یک طبقه به مدرک به کار می‌رود)، نزدیک‌ترین همسایه<sup>۱</sup>، درخت تصمیم‌گیری<sup>۲</sup>، طبقه‌بندی بیزی<sup>۳</sup>، تشابه بر اساس بازخورد، شبکه‌های عصبی و غیره.

### سامانه‌های طبقه‌بندی متون

طبقه‌بندی متون یا عمل برچسب‌گذاری موضوعی متون زبان طبیعی یکی از کاربردهای پردازش زبان طبیعی است که در بسیاری زمینه‌ها از جمله نمایه‌سازی متون بر اساس یک واژه‌نامه کنترل‌شده، پالایش متون، تولید خودکار فراداده، ابهام‌زدایی از واژه، تولید فهرست‌های سلسله‌مراتبی از منابع وبی و به‌طور کلی هر حوزه‌ای که نیاز به ساماندهی مدارک و یا توزیع انتخابی و تطبیقی خاصی از مدارک مدنظر باشد، کاربرد دارد. مزیت این شیوه کاهش هزینه‌های ایجاد شده ناشی از دخالت عامل انسانی یا فرد خبره است، چراکه طبقه‌بندی دستی متون علاوه بر زمان بر بودن، هزینه‌های زیادی به همراه دارد. همچنین علاوه بر موارد فوق طبقه‌بندی به شیوه دستی دارای معایب دیگری نیز می‌باشد، از جمله آنکه برای زمینه‌های تخصصی نیازمند دانش افراد متخصص و خبره است و از آنجایی که برچسب‌گذاری مبتنی بر دانش انسانی است، عاری از خطا نمی‌باشد، همچنین

- 
1. K Nearest neighbor algorithm
  2. Decision tree algorithm
  3. Bayes classification



دیدگاه و تصمیم دو فرد خبره می‌تواند متفاوت باشد.

### طبقه‌بندی متون فارسی

از آنجایی تعداد مدارک الکترونیکی فارسی به‌ویژه در کتابخانه‌های دیجیتال رو به فزونی است، به‌کارگیری روش‌های کارآمد جهت طبقه‌بندی متون فارسی بسیار توجه واقع شده است. کلمات کلیدی مجموعه‌ای از واژگان مهم در یک مدرک هستند که توصیفی از محتوای مدرک را ارائه می‌دهند و جهت اهداف مختلفی قابل استفاده هستند. استخراج کلیدواژه‌های نمایه‌ای از مدارک، یکی از فرایندهای مهم در طبقه‌بندی متون و استخراج اطلاعات محسوب می‌شود. این واژه‌های کلیدی و شاخص‌ها نکات اصلی متن را توصیف می‌کنند، لذا به‌منظور طبقه‌بندی متون مورداستفاده قرار می‌گیرند. استخراج کلمات کلیدی به‌طور دستی فرایندی بسیار دشوار و زمان‌بر است، لذا نیاز به ابزارهایی جهت خودکارسازی فرایند است. طبقه‌بندی متون بر اساس کلیدواژه‌های نمایه‌ای یک مسئله بسیار مهم در پردازش زبان فارسی است. بیشتر نظام‌های طبقه‌بندی خودکار برای زبان انگلیسی طراحی شده اند و معمولاً قابل استفاده برای متون فارسی نیستند. در زبان فارسی کلمات، صورت‌های نگارشی پیچیده‌ای دارند و پوشش کلیه حالات دستوری واژه‌ها با به‌کارگیری یک سری قواعد مشخص، ناممکن است، همچنین عدم وجود رسم‌الخط یکسان (برای مثال کتاب‌ها و کتابها) و وجود کدهای متفاوت برای حروف فارسی، تحلیل واژگانی را با چالش روبه‌رو ساخته است، به همین دلیل استخراج کلیدواژه‌های نمایه‌ای و طبقه‌بندی خودکار متون در زبان فارسی دشوار و پیچیده است. بدون استخراج نمایه، بسیاری از کاربردهای بازیابی اطلاعات مانند جستجوی و طبقه‌بندی متون، پالایش اطلاعات و خلاصه‌سازی متون به نتایج مطلوبی نمی‌تواند دست پیدا کنند. توسعه نظام طبقه‌بندی خودکار متون فارسی به دلیل ماهیت زبان فارسی و در دسترس نبودن مجموعه‌ای شامل ریشه واژه‌ها و واژه‌های پرکاربرد زبان و مجموعه‌ای برای آزمون نظام عملکردی نسبت دشوار است. در برخی پژوهش‌های انجام‌گرفته در زبان فارسی، با استفاده از مدل‌های بازیابی اطلاعات مانند مدل فضا برداری آن-گرام و روش‌های وزن دهی

اجرا شده است و یا نوعی ریشه‌یاب فارسی شرح داده شده است و نیز از شیوه‌های شاخص‌گذاری متون ۳-گرام و ۴-گرام و الگوریتم یادگیری ماشینی موسوم به نزدیک‌ترین همسایه به کاررفته است. بینا، رهگذر و دهموبد (۱۳۸۶) روش ۴-گرام را بهترین روش شاخص‌گذاری متون با معیار تشابه ضرب داخلی معرفی کرده‌اند. مزدک و هسل<sup>۱</sup> (۲۰۰۴) سامانه‌ای به نام فارسی سام برای زبان فارسی ارائه دادند که بر مبنای یک سیستم خلاصه‌سازی متون سوئدی به نام سوئی سام است.

### اصطلاح‌نامه<sup>۲</sup>

توجه داشته باشید که اصطلاح «اصطلاح‌نامه» از دیدگاه متخصصین علم اطلاعات یا کتابداران با دیدگاه متخصصان کامپیوتر، یا زبان‌شناسان متفاوت است. از دیدگاه عموم، اصطلاح‌نامه نوعی فرهنگ لغات مترادف است؛ اما در واقع اصطلاح‌نامه چیزی فراتر از این است. اصطلاح‌نامه نه تنها اصطلاحات مترادف را، بلکه روابط سلسله‌مراتبی و سایر ارتباطات معنایی مرتبط با منابع را نیز رمزگذاری می‌کند (به این معنا که اصطلاحات اعم و اخص را مشخص می‌کند). اصطلاح‌نامه مجموعه‌ای از واژه‌ها و اصطلاحات مربوط به یک حوزه موضوعی خاص از دانش بشری است که شامل واژگان زبان نمایه‌ای کنترل شده است و به گونه‌ای سازمان‌یافته تا روابط پیشین مفاهیم را روشن سازد واحد تشکیل دهنده اصطلاح‌نامه واژه‌هایی است که نمایانگر اطلاعات آن مدرک یا متن است و «کلیدواژه» نامیده می‌شوند. یکی از کارکردهای اصطلاح‌نامه نشان دادن روابط واژه‌ها با یکدیگر و روابط مفاهیمی است که این واژه‌ها بر آن‌ها دلالت دارند. بر این اساس سه نوع رابطه هم‌ارز (رابطه تعادل و ترادف)، رابطه سلسله‌مراتبی (رابطه کل و جزء) و رابطه هم‌بسته (مرتبط) در هر اصطلاح‌نامه وجود دارد (حری، ۱۳۸۳).

اصطلاح‌نامه ابزار کلیدی در فرایند سازمان‌دهی و بازیابی اطلاعات است، از این نقش مهمی در نظام‌های ذخیره و بازیابی اطلاعات ایفا می‌کند. کاربرد اولیه این ابزار در

1. Hassel, Martin and Mazdak, Nima

2. Thesaurus

سازمان‌دهی مدارک اطلاعاتی و به‌منظور بازنمودن نظام‌یافته محتوای مدارک اطلاعاتی بوده است، اما با توسعه نظام‌های اطلاعاتی، نقش پررنگ‌تری در بازیابی اطلاعات یافته است (شیری و روی ۲۰۰۵). باوجود توانمندی‌های ذاتی اصطلاح‌نامه‌ها، بررسی تاریخی و کارکردی آن‌ها در طول سال‌ها نشان می‌دهد که باوجود تغییر در محیط‌های اطلاعاتی، اهداف و ساختار آن‌ها همواره ثابت بوده است (صنعت جو، ۱۳۸۴). به‌گونه‌ای که هدف اصلی آن بازیابی اطلاعات و سایر اهداف آن شامل کمک درزمینه درکلی یک حوزه موضوعی خاص، فراهم ساختن نقشه معناشناختی از طریق نمایاندن روابط مفهومی و کمک برای ارائه تعاریف اصطلاحات است (ایچیسن، گیلکریست و بادن، ۲۰۰۰). به‌این ترتیب اصطلاح‌نامه با ارائه مفاهیم به شیوه‌های مختلف و بر اساس ماهیت متفاوت مقوله‌های مرتبط و نشان دادن ارزش‌گذاری و اهمیت آن‌ها، بر اساس اطلاعات دریافت شده از پدیدآوران در نظام‌های ذخیره و بازیابی اطلاعات، به‌عنوان عنصری کلیدی عمل می‌کند (لوپز-هورتاس، ۱۹۹۷). کاربرد اصلی اصطلاح‌نامه یعنی بازیابی اطلاعات ممکن است با به‌کارگیری آن در نمایه‌سازی یک پایگاه اطلاعاتی و یا کاوش در آن پایگاه اطلاعاتی محقق شود. در استفاده کلاسیک از اصطلاح‌نامه، از آن‌هم در نمایه‌سازی و هم کاوش در یک پایگاه اطلاعاتی استفاده می‌کنند، اما در حالت دوم از اصطلاح‌نامه برای نمایه‌سازی و نه کاوش استفاده می‌کنند و در حالت سوم عکس آن عمل می‌شود. با رواج پایگاه‌های متن کامل، اصطلاح‌نامه نقش‌های جدیدی در نمایه‌سازی و کاوش ایفا می‌کند. اصطلاح‌نامه‌ها مبنای نمایه‌سازی ماشینی هستند. میوزل و دیگران<sup>۱</sup> (۲۰۱۰) روش‌هایی برای گسترش اصطلاح‌نامه‌های موجود با استفاده از ترکیب یادگیری ماشین و پردازش زبان طبیعی ارائه دادند که روش‌های خود را بر روی مش<sup>۲</sup> (اصطلاح‌نامه پزشکی) و وردنت مورد آزمایش قرار دادند (دا سیلوا<sup>۳</sup>، ۲۰۱۲).

از موارد کاربرد اصطلاح‌نامه طبقه‌بندی خودکار متون، نمایه‌سازی ماشینی، ذخیره و

1. Meusel, R. Niepert, M. Eckert, K. Stuckenschmidt, H.

2. MESH

3. Da Sylva, Lyne

بازیابی اطلاعات در بانک‌های اطلاعاتی است. راد و همکاران (۱۳۹۵) روشی جدید برای نمایه‌سازی خودکار و استخراج کلمات کلیدی جهت بازیابی اطلاعات و طبقه‌بندی متون ارائه داده‌اند، آن‌ها تلاش کردند که با استفاده از اطلاعات زبان‌شناختی و اصطلاح‌نامه که از نظامی ساختارمند برخوردار است، شبکه واژگان کلیدی شامل واژگان هم‌ارز، سلسله مراتبی و وابسته را تکمیل کرده و افزایش دهند. در این حالت جامعیت و توافق میان واژگان کلیدی با واژگان جستجوی کاربر افزایش می‌یابد. در مرحله اول واژه‌های غیرمبهم و عام حذف می‌شوند، سپس واژه‌ها ریشه‌یابی می‌شوند و در ادامه با استفاده از شیوه‌های وزن دهی، به هر واژه وزنی اختصاص داده می‌شود که بیانگر میزان تأثیر واژه در ارتباط با موضوع متن، در مقایسه با سایر واژگان به‌کاررفته در متن است. استفاده از اصطلاح‌نامه باعث می‌شود که طبقه‌بندی متون دقیق‌تر انجام گیرد.

رضایی و همکاران (۱۳۹۶) روشی برای طبقه‌بندی خودکار متون با استفاده از اصطلاح‌نامه ارائه دادند. در این پژوهش با استفاده از اصطلاح‌نامه برای ویژگی‌های اصلی استخراج‌شده از بخش پیش‌پردازش، کلمات هم‌خانواده، مترادف‌ها و وابسته‌ها (اعم و اخص) استخراج شدند؛ به عبارت دیگر برای تک‌تک کلمات اصلی متن کلمات هم‌خانواده، مترادف، متضاد، اعم و اخص استخراج و درجایی نگهداری شدند. هدف این کار بوده است که در صورت مشاهده کلمات هم‌خانواده در متن، به‌جای اینکه به صورت مجزا برای هر یک وزنی در نظر گرفته شود، یک کلمه از میان آن‌ها به‌عنوان نماینده انتخاب شود و ضریب وزنی مشخصی به آن اختصاص داده شود و با استفاده از الگوریتم نزدیک‌ترین همسایه متون را طبقه‌بندی نموده‌اند.

### دسترس‌پذیری کاربران به مدارک: بازیابی مدرک یا سند

در کتابخانه‌های سنتی بازیابی مدرک اغلب با فرایند مصاحبه مرجع همراه است، به این مفهوم که یک کتابدار مرجع می‌کوشد نیازهای اطلاعاتی کاربر را به‌طور دقیق تشخیص دهد و مطابق آن در راهبرد جستجو موفق عمل کند که شامل جستجوی برخط یا پیوسته و جستجو در سایر منابع است؛ اما در کتابخانه دیجیتال چنین مرحله‌ای وجود ندارد. کاربران

راهبرد جستجوی خود را به تدریج با پاسخ‌هایی که از نظام دریافت می‌کنند اصلاح می‌کنند. علاوه بر این، ویژگی‌های خاصی از نظام کتابخانه دیجیتال طراحی شده است که وسعت و پیچیدگی جستجویی که کتابدار انجام می‌دهد را شبیه‌سازی می‌کند و بنابراین عبارت «بازیابی مدرک» در یک کتابخانه دیجیتال به «بازیابی اطلاعات» قابل کاستن است. این مسئله احتمالاً بهترین موضوع قابل تحقیق در حوزه مدیریت مدارک است. این امر نشانگر اهمیت نقش فناوری‌های پردازش زبان طبیعی است.

### نظام‌های پرسش و پاسخ

باگذشت زمان خدمات مرجع در کتابخانه تغییراتی نموده و امروزه کتابداران مرجع، نه تنها به شکل سنتی در میز مرجع بلکه از راه دور، در فضای مجازی و در قالب کتابخانه‌های دیجیتال پاسخ‌گوی پرسش‌ها و نیازهای کاربران هستند. یکی از مهم‌ترین خدمات مرجع دیجیتالی پاسخ‌گویی به پرسش‌های کاربران در قالب نظام‌های پرسش و پاسخ است. از آنجایی که در کتابخانه‌های سنتی، بخش عمده‌ای از زمان کتابدار مرجع صرف پاسخ‌گویی به سؤالات کاربران می‌شود، در شیوه‌های مدرن‌تر، به‌ویژه در نظام کتابخانه‌های دیجیتال، اتاق‌های گفتگویی با عنوان «از کتابدار پرس» طراحی کرده‌اند که توسط یک انسان از طریق اینترنت پاسخ‌گویی انجام می‌شود؛ اما ابزارهای پردازش زبان طبیعی می‌توانند به شیوه‌های مختلفی در این زمینه یاری رسانند. نظام پرسش و پاسخ، پرسش کاربر به زبان طبیعی را دریافت کرده و پاسخ او را ارسال می‌کند، در واقع نظام‌های پرسش و پاسخ نوع پیچیده‌تری از نظام‌های بازیابی اطلاعات هستند. در نظام‌های بازیابی اطلاعات فهرستی از مدارک مرتبط با پرسش کاربر بر اساس ربط فراهم می‌شود؛ اما نظام‌های پرسش و پاسخ به‌منظور برطرف ساختن نیاز کاربر و پاسخ به پرسش‌های خاص اوست که به لحاظ تاریخی سابقه طولانی‌ای دارد (گرین و دیگران<sup>۱</sup>، ۱۹۶۱؛ وودز<sup>۲</sup>، ۱۹۷۳). نظام‌های پرسش و پاسخ، پرسش کاربر نه تنها در سطح لغوی، بلکه در سطح نحوی و معنایی نیز

- 
1. Green, Bert F. Wolf Alice K. Chomsky, Carol, Laughery, Kenneth
  2. Woods, William A.

پردازش می‌شوند و پاسخی به زبان طبیعی تولید می‌کنند.

نظام‌های پرسش و پاسخ به دو گونه: نظام‌های پرسش و پاسخ با دامنه بازی نامحدود و نظام‌های پرسش و پاسخ با دامنه محدود تقسیم می‌شوند. در حالت اول نظام‌های پرسش و پاسخ قادر است به تمامی انواع سؤالات در هر زمینه‌ای پاسخگو باشد و مستلزم بهره‌گیری از شبکه‌های معنایی جهت استخراج ارتباطات معنایی می‌باشند (برای مثال نظام انسرباس<sup>۱</sup>)، در حالت دوم نظام‌های پرسش و پاسخ صرفاً قدر به پاسخگویی به سؤالات در زمینه‌های خاصی هستند (برای نمونه نظامان آل اچ پرایمری کر<sup>۲</sup> در حوزه بهداشت و پزشکی) (کاوه یزدی، زارع میرک آباد و بحرانی، ۱۳۸۶). نمونه‌هایی از این نظام‌ها در سال ۱۹۶۰ برای نخستین بار اجرا شدند. از آن جمله می‌توان به «بیسبال<sup>۳</sup>» (پاسخگوی سؤالات در زمینه بیسبال در آمریکا بوده است) و «لونار<sup>۴</sup>» (به سؤالات در حوزه زمین‌شناسی پاسخ گو بوده است).

امروزه تعداد وب‌سایت‌های ارائه‌دهنده خدمات پرسش و پاسخ برخط رو به افزایش است. از جمله این وب‌سایت‌ها می‌توان به گوگل انسر<sup>۵</sup>، چاچا<sup>۶</sup> و واندیر<sup>۷</sup> اشاره کرد. همچنین در زبان فارسی خدمات پرسش و پاسخ برخط در زمینه‌های مختلف علوم و نیز مشاوره با کاربران وجود دارد. مهم‌ترین هدف این خدمات ارائه پاسخی به کاربر است که بیشترین تطابق محتوایی و معنایی را با پرسش کاربر داشته باشد. بدین منظور از فنون معنایی و لغوی استفاده می‌شود. خانی جزنی و ساجدی (۱۳۹۵) یک نظام‌های پرسش و پاسخ فارسی به نام «جويا» را در پژوهش خود ارائه دادند، این سامانه از نوع نظام‌های دامنه نامحدود بوده اما نمی‌تواند به پرسش‌های پیچیده‌ای که نیاز به استدلال دارد پاسخ دهد، همچنین وب-مبنا بوده و اطلاعات خود را از وب اخذ می‌کند.

پاسخ‌گویی به سؤالات کاربران به شیوه‌های مختلفی انجام می‌شود:

1. AnswerBus
2. NLH primary care
3. Baseball
4. Lunar
5. Google Answer
6. Chacha
7. Wondir

سؤالاتی که پاسخ بله یا خیر دارند: برای مثال آیا جرج دبلیو بوش رئیس‌جمهور فعلی ایالات متحده است؟

سؤالاتی که در پاسخ نیازمند ارائه فهرستی هستند: برای مثال کدام جاده‌ها به رم ختم می‌شوند؟ یا کدام تیم‌های فوتبال در این دهه قهرمان لیگ شده‌اند؟

سؤالاتی که در پاسخ مستلزم ارائه دستورالعمل هستند: برای مثال چگونه لازانیا بپزم؟ یا بهترین روش پل ساختن چیست؟

سؤالاتی که در پاسخ نیازمند ارائه توضیح هستند: برای مثال چرا جنگ جهانی اول آغاز شد؟ یا چگونه یک رایانه داده‌ها را پردازش می‌کند؟

سؤالاتی که به شکل امری طرح می‌شوند: برای مثال ارتفاع برج ایفل را به من بگویید.

برای نزدیک شدن به پاسخ پرسش به‌طور استاندارد سه مرحله انجام می‌شود: ۱. تجزیه و تحلیل سؤال، ۲. بازیابی مدارک، ۳. استخراج پاسخ. از آنجایی که سؤالات کاربران به زبان طبیعی مطرح می‌شود، پردازش زبان طبیعی در مرحله اول (تجزیه و تحلیل سؤال) مفید شناخته شده است (گرین وود و دیگران، ۲۰۰۲). در مراحل فوق از فنون مختلف پردازش زبان طبیعی استفاده می‌شود، روش‌های مستقیم شامل عبارات ساده‌ای که با متن مرتبطاند (راویچندران و هوی، ۲۰۰۲) و روش‌های پیچیده‌تری که از پردازش زبانی عمیق استفاده می‌کنند. از مهم‌ترین بخش‌های نظام‌های پرسش و پاسخ، بخش تجزیه و تحلیل سؤال است که به دو گونه تحلیل دستوری و تحلیل معنایی انجام می‌پذیرد. اسکات و گایزوسکاس<sup>۲</sup> (۲۰۰۱) تحلیل‌های معنایی و دقیقی از مدارک بازیابی شده انجام دادند تا بتوانند پاسخ‌های خاص سؤالات را شناسایی کنند. ماهیت وظیفه پاسخ‌گویی به سؤالات که شامل تفسیرسؤالات به زبان طبیعی و شناسایی اطلاعات خاص در مدارک است، رویکردهای استاندارد در بازیابی را ایجاد کرده است که ناکافی می‌باشد، فنون پردازش زبان طبیعی مستلزم انجام برخی فرایندهای افزوده است.

1. Ravichandran, Deepak and Hovy, Eduard

2. Scott, S. and Gaizauskas, R.

برخی سؤالات معنای مشابهی دارند اما با واژگان متفاوت بیان می‌شوند، برای مثال سؤال «آیا بارگذاری فیلم غیرقانونی است» با سؤال «آیا می‌توانم یک کپی از یک لوح فشرده به اشتراک گذارم؟» گرچه از لحاظ واژگان متفاوت‌اند، اما معنای مشابهی دارند. به این منظور روش‌های سنی مانند کمک گرفتن از فرهنگ واژگان یا الگوهای دست‌نویس جوابگو نمی‌باشد، لذا از شیوه‌های پردازش زبان طبیعی در بازیابی اطلاعات استفاده می‌شود. رویکرد معنایی در پی اجرای درجه‌ای از تحلیل‌های معنایی و نحوی و به دنبال راهکارهایی جهت درک زبان طبیعی است. امروزه روش‌های آماری به همراه شیوه‌های معنای همراه باهم مورد استفاده قرار می‌گیرند. یکی از شیوه‌های بهبود کارایی در بازیابی اطلاعات، به دست آوردن روابط میان کلمات است، به این منظور از ساختارهای محاسباتی پیچیده پردازش زبان طبیعی مانند گراف مفهومی در استخراج، استفاده می‌شود. روش‌های مختلفی برای تطابق معنایی ارائه شده است. نمونه‌ای از این نظام‌ها اف‌ای کیو فایندر<sup>۱</sup> است که با استفاده از تکنیک‌های آماری و پردازش زبان طبیعی مبتنی بر هستان‌شناسی به ارائه پاسخ بر اساس پایگاه دانش خود می‌پردازد. این نظام برخلاف سایر نظام‌ها که بر تولید پاسخ جدید تأکید دارند، این نظام پاسخ‌هایی را که از قبل در پایگاه دانشی خود داشته به کاربر ارائه می‌دهد که از سه معیار مقایسه آماری، امتیاز تشابه معنایی و امتیاز شمول استفاده می‌کند. روش دیگر جهت تشخیص تشابه معنایی دو پرسش استفاده از تابع شالوده درخت<sup>۲</sup> است. (ایزدی، ۱۳۹۱) با استفاده از آزمون هجده هزار پرسش و پاسخ ثبت شده در سرویس پرسش و پاسخ راسخون<sup>۳</sup> مدلی بر مبنای شباهت کسینوسی و مدل فضا-برداری ارده کرده که نتایج حاصله بهبود قابل ملاحظه‌ای در تطبیق سؤال با استفاده از مدل‌های زبانی تعمیم یافته است.

- 
1. FAQ Finder
  2. Tree kernel
  3. www.rasekhoon.net



## تصحیح خودکار املاي واژه‌ها<sup>۱</sup>

تصحیح خودکار املاي واژه از دهه ۱۹۶۰ به‌عنوان چالشی در پژوهش‌های پردازش زبان طبیعی مطرح بوده است و در این راستا فنون و الگوریتم‌های فراوانی ایجاد شده است. استفاده مؤثر از راه‌کارهای تصحیح املاي متون در افزایش کارایی سامانه‌های پردازش واژگان، بازیابی اطلاعات، برنامه‌های تصحیح دستور زبان و بازشناسی نوری نویسه‌ها به اثبات رسیده است (کوکیچ<sup>۲</sup>، ۱۹۹۲).

خطاهای املايی به دودسته تقسیم می‌شوند: خطاهای غیر واژه‌ای (خطاهای مفرد و چندگانه) که شامل واژگان فاقد معنا هستند و در واژه‌نامه‌ها وجود ندارند و خطاهای واژه حقیقی که واژه بررسی شده در واژه‌نامه‌ها وجود دارد اما از لحاظ معنایی یا دستوری دارای خطا است. در هر دو حالت فوننی جهت رفع خطا و تصحیح واژه ارائه می‌شود که در خطاهای غیر واژه‌ای فنون تحلیل مدل‌های چندوزنی و استفاده از واژه‌نامه، به کار می‌رود. برای شناسایی خطاهای دستوری از روش‌های مبتنی بر قاعده و برای شناسایی خطاهای املايی از روش‌های آماری استفاده می‌شود. چندین سامانه پردازش زبان طبیعی به‌عنوان خطایاب برای متون فارسی ارائه شده است از جمله ویراستیار که نرم‌افزار خطایاب املاي فارسی بوده و قادر به شناسایی خطاهای غیر واژه‌ای است. سامانه دیگر با عنوان واری‌گر فارسی (وفا) وجود دارد که خطاهای غیر واژه‌ای، خطاهای دستور زبانی و خطاهای واژه حقیقی را شناسایی می‌کند. این خطایاب بر اساس یک واژه‌نامه حجیم که شامل واژه‌های رایج زبان فارسی است، عمل می‌کند که در آن تمام واژگان، ریشه‌ها و مشتقات آن‌ها وجود دارد. به عبارت ساده‌تر این خطایاب برای تشخیص صحیح بودن املاي واژه‌ها به واژه‌نامه خود مراجعه کند، چنانچه واژه موردنظر را نیافت، احتمال خطای نوشتاری می‌دهد و در جهت تصحیح آن اقدام می‌کند. (دستغیب، کلینی و فخر احمد، ۱۳۹۸).

سامانه‌های تصحیح خودکار املاي واژه‌ها در متون در زمینه‌های مختلف، از جمله

- 
1. Automatic spell checking
  2. Kukich, K

تصحیح عبارت پرسش کاربر در نظام‌های پرسش و پاسخ به کار می‌رود. تصحیح خودکار املاي واژه‌ها موجب می‌شود که اشتباه تايي کاربران کتابخانه‌ها و مراکز اطلاع‌رسانی مانع از بازیابی و دسترسی آن‌ها به پاسخ و مدرک مورد جستجو نشود و در نهایت رضایت خاطر بیشتر کاربران را در پی خواهد داشت.

### پیش‌بینی واژه بعدی<sup>۱</sup> در پرسش کاربران

یکی از کاربردهای پیش‌بینی واژه بعدی در کتابخانه‌ها برای افرادی است که دارای ناتوانی جسمی اعم از اختلالات عدم یادگیری در خواندن و نوشتن هستند، در این حالت نیازی نیست که فرد معلول کل عبارت را تایپ کند (ساندار کانتهم و شالینی، ۲۰۰۷). یکی از روش‌های مورد استفاده در پردازش زبان طبیعی استفاده از آن-گرام است. این شیوه بر مبنای پیش‌بینی واژه‌ها استوار است. برای مثال در ادامه عبارت «چگونه تکالیف را...» احتمال زیاد عبارت «انجام بدهم» خواهد آمد. مطابق مدل آن-گرام می‌توان واژه بعدی یا قبلی عبارات را پیش‌بینی کرد. به چنین مدل‌های آماری از دنباله کلمات، مدل‌ها زبانی<sup>۲</sup> نیز گفته می‌شود. بر طبق این مدل می‌توان واژه بعدی را با توجه به  $N-1$  واژه قبلی پیش‌بینی کرد. این مدل در زمینه پردازش گفتار و زبان مهم و کارآمد است، به‌ویژه زمانی که نظام پرسش و پاسخ مجبور به واژه‌های گفتاری از یک ورودی دارای اختلال باشد. همچنین در شناسایی دست خط و غلط‌یابی نیز کاربرد دارد.

### تبدیل گفتار به نوشتار و تبدیل نوشتار به گفتار<sup>۳</sup>

در ادامه تلاش برای برقراری ارتباط میان انسان و ماشین، نظام‌های ارائه خدمات تبدیل گفتار به نوشتار و تبدیل نوشتار به گفتار با استفاده از پردازش زبان طبیعی پا به عرصه گذاشته‌اند. استفاده از خدمات تبدیل صوت به متن و بالعکس، به‌عنوان نوعی از خدمات رابط کاربری باعث افزایش دسترسی طیف وسیعی از کاربران به منابع کتابخانه‌ها خواهد

1. Next word prediction
2. Language Models
3. Speech to text and text to speech

شد، به‌ویژه کاربرانی که دارای اختلالات و ناتوانایی‌های جسمی و حرکتی هستند، می‌توانند از این نوع خدمات در بازیابی اطلاعات مناسب خود، بهره‌جویند. برای مثال استفاده از صوت در سامانه‌های پرسش و پاسخ به‌جای تایپ کردن و ورود اطلاعات نوشتاری برای نابینایان و کم‌بینایان می‌تواند نحوه ارائه خدمات به این گروه از جامعه را سهولت بخشد. همچنین افرادی که سواد نوشتن ندارند می‌توانند از طریق صوت پرسش خود را مطرح کنند. در نظام‌های تبدیل متن به گفتار با پردازش زبان طبیعی، امکان تعیین مشخصه دستوری واژه‌ها، تبدیل متن به واج و رفع ابهام از واژه‌ها را فراهم می‌سازند و با یک نظام مولد صوت با امکان تولید هم‌زمان پارامترهای صوتی گفتار و نیز به کمک یک تحلیل‌گر گفتار، قادرند متن را به گفتار تبدیل کنند.

### پردازش زبان طبیعی و کتاب‌سنجی

مطالعاتی که در آن‌ها فنون پردازش زبان طبیعی و تحلیل‌های کتاب‌سنجی مکمل هم انجام می‌گیرد، در سال‌های اخیر بسیار رواج یافته است. این مطالعات در دو گروه طبقه‌بندی می‌شوند: مطالعاتی که در آن پردازش زبان طبیعی به منظور افزایش عملکرد مطالعات کتاب‌سنجی استفاده می‌شود و مطالعاتی که در آن مقالات پردازش زبان طبیعی با استفاده از روش‌های کتاب‌سنجی تجزیه و تحلیل می‌شوند. از فنون پردازش زبان طبیعی برای بهبود عملکرد مطالعات کتاب‌سنجی استفاده شده است. استخراج اطلاعات که از پرکاربردترین موارد استفاده پردازش زبان طبیعی برای دسترسی به نام نویسندگان، مؤسسات و کشورهای مربوط به آثار یا انتشارات است به‌ویژه هنگامی که مشکلات استاندارد در نمایه‌های استنادی وجود دارد. در چنین مواردی فرایند استخراج خودکار اطلاعات می‌تواند با نرم‌افزارهای پردازش زبان طبیعی مانند نوج<sup>۱</sup>، سافرون<sup>۲</sup> و ایندکس انجام پذیرد. در مطالعات اخیر بیان شده است که از سیستم استخراج اطلاعات مبتنی بر پردازش زبان طبیعی جهت تحلیل و تفسیر عبارات و اصطلاحات خاص در یک مجموعه داده در پایگاه وب آو

---

1. Nooj  
2. Saffron

ساینس<sup>۱</sup> استفاده شده است (لی، رولینز و یان<sup>۲</sup>، ۲۰۱۸). میان کتاب‌سنجی و بازیابی اطلاعات رابطه قوی‌ای وجود دارد به این ترتیب که نمایه‌های استنادی منبع اصلی داده برای مطالعات کتاب‌سنجی است. کیفیت داده‌ها در نمایه‌های استنادی می‌تواند موفقیت این مطالعات را افزایش دهد. ارتباط قوی میان مفاهیم، کتاب‌سنجی‌ها و بازیابی اطلاعات، سازمان‌دهی رویدادهای جدیدی که به این ارتباط می‌پردازند را تسهیل بخشیده است.

فصل مشترک میان کتاب‌سنجی و بازیابی اطلاعات با جزئیات در سال ۲۰۱۳ در کنفرانس بین‌المللی انجمن علم‌سنجی و اطلاع‌سنجی در وین با موضوع «ترکیب کتاب‌سنجی و بازیابی اطلاعات» ارائه شد و مقالات منتخب آن در حوزه علم‌سنجی در سال ۲۰۱۵ منتشر شده است. تمام مطالعات انجام گرفته در راستای کاربردهای پردازش زبان طبیعی به منظور افزایش کیفیت تحقیقات کتاب‌سنجی می‌باشد. کلمات در حال ظهور به وسیله وب اسکریپینگ<sup>۳</sup>، مدل‌های هستی‌شناسی<sup>۴</sup>، کتاب‌سنجی پیشرفته، تحلیل‌های معنایی و حسی تشخیص داده می‌شوند (تاسکین و آل، ۲۰۱۹).

### بازیابی اطلاعات و کتاب‌سنجی

به کارگیری روش‌های کتاب‌سنجی برای تحقیقات بازیابی اطلاعات و برعکس آن در سال‌های اخیر توسعه یافته است، از مدل‌های کتاب‌سنجی فرایندهای سیستم‌های بازیابی به بهره‌برداری روابط استنادی تا قابلیت‌های مرور گسترده جهت شناسایی بالقوه مدارک مرتبط مبتنی بر پیوندهای استنادی برخوردار باشد. کاربرد رویکردهای مبتنی بر زبان‌شناسی هنوز در متون کتاب‌سنجی نسبتاً جدید است. رویکردهای کاربردی و عملی در هر دو حوزه بازیابی اطلاعات و کتاب‌سنجی، مستقیماً از پیشرفت در پردازش زبان طبیعی که بسترهای جدیدی را فراهم کرده است، بهره‌مند شده‌اند. در پایگاه‌های داده‌ای بزرگ تمام متن، هم بازیابی اطلاعات و هم کتاب‌سنجی از پیشرفت‌های پردازش زبان طبیعی و

1. Web of science
2. Li, Kai, Rollins, Jason and Yan, Erjia
3. Web scraping
4. Antology models

زبان‌شناسی محاسباتی بهره می‌برند که در آن تکنیک‌های مبتنی بر متن‌بازیابی مدرک را بهبود بخشیده و توانایی کشف روابط در میان موجودیت‌های موردعلاقه در تحقیقات متریک را داراست. حتی اگر که کتابخانه‌های دیجیتال مخازن انتشارات رسمی یا مجموعه‌ای از مدارک چندرسانه‌ای ناهمگون باشند، بازهم یک محیط ایدئال برای مطالعه بر سه حوزه‌بازیابی اطلاعات، کتاب‌سنجی و پردازش زبان طبیعی هستند. با نمایش محتوا و جستجو که بیشتر در محیط‌های سنتی‌بازیابی وجود دارد، پیوندها، روابطی مشابه آنچه که در تحلیل‌های استنادی یافت می‌شود را تقلید کرده و محتوای تمام متن که معطوف به تجزیه و تحلیل پردازش زبان طبیعی است را نمایان می‌سازد. یک مرور اجمالی مشخص می‌کند که چگونه تحولات بازیابی اطلاعات و کتاب‌سنجی، همراه با روش‌های مبتنی بر زبان‌شناسی، به پیشرفت تحقیقات در هر دو زمینه کمک کرده است. بهره‌گیری از روش‌های فناوری‌های پردازش زبان طبیعی پیشرفته، زمینه جستجوی طبیعی‌تر را برای کاربران فراهم می‌آورد و ارزیابی‌بازیابی را به فراتر از مدل‌های مستقل عبارت ساده بسط می‌دهد. تحقیقات بازیابی اطلاعات بیشتر پاسخگوی فایده‌جستن از فنون پردازش زبان طبیعی به منظور فراهم آوردن محیط طبیعی‌تر برای جستجوی کاربران و ارزش‌افزوده برای بازیابی اطلاعات فراتر از مدل‌های مستقل با عبارات ساده است. اخیراً اقتباس از روش‌های مبتنی بر زبان اعم از پردازش زبان طبیعی و زبان‌شناسی محاسباتی، تحقیقات کتاب‌سنجی و بازیابی اطلاعات را کارآمدتر ساخته است. کاربرد رویکردهای مبتنی بر زبان‌شناسی همچنان در متون کتاب‌سنجی جدید است (ولفرام، ۲۰۱۶).

### بحث و نتیجه‌گیری


در سال‌های اخیر افزایش روزافزون اطلاعات و چالش‌های مربوط به پردازش و بازیابی اطلاعات در حوزه علم اطلاعات و دانش‌شناسی منجر به استفاده از فنون پردازش زبان طبیعی شده است. به کارگیری فناوری‌های رایانه‌ای به‌طور کلی و به‌ویژه استفاده از ابزارهای

پردازش زبان طبیعی زمینه‌ای را فراهم ساخته که کتابخانه‌ها و مراکز اطلاع‌رسانی در جهت کاهش نیروی انسانی در انجام فرایندهای خدمت‌رسانی به کاربران و حرکت به سوی خودکارسازی فرآیندها پیش روند و عملکردی از خود ارائه دهند که در شیوه‌های سنتی امکان‌پذیر نبود. کتابخانه‌های دیجیتال چالش‌های منحصر به فردی جهت ارائه خدمات به کاربران خود دارند که موجب شده به محیطی ایدئال جهت به کارگیری روش‌های پردازش زبان طبیعی در کارکردهای مختلف خود به ویژه بازیابی اطلاعات، تبدیل شوند. ابزارهای فراهم آوری، توصیف و اشاعه منابع که از نیازهای اساسی کتابخانه‌های دیجیتال است، متکی بر منابع سازمان‌دهی دانش است؛ بسیاری از مسائل مرتبط با زبان برای مدیریت کتابخانه دیجیتال باید مدنظر قرار گیرد و از این روی کتابخانه‌های دیجیتال از ابزارهای پردازش زبان طبیعی بهره می‌برند. کتابخانه دیجیتال نشان‌دهنده یک فرصت برای ظهور فناوری‌های زبانشناسی محاسباتی و به ویژه پردازش زبان طبیعی است که می‌تواند بسیاری کاربردهای بالقوه (به ویژه، کاربردهای مالی یا اقتصادی، با توجه به ارزش اقتصادی جدید اطلاعات) را مورد استفاده قرار دهد. تمرکز بر کتابخانه‌های دیجیتال بسیار بزرگ می‌تواند پایداری فناوری به ظاهر تکامل یافته پردازش زبان طبیعی را مورد آزمون قرار دهد. در پژوهش‌های گذشته، نحو نقش مهمی در توسعه پردازش زبان طبیعی داشته است، به ویژه در رویکردهای مبتنی بر ترجمه ماشینی که نظام‌ها با قوانین ترجمه از ساختارهای نحوی یک زبان به زبان دیگر گسترش یافته‌اند؛ اما امروزه نحو نقش بسیار اندکی در پردازش زبان طبیعی برای مدیریت مدارک داشته است. تمرکز تحقیقات کنونی بیشتر بر معناشناختی محاسباتی شامل معناشناسی واژگانی، عبارتی و جمله‌ای و حتی واحدهای سطوح بالاتر است. در واقع، زبانشناسی متن یا تجزیه و تحلیل گفتار، به ویژه برای خلاصه‌سازی و برخی از رویکردهای خاص به رده‌بندی، تحقیقات جدیدی را پیش خواهد برد. در درازمدت، چالش نهایی بیش از ابعاد زبانشناسی صرف، در مدیریت کتابخانه‌های دیجیتال مدل‌سازی خواهد شد و ابعاد شناختی، ارتباطی، عملی، اجتماعی یا ابعاد نشانه‌شناسی و... به آن افزوده خواهد شد. این موارد به طور کلی برای علوم شناختی و هوش مصنوعی می‌تواند جذاب و

مورد توجه واقع شود، اما هنوز در ابعاد زبان‌شناختی دارای چالش‌های فراوانی است. نتایج این مطالعه نشان داد پردازش زبان طبیعی در بسیاری از حوزه‌های فرعی و مرتبط با علم اطلاعات مانند بازیابی اطلاعات، کتاب‌سنجی، استخراج خودکار اطلاعات، نمایه‌سازی خودکار، خلاصه‌سازی خودکار متون، طبقه‌بندی خودکار متون، نظام‌های پرسش و پاسخ و به‌کارگیری فناوری خطایاب املائی، ابهام‌زدایی از عبارات پرسش کاربران و پیش‌بینی واژه‌های مورد نظر آن‌ها، تبدیل گفتار به متن و بالعکس و یاری‌رساندن به افراد دارای معلولیت‌های جسمی مانند کم‌بینایان و نابینایان، نظر کاوی و تحلیل عقیده و احساس کاربران کتابخانه‌ها، مدیریت اسناد و مدارک و... قابل استفاده است. بررسی‌ها نشان می‌دهد، گرچه بازیابی اطلاعات همچنان به‌عنوان یکی از کاربردهای اصلی پردازش زبان طبیعی در علم اطلاعات و دانش‌شناسی است، اما پژوهش‌ها در حوزه مشترک پردازش زبان طبیعی و علم اطلاعات در سال‌های اخیر برای رسیدن به هدف اصلی ارائه خدمات بهینه به کاربران در کتابخانه‌ها و مراکز اطلاع‌رسانی به سمت تحلیل احساس و عقیده کاوی کاربران سوق یافته است. تحلیل نظرات کاربران بر مبنای تحلیل احساس واژگان به افزایش میزان رضایت‌مندی کاربران کتابخانه‌ها منجر خواهد شد و بهینه‌سازی ارائه خدمات را در پی خواهد داشت.

#### ORCID

Mahboubeh Rabiei  <https://orcid.org/0000-0001-8177-4646>

Vahid Reza Mirzaian  <https://orcid.org/0000-0002-51421732>

## منابع

- آقاکاردان، احمد.، کیهانی‌نژاد، مینا. (۱۳۹۱). ارائه مدلی برای استخراج اطلاعات از مستندات متنی، مبتنی بر متن کاوی در حوزه یادگیری الکترونیکی، فصلنامه علمی پژوهشی فناوری اطلاعات و ارتباطات ایران، ۴(۱۱، ۱۲): ۴۷-۵۴.
- اسمعیلی تفت، شیمیا.، شاکری، آزاده. (۱۳۹۴). نظر کاوی بین زبانی با استفاده از ویژگی‌های معنایی، نشریه علمی انجمن کامپیوتر ایران، ۱۳(۲): ۴۷-۵۹.
- ایزدی، سارا. (۱۳۹۱). به‌کارگیری تکنیک‌های پردازش زبان طبیعی برای تطبیق سؤال در سیستم‌های پرسش و پاسخ فارسی، پایان‌نامه کارشناسی ارشد، دانشگاه یزد.
- برادران، راضیه.، گلپر رابوکی، عفت. (۱۳۹۸). فصلنامه علمی پژوهشی پردازش علائم و داده‌ها، ۴۱(۳): ۷۹-۸۸.
- بهره‌ور، مجید.، مهدی‌پور، الهام.، کامل، آزاد. (۱۳۸۷). سیستم خلاصه‌ساز خودکار متن‌های فارسی، چهاردهمین کنفرانس ملی سالانه انجمن کامپیوتر ایران، دانشگاه صنعتی امیرکبیر، تهران.
- بینا، بهاره.، رهگذر، مسعود.، ده‌موبد، آذین. (۱۳۸۶). طبقه‌بندی خودکار متون فارسی، سیزدهمین کنفرانس ملی انجمن کامپیوتر ایران، جزیره کیش.
- پرتی، اعظم‌السادات.، حمیدی، حجت‌الله. (۱۳۹۶). ارائه رویکردی برای مدیریت و سازمان‌دهی اسناد متنی با استفاده از تجزیه و تحلیل هوشمند متن، فصلنامه علمی پژوهشی پژوهشگاه علوم و فناوری اطلاعات ایران، ۳۲(۴): ۱۱۷۱-۱۲۰۲.
- جلالی منش، عمار.، علیدوستی، سیروس.، خسروجردی، محمود. (۱۳۹۲). نمایه‌ساز ماشینی منابع فارسی: مدلی یکپارچه برای پژوهشگاه علوم و فناوری اطلاعات ایران، فصلنامه علمی پژوهشی پژوهشگاه علوم و فناوری اطلاعات ایران، ۲۹(۲): ۴۲۵-۴۵۱.
- جمالی، ایمان.، میرعابدینی، سیدجواد.، هارون‌آبادی، علی. (۱۳۹۶). ارائه یک مدل دسته‌بندی متون فارسی با استفاده از ترکیب روش‌های دسته‌بندی، مجله مهندسی مخابرات، ۷(۲۳): ۳۴-۴۴.
- حریر، نجلا.، هراتی‌زاده، سائنا. (۱۳۹۴). سنجش رضایت‌مندی کاربران از اصطلاح‌نامه علوم اسلامی به‌عنوان ابزار بازیابی اطلاعات، فصلنامه مطالعات ملی کتابداری و سازمان‌دهی



اطلاعات، ۲۶(۲): ۱۴۱-۱۶۰.

حسینی خواه طیه، احمدی، عباس، محبی، آزاده. (۱۳۹۶). بهبود خلاصه‌سازی خودکار متون فارسی با استفاده از روش‌های پردازش زبان طبیعی و گراف شباهت، ۳۳(۲): ۸۸۵-۹۱۴.

خاصه، علی‌اکبر. (۱۳۸۹). داده‌کاوی، متن‌کاوی و وب‌کاوی: تعاریف و کاربردها، مجله الکترونیکی ارتباط علمی، ۵۵: ۱-۶.

خانی جزنی، ایمان، ساجدی، هدیه. (۱۳۹۵). جويا: يك سيستم پرسش و پاسخ فارسی، علوم رایانشی، ۵۱: ۳-۶۶.

داوودآبادی، مرضیه. (۱۳۸۴). پردازش معنایی جملات و اجرای دستورات صادرشده به زبان فارسی، پایان‌نامه کارشناسی ارشد هوش مصنوعی، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی اصفهان.

دستغیب، محمدباقر، کلینی، سارا، فخراحمد، سیدمصطفی. (۱۳۹۸). طراحی و پیاده‌سازی سامانه شناسایی و تصحیح خطای املايي متون فارسی مبتنی بر معنای واژگان، فصلنامه علمی پژوهشی پردازش علائم و داده‌ها، ۴۱(۳): ۱۱۷-۱۲۷.

دولانی، عباس، فرهادپور، محمدرضا. (۱۳۸۸). مروری بر نمایه‌سازی خودکار و نرم‌افزارهای رایج در تولید آن، فصلنامه کتاب، ۷۹(۳): ۲۹۱-۳۱۰.

راد، فرهاد، پروین، حمید، دهباشی، آتوسا، مینائی، بهروز. (۱۳۹۵). ارائه روشی جدید برای شاخص‌گذاری خودکار و استخراج کلمات کلیدی برای بازیابی اطلاعات و خوشه‌بندی متون، فصلنامه علمی پژوهشی پردازش علائم و داده‌ها، ۲۷(۱): ۷۸-۱۰۰.

رضایی، وحیده، محمدپور، مجید، پروین، حمید، نجاتیان، صمد. (۱۳۹۶). ارائه روشی برای استخراج کلمات کلیدی و وزن دهی کلمات برای بهبود طبقه‌بندی متون فارسی، فصلنامه علمی پژوهشی پردازش علائم و داده‌ها، ۳۴(۴): ۵۵-۷۸.

زبردست، مریم. (۱۳۸۹). خدمات مرجع دیجیتال، با نگاهی به سامانه پرسش از کتابدار سازمان کتابخانه‌ها، موزه‌ها و مرکز اسناد آستان قدس رضوی، ۲(۶): ۱-۲۰.

سپهریان، زهرا، سدیدیپور، سعیده‌سادات، شیرازی، حسین. (۱۳۹۳). روش مبتنی بر شباهت معنایی در خلاصه‌سازی متون فارسی بر اساس عبارت پرس و جوی کاربر، مجله علمی پژوهشی پدافند الکترونیکی و سایبری، ۲(۳): ۵۱-۶۳.

سنجی، مجید، داورپناه، محمدرضا. (۱۳۸۸). **شناسایی واژه‌های غیر مفهومی (رایج) در نمایه‌سازی خودکار مدارک فارسی**، فصلنامه کتابداری و اطلاع‌رسانی، ۴۸(۴): ۹-۳۶.

شیخان، منصور، نصیرزاده، مجید، دفتریان، علی. (۱۳۸۴). طراحی و پیاده‌سازی سیستم تبدیل متن به گفتار طبیعی برای زبان فارسی، نشریه دانشکده مهندسی، ۱۷(۲): ۳۱-۴۸.

صنعت‌جو، اعظم. (۱۳۸۴). ضرورت بازنگری در ساختار اصطلاح‌نامه‌ها: بررسی عدم کارایی اصطلاح‌نامه‌ها در محیط اطلاعاتی جدید و قابلیت‌های هستی‌شناسی‌ها در مقایسه با آن، فصلنامه کتاب، ۶۴: ۷۹-۹۲.

طالبیان کوچکسرایبی، مهدی. (۱۳۸۶). استخراج اتوماتیک اطلاعات بر اساس آنتالوژی، پایان‌نامه کارشناسی ارشد، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات.

فرهادپور، محمدرضا، مطلبی، داریوش. (۱۳۹۰). اینفوکریمستال‌ها و کاربرد آن‌ها در بازیابی اطلاعات، نشریه مطالعات ملی کتابداری و سازمان‌دهی اطلاعات، ۲۲(۳): ۲۴-۴۵.

کاوه‌یزدی، فاطمه، زارع میرک‌آباد، محمدرضا، بحرانی، محمد. (۱۳۸۶). طراحی و پیاده‌سازی یک نمونه سیستم پرسش و پاسخ، سیزدهمین کنفرانس ملی انجمن کامپیوتر ایران، جزیره کیش.

گیلوری، عباس. (۱۳۷۹). نمایه‌سازی خودکار (گذشته، حال، آینده)، تحقیقات اطلاع‌رسانی و کتابخانه‌های عمومی، ۱۰(۴): ۱۷-۲۵.

نعیمی، فاطمه، قدس، وحید. (۱۳۹۸). سنتز گفتار فارسی با استفاده از فرکانس گام در نرم‌افزار Flite، پردازش سیگنال پیشرفته، ۳(۱): ۹۷-۱۰۷.

نوربهبهانی، سیدفخرالدین. (۱۳۹۷). نظر کاوی افزایشی با استفاده از یادگیری فعال بر روی جریان متون، نشریه مهندسی برق و مهندسی کامپیوتر ایران، ۱۶(۴): ۲۹۱-۳۰۰.

نیاکان، شهرزاد. (۱۳۸۳). نمایه‌سازی ماشینی، تهران، مرکز اطلاعات و مدارک علمی ایران.

ویسی، هادی، پارسافرد، پویان. (۱۳۹۸). مروری بر روش‌ها و پژوهش‌های دسته‌بندی خودکار متون، علوم رایانشی، ۱۳(۱): ۲-۲۳.

## References

Baxendale, P.B. (1958). Machine-made index for technical literature: An experiment, *IBM journal of research and development*, 2(4): 354-361.

- Berger, Helmut and Merkl, Dieter (2004). *A comparison of text categorization methods applied to N-gram frequency statistics*, in Webb, G.I. Yu X. (eds) AI 2004, Advances in artificial intelligence, AI 2004, Lecture notes in computer science, Springer, Berlin, Heidelberg, 3339:998-1003.
- Black, Catherine. (2011). *Text mining annual review of information science and technology*, 44(1):121-155.
- Blitzer, John. (2008). *A survey of dimensionality reduction techniques for natural language*, [on line], available on [http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.379.3965&rep=rep1](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.379.3965&rep=rep1&type=pdf)
- Borgman, Christine.L. (1997). *Multi media, multi cultural and multi lingual digital libraries: or how do we exchange data in 400 languages?* Dlib Magazine [on line], available on <http://www.dlib.org>.
- Borlund, Pia. (2003). The concept of relevance in IR, *Journal of the American society for information science and technology*, 54(10): 913-925.
- Chanod, Jean Pierre (1999). Natural language processing and digital libraries, in proceeding of applied natural language processing Washington DC, *Information extraction, toward scalable, adaptable system*:17-31
- Chowdhury, G. (2003). *Natural language processing*, *Annual review of information science and technology*, 37: 51-89.
- Cunningham, Hamish (2005). *Information extraction, automatic*, *Encyclopedia of language and linguistics*, 3(8): 10.
- Da Sylva, Lyne (2012). NLP and digital library management, in Bandyopadhyay, Sivaji; Naskar, Sudip Kumar; Ekbal, Asif (réds), *Emerging applications of natural language processing: Concepts and New Research*. Hershey, PA: IGI global, 265-290 (<https://doi.org/10.4018/978-1-4666-2169-5.ch011>).
- Dunning, Ted (1993). *Accurate methods for statistics of surprise and coincidence*, *Computational linguistics*, 19(1): 61-74.
- Evans, David A. Ginther Webster, Kimberly, Hart, Mary, Lefferts, Robert G. and Monarch, Ira A. (1991). Automatic indexing using selective NLP and first order thesauri, RIAO (conference), *intelligent text and image handling*, v.2: 624-643.
- Fang, X. Zhan, J. (2015) Sentiment analysis using product review data. *Journal of Big Data* 2, 5. <https://doi.org/10.1186/s40537-015-0015-2>
- Feldman, S. (1999). *NLP meets the jabberwocky*. Online, 23, 62-72.
- Green, Bert F. Wolf Alice K. Chomsky, Carol, Laughery, Kenneth (1961). Baseball: an automatic question-answer, proc Western joint

- IREAIEE-ACM *computer conference, New York, NY, USA*:219-224.
- Gupta, Anupama, Banerjee Imon and Rubin, Daniel L. (2018). Automatic information extraction from unstructured mammography reports using distributed semantics, *Journal of Biomedical informatics*,78: 78-86 (<https://doi.org/10.1016/j.jbi.2017.12.016>).
- Gupta, Vishal. and Lehal,Gurpreet.S. (2009). A survey of text mining techniques and applications, *Jurnal of imerging technologies in web intelligence*:60-77.
- Hassel,Martin and Mazdak,Nima(2004). *FarsiSum: A Persian text summarizer*, Proceedings of the workshop on computational approaches to Arabic script-based languages, Geneva, Switzerland: 82-84.
- Ingwersen,Peter (2002). *Information retrieval interaction*, London, Tylor Graham.
- Karoo, Krishna (2017). Natural language processing and digital library management system, *International journal of science and research (IJSR)*, 7(11): 1580-1584.
- Knight, Kevin (1999).*Mining online text*,*Communication of the ACM*,42(11): 58-61.
- Korycinsky, C. and Newell, Alan F. (1990). *Natural language processing and automatic indexing*, 17(1): 21-29.
- Krovetz, R. (1997) Homonymy and Polysemy in Information Retrieval. Proceedings of the 35th Meeting of the Association for Computational Linguistics and the 8th Meeting of the European *Chapter of the Association for Computational Linguistics (ACL/EACL-97)*.
- Kukich, K. (1992). Techniques for automatically correcting words in text, *ACM computer surveys (CSUR)*, 24: 377-439.
- Lesk, Michael. (1986). Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. *In proceedings of ACM CIGDOC conference, Toronto, Canada*: 24-26.
- Lee, ChristopherA. Woods, Kam (2017). Diverse digital collections meet diverse uses: applying natural language processing to born digital primary sources,presented at proceeding of the 14<sup>th</sup> *international conference on digital preservation(iPRES)*,Kyoto,Japan.
- Lee, Richard (1998). Automatic information extraction from documents: A tool for intelligence and law inforcement analysts, in proceeding of 1998 AAAI fall symposium on artificial intelligence and link analysis (Menlo park CA), *American association for artificial intelligence*,CA:AAAI press: 63-65.
- Lewis, David D. and Jones, Karen Sparck (1996).*Natural language processing for information retrieval, Communications of*

- ACM,39(1):92- 101.
- Li, Kai, Rollins, Jason and Yan, Erjia (2018).Web of science use in published research and review papers 1997-2017: a selective, dynamic, cross domain, *content-based analysis*,*Scientometrics*, 115: 1-20.
- Liddy,Elizabeth D. (2018).Natural language processing,in Encyclopedia of library and information science,4<sup>th</sup> ed,CRC press (DOI<https://doi.org/10.1081/E-ELIS4>).
- Liu, Bing (2012). Sentiment analyzing and opinion mining, Synthesis lectures on human language technology, *Morgan & Claypool*,5(1): 1-167.
- Luhn,H. (1957). A statistical approach to mechanized encoding and searching of literacy information,*IBM journal of research and development*, 1(4):309-317.
- Mayr, Philipp, Frommholz, Ingo, Cabanac, Guillaume, Chandrasekaran,Kumar, Muthu, Jaidka,Kokil, Kan, Min Yen and Wolfram, Dietmar(2018). Introduction to special issue on bibliographic enhanced information retrieval and natural language processing for digital libraries (BIRNDL), *International journal of digital libraries*, 19(2-3):107-111.
- Meusel, R. Niepert, M. Eckert, K. Stuckenschmidt, H. (2010). Thesaurus extension using web search engines, *In proceeding of ICADL 2010*: 198-207.
- Moens, Marie Francine (2003). *Automatic indexing and abstracting of document texts*, second edition,Massachusetts, MA,Kluwer.
- Morse, Emile, Lewis, Michael and Olsen, Kai A. (2001). Testing visul information retrieval methodologies case study: comparative analysis of textual icon, graphical and spring display, *Journal of American society for information science and technology*, 53(1): 28-40.
- Moschitti, Alessandro (2003).*Natural language processing and automated text categorization: a study on the reciprocal beneficial interactions*,P.hDDisertationdepartment of computer science university of rome.
- Nenkova, Ani and Mackeown, Kathleen (2012). *A survey of text summarization techniques*, IN C.C.Aggarwal& C. Zhai, ed. Mining text data, Boston, MA Springer US: 43-76.
- Peters,C. and Picchi,E. (1997).Across languages, across cultures: issues in multilinguality and digital libraries, *D-Lib Magazine*. [on line] available on <http://www.dlib.org>.
- Rabertson, S. and Spark, Jones, K. (1997).*Simple proven approaches to text retrieval*.Technical report TR 356, Cambridge university computer laboratory.

- Ravichandran, Deepak and Hovy, Eduard (2002). *Learning surface text patterns for a question answering system*, In ACL,02, Proc, 40<sup>th</sup> annual meeting on association for computational linguistics: 41-47.
- Rodgers, Peter, Gaizauskas, Robert, Humphreys, Kevin and Cunningham, Hamish (1997). Visual execution and data visualization in natural language processing, in *proceeding of IEEE symposium on visual language*: 338-343.
- Rajman, M. and Besancon, R. (1998). text mining: natural language techniques and text mining applications. In: Spaccapietra, S. Maryanski, F. (eds). *Data mining and reserve engineering. IFIP, Springer, Boston*: 50-64.
- Rubin, Victoria L. Chen, Yimin (2013). Information manipulation classification theory for LIS and NLP, in *proceeding of American society and information science and technology (ASIST)*, 49(1): 1-5 (<https://doi.org/10.1002/meet.14504901353>).
- Russel-Ross, Tony, Stevenson, Mark (2009). The role of natural language processing in information retrieval: searching for meaning and structure, in A. Goker and J. Davies, eds, *information retrieval: searching in the 21<sup>st</sup> century*, Wiley, 2: 215-232.
- Saracoglu, Ridvan, Tutuncu, Kemal and Allahverdi, Novruz (2008). A new approach on search for similar documents with multiple categories using fuzzy clustering, *Expert systems with applications*, 34(4): 600-605 (<https://doi.org/10.1016/j.eswa.2007.04.003>).
- Scott, S. and Gaizauskas, R. (2001). *University of sheffield TREC-9 Q & A system*. Proceeding 9<sup>th</sup> text retrieval conference. NIST special publication, 500-249: 635-644.
- Shiri, A. Revie, C (2005). Usability and user perceptions of thesaurus enhanced search interface, *Journal of documentation*, 61(5): 640-656.
- Shruthi, Jand Swamy, Suma (2019). Effectiveness of recent research approaches in natural language processing on data science-an insight, *Springer nature Switzerland*: 172-182.
- Smeaton, Alan F. (1995). *Natural language processing and information retrieval*, a lecture presented at the European summer school in information retrieval Glasgow.
- Smeaton, Alan F. (1992). Progress in the application of natural language processing to information retrieval tasks, *the computer journal*, 35(3): 268-278.
- Taskin, Zehra, Al, Umut (2019). Natural language processing applications in library and information science, *online information review*, 43(4): 676-690 (<https://doi.org/10.1108/oir-07-2018-0217>).
- Tsatsaronis, George, Varlamis, Iraklis and Vazirgiannis, Michalis (2010). Text relatedness based on word thesaurus, *Journal of artificial*

*intelligence research*, 37: 1-39.

- Vickery, Alina and Vickery, Brian C. (2005). *Information science in theory and practice*, 3<sup>rd</sup> ed, Walter de Gruyter.
- Voorhees, Ellen M. (1999). natural language processing and information retrieval, in proceeding of applied natural language processing and information retrieval, *Information extraction, toward scalable, adaptable system*: 32-48.
- Wolfram, Dietmar (2016). Bibliometrics, information retrieval and natural language processing: natural synergies to support digital research, in proceeding of the joint workshop on bibliometric enhanced information retrieval and natural language processing for digital libraries (BIRNDL): 6-13.
- Woods, William A. (1973). Progress in natural language understanding: an application to lunar geology, In American Federation of Information Processing Societies, *National computer conference*, 42: 441-450.
- Xi, Sumei (2013). Application of natural language processing for information retrieval, *Applied mechanics and materials, Trans tech*, 380,384: 2614-2618.
- Zerual, Imad and Lakhouaja, Abdelhak (2018). Data science in light of natural language processing: *an overview, Procedia computer science*, 127: 82-91.
- Zuo, Feng, Wang, Fu Lee, Deng, Xiaotie, Han, Song, and Wang, Lu Sheng (2006), WSEAS transaction on information science and applications, 3(6): 1036-1044.

#### **References [in Persian]**

- Aghakardan, Ahmad and Keyhaninejad, Mina (2011). Presenting a model for extracting information from text documents, based on text mining in the field of e-learning, Iran Information and Communication Technology *Scientific Research Quarterly*, 4(12,11): 47-54. [in Persian]
- Baradaran, Razieh and Golpar Rabooki, Effat (2018), Scientific Research Quarterly *Journal of Signal and Data Processing*, 41(3): 79-88. [in Persian]
- Bahrevar, Majid, Mahdipour, Elham and Kamel, Azad (2007). *Automatic Persian text summarization system*, 14th annual national conference of Iran Computer Association, Amirkabir University of Technology, Tehran. [in Persian]
- Bina, Bahareh, Rahgozar, Masoud, and Dehmoobad, Azin (2006). *Automatic Classification of Persian Texts*, 13th National Conference of Iranian Computer Association, Kish Island. [in Persian]
- Davoodabadi, Marzieh (2005). *Semantic processing of sentences and*

- execution of commands issued in Farsi language*, artificial intelligence master's thesis, Faculty of Electrical and Computer Engineering, Isfahan University of Technology. [in Persian]
- Dastgheib, Mohammad Baqer, Keilini, Sara and Fakhrahamd, Seyed Mostafa (2018). Designing and implementing a system for identifying and correcting spelling errors in Persian texts based on the meaning of words, *Scientific Research Quarterly of Signal and Data Processing*, 41(3): 117-127. [in Persian]
- Dolani, Abbas and Farhadpour, Mohammadreza (2008). A review of automatic indexing and common software in its production, *Book Quarterly*, 79(3): 291-310. [in Persian]
- Farhadpour, Mohammad Reza and Motalebi, Dariush (2018). Infocrystals and their application in information retrieval, *Journal of National Library Studies and Information Organization*, 22(3):24-45. [in Persian]
- Guillory, Abbas (1379). Automated Indexing (Past, Present, Future), *Information Research and Public Libraries*, 10(4): 17-25. [in Persian]
- Harir, Najla and Heratizadeh, Saina (2014). Measuring users' satisfaction with the thesaurus of Islamic sciences as an information retrieval tool, *National Library and Information Organization Studies Quarterly*, 26(2): 141-160. [in Persian]
- Hosseinikhah, Tayyebeh, Ahmadi, Abbas and Mohebi, Azadeh (2016). Improving automatic Persian text summarization using natural language processing and similarity graph methods, 33 (2): 885-914. [in Persian]
- Ismaili Taft, Shima and Shakeri, Azadeh (2014). Cross-language analysis using semantic features, *Scientific Journal of Iran Computer Association*, 13(2): 47-59. [in Persian]
- Izadi, Sara (2011). *Application of natural language processing techniques for matching questions in Persian question and answer systems*, Master's thesis, Yazd University. [in Persian]
- Jalalimanesh, Ammar, Alidousti, Sirous and Khosrojerdi, Mahmoud (2012). Machine indexing of Persian sources: an integrated model for Iran Research Institute of Science and Information Technology, *Scientific Research Quarterly of Iran Research Institute of Science and Information Technology*, 29(2): 425-451. [in Persian]
- Jamali, Iman, Miraabedini, Seyed Javad and Haroonabadi, Ali (2016). Presenting a classification model of Persian texts using a combination of classification methods, *Journal of Telecommunication Engineering*, 7(23): 34-44. [in Persian]
- Khaseh, Ali Akbar (2010). Data mining, text mining and web mining: definitions and applications, *electronic journal of scientific*



- communication*, 55: 1-6. [in Persian]
- Khani Jazni, Iman and Sajedi, Hedieh (2015). Jouya: A Persian Question and Answer System, *Computer Science*, 3:51-66. [in Persian]
- Kaveh Yazdi, Fatemeh, Zare Mirakabad, Mohammad Reza and Bahrani, Mohammad (2006). *Designing and implementing a sample question and answer system*, the 13th National Conference of the Iranian Computer Association, Kish Island. [in Persian]
- Naeimi, Fatiemeh and Qods, Vahid (2018). Farsi speech synthesis using step frequency in Flite software, *Advanced Signal Processing*, 3(1): 97-107. [in Persian]
- Noorbehbahani, Seyed Fakhreddin (2017). Incremental analysis using active learning on text flow, *Iranian Journal of Electrical Engineering and Computer Engineering*, 16(4): 291-300. [in Persian]
- Niakan, Shahrzad (2013). Machine indexing, Tehran, *Iran Scientific Information and Documents Center*. [in Persian]
- Parei, Azam al-Sadat and Hamidi, Hojatolah (2016). Presenting an approach for managing and organizing text documents using intelligent text analysis, *Scientific Research Quarterly of Iran Information Science and Technology Research Institute*, 32 (4): 1171-1202. [in Persian]
- Rad, Farhad, Parveen, Hamid, Dehbashi, Atusa and Minaei, Behrouz (2015). Presenting a new method for automatic indexing and keyword extraction for information retrieval and text clustering, *Journal of Signal and Data Processing*, 27(1): 78-100. [in Persian]
- Rezaei, Vahideh, Mohammadpour, Majid, Parvin, Hamid and Nejatian, Samad (2016). Presenting a method for extracting keywords and weighting words to improve the classification of Persian texts, *Scientific Research Quarterly of Signal and Data Processing*, 34(4): 55-78. [in Persian]
- Sepehrian, Zahra, Sadidpour, Saeideh Sadat and Shirazi, Hossein (2014). The method based on semantic similarity in summarizing Persian texts based on the user's query phrase, *Scientific Research Journal of Electronic and Cyber Defense*, 2(3): 51-63. [in Persian]
- Sanji, Majid and Davarpanah, Mohammad Reza (2008). Identification of non-conceptual (common) words in automatic indexing of Persian documents, *Library and Information Quarterly*, 48(4): 9-36. [in Persian]
- Sheikhan, Mansour, Nasirzadeh, Majid and Daftarian, Ali (2004). Design and implementation of text to natural speech conversion system for Persian language, *Journal of Faculty of Engineering*, 17(2): 31-48. [in Persian]
- Sanatjo, Azam (2005). The necessity of revising the structure of thesauruses: investigating the ineffectiveness of thesauruses in the new information

- environment and the capabilities of ontologies compared to it, *Book Quarterly*, 64: 79-92. [in Persian]
- Talebian Khouchaksaraei, Mehdi (2006). *Automatic information extraction based on an ontology*, Master's thesis, Islamic Azad University, Science and Research Unit. [in Persian]
- Veysei, Hadi and Parsafard, Pouyan (2018). A review of automatic text classification methods and researches, *Computer Science*, 13(1):2-23. [in Persian]
- Zebardast, Maryam (2009). Digital reference services, with a look at the system of asking the librarians of the Organization of Libraries, *Museums and Documents Center of Astan Quds Razavi*, 2 (6): 1-20. [in Persian]



پژوهشگاه علوم انسانی و مطالعات فرهنگی  
رتال جامع علوم انسانی

**استناد به این مقاله:** ربیعی، محبوبه، میرزاییان، وحیدرضا. (۱۴۰۱). کاربردهای پردازش زبان طبیعی در علم اطلاعات و دانش‌شناسی با تأکید بر کتابخانه‌های دیجیتال، فصلنامه علمی بازیابی دانش و نظام‌های معنایی، ۹(۳۳)، ۱۹۷-۲۶۲.

DOI: 10.22054/jks.2022.64237.1478



Name of Journal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.