

قاموسی نور

واژه‌نامه هوشمند عربی



مسیح توحیدی وحدت*

mtohidi@gmail.com

حبیب سریانی**

hseryani@noornet.net



چکیده

مراجعه به منابع لغوی برای درک صحیح معانی کلمات، یکی از نیازهای اولیه در فرآیندهای پژوهشی حوزه علوم اسلامی است. هر پژوهشگر، جهت فهم معانی کلمات یا عبارات مد نظر خود، هنگام مراجعه به منابع لغوی در نسخه‌های چاپی سنتی، فرآیندی را در چندین مرحله طی می‌نماید تا به مقصود نهایی خویش، یعنی یافتن توصیفات مندرج ذیل مداخل لغوی، دست یابد. این فرآیندها، در خصوص زبان عربی که کلمات دست‌خوش تغییرات متعددی در ساختار صرفی خود شده‌اند، پیچیده‌تر است. هدف از طراحی و تولید «قاموس نور»، به عنوان یک لغت‌نامه هوشمند الکترونیکی، گذر از شیوه‌های سنتی مراجعه به منابع لغوی متعدد و پیاده‌سازی حداکثری الگوهای ماشینی هوشمند جهت ارائه سریع پاسخ‌های صحیح به پژوهشگران است.

ساماندهی لغت‌نامه‌های متقدم عربی و نظام‌بندی این منابع ارزشمند بر اساس یک ساختار واحد و همچنین تحلیل صرفی کلمات ورودی کاربران توسط دو موتور قدرتمند تحلیلگر صرفی نور و استمر هوشمند نور جهت اتصال ماشینی به نزدیک‌ترین مداخل لغوی، از دستاوردهای منحصر به فرد این واژه‌نامه هوشمند الکترونیکی است که در قالب پایگاه اینترنتی، نرم‌افزار رومیزی و تلفن‌های همراه قابل استفاده است. همچنین، با ایجاد شبکه واژگانی بین مداخل، تطبیق بر آیات قرآن و روایات سرشار از معارف غنی اهل بیت (ع)، اطلاعات متنوع و ارزشمند دیگری نیز در اختیار پژوهشگر قرار می‌گیرد که این واژه‌نامه هوشمند را از دیگر رقبا خود متمایز می‌سازد.

کلیدواژگان: زبان عربی، مدخل لغوی، لغت‌نامه هوشمند، پیکره‌نگاری، شبکه واژگان عربی، تحلیل صرفی.

**مدیر گروه ادبیات و متن‌کاوی مرکز تحقیقات کامپیوتری علوم اسلامی.

**پژوهشگر ادبیات و متن‌کاوی مرکز تحقیقات کامپیوتری علوم اسلامی.

مراحل طراحی و تولید لغت‌نامه هوشمند عربی

مراحل طراحی و تولید لغت‌نامه هوشمند «قاموس نور»، مشتمل بر سه امر است:

۱. ساماندهی ماشینی و دستی لغت‌نامه‌های معتبر زبان عربی کلاسیک و عربی معاصر در ساختاری واحد؛ به گونه‌ای که معنای لغات مد نظر پژوهشگران، به راحتی در اختیار آنان قرار گیرد و دیگر نیازی به تورق در منابع متعدد نباشد.

۲. ارائه یک سیستم هوشمند قوی جهت شناسایی و تحلیل صرفی کلمات و اتصال آن کلمه به نزدیک‌ترین مشتق مدخل لغوی.

۳. ارائه اطلاعات منحصر به فردی از کلمات عربی در دایره شبکه واژگانی، آیات، احادیث و اصطلاحات، به صورت خودکار.

تهیه ساختاری نظام‌مند از مداخل لغوی

دستیابی به یک پیکره زبانی منسجم از داده‌های موجود در فرهنگ‌نامه‌ها، همواره از دغدغه‌های جدی صاحب‌نظران این عرصه بوده است (مطهری و شمس‌فرد، ۱۳۹۳: ۱۱۷۸-۱۱۸۲). از این رو، نخستین مرحله در طراحی و ساخت یک لغت‌نامه هوشمند، ایجاد ساختاری دقیق و منسجم از مداخل لغوی مشترک میان واژه‌نامه‌های مختلف و تهیه داده‌ای مستقل از تمامی این مداخل به همراه توضیحات ضبط‌شده برای آنهاست. تاکنون استفاده از لغت‌نامه در نرم‌افزارهای مرکز نور، بر محور «ریشه» بوده است؛ یعنی کاربران برای پیدا کردن معنای یک کلمه، با واسطه ریشه - آن هم در تعداد کمی از لغات - به لغت‌نامه‌ها منتقل شده و می‌بایست به صورت دستی، به نزدیک‌ترین مشتق مورد نظر خود

می‌رسیدند. حال با استفاده از لغت‌نامه هوشمند، تمام سعی طراحان بر این است با استفاده از فرآیندهای هوشمندسازی، شیوه استفاده از لغت‌نامه از ریشه به «مدخل محور» ارتقا یابد؛ یعنی کاربران به صورت هوشمند به نزدیک‌ترین مدخل لغوی دست یابند. از این رو، هدف نهایی از ایجاد یک پیکره زبانی یکپارچه و منسجم، در اختیار داشتن تمامی مداخل لغوی موجود در زبان عربی به همراه تمامی معانی و مفاهیم مندرج در لغت‌نامه‌های معتبر به شکلی منظم و رتبه‌بندی‌شده همراه نشانی دقیق به منابع اصیل خود است تا بتوان در سیستم‌های بازبازی اطلاعات نیز به بهترین نتایج دست یافت. از این داده، در طول این نگاشته، تعبیر به «بانک مداخل» می‌شود. جهت دستیابی به این مهم، مراحل ذیل در قالب فعالیت‌های ماشینی و بازبینی نیروی خیره انسانی صورت گرفته است:

۱. استفاده از علایم ویرایشی در منابعی که دارای نظم قابل توجهی در چاپ و نشر بوده‌اند

به منظور تسریع و دقت در استخراج مداخل و توصیفات و اجتناب حداکثری از انجام کار دستی، منابع لغوی که علاوه بر دارا بودن اتقان محتوایی، دارای ساختار منسجم و تقریباً یکپارچه نیز هستند، شناسایی شده و با بررسی دقیق متن این کتب، قواعدی برای شناسایی ماشینی مداخل و توصیفات استخراج می‌گردد. این قواعد، متناسب با هر کتاب و شیوه نگارش و میزان علایم ویرایشی موجود همچون: رنگ، نقطه، ویرگول، دو نقطه و غیره، در پردازش ماشینی متن آن کتاب، کلید راهنمای استخراج مداخل و توصیفات قرار گرفته است. بدیهی است که با بهره‌گیری از این شیوه و



تصویر شماره ۱: نمونه‌ای از نسخه‌های اولیه ابزار مداخل

مدخل اصلی عنوان کرده و توصیف را برای مجموع عبارت قرار داده‌اند. بنابراین، مداخل اصلی در مجموع منابع لغوی، هرچند چیزی در حدود یک صد هزار واژه است، ولی با توجه به تنوع در نگارش و انتقال مطلب به مخاطب، با تعدادی بالغ بر نهصد هزار کلمه و عبارت مختلف مواجه هستیم؛ به عنوان نمونه، برای بیان مفهوم مدخل اصلی «سَمَن» در کتب مختلف لغوی از تعابیر: «سمن الطعام»، «سمن الخبز»، «سمنت له»، «سمنت الطعام»، «سمن القوم» و «سمن له» به عنوان مدخل استفاده شده و متناسب با هریک از این مداخل، توصیفات نیز ذکر گردیده است.

در مواردی که مداخل لغوی چندکلمه‌ای بوده و ترکیب این کلمات به عنوان یک «مدخل مرکب» مورد نظر مؤلف نبوده و تنها از جهت مثال، آن را به شکل چندکلمه‌ای بیان داشته، اقدام به تعیین «مدخل نماینده» برای هریک از این مداخل چندکلمه‌ای می‌گردد. «مدخل نماینده»، کلمه‌ای در بین چند کلمه به کاررفته در «مدخل» است که بار اصلی مفهوم، متکی بر آن بوده و سایر کلمات، برای تکمیل آن به کمک گرفته شده‌اند؛ به عنوان مثال، در عبارت «نَهَلَ الرَّجُلُ: عَطَشٌ أَشَدُّ الْعَطَشِ» (العین، ج ۴: ۱۵۲)، کلمه «الرَّجُلُ» موضوعیتی نداشته و مدخل مرکب «نَهَلَ الرَّجُلُ» ذیل مدخل نماینده «نَهَلَ» قرار خواهد گرفت؛ برخلاف مثال «المسجد الحرام» که از همنشینی این دو کلمه، یک کلمه جدید با مفهومی جدید ایجاد شده است. در مداخل لغوی غیرمرکب، پیراسته مداخل به عنوان «مدخل نماینده» در نظر گرفته شده است.

جهت تعیین ماشینی مدخل نماینده از کلمات و عبارات لغوی، از

استفاده از ماشین برای استخراج اطلاعات، در بسیاری موارد از شیوه برچسب‌گذاری دستی متون توسط محققان بی‌نیاز گشته و علاوه بر کاهش هزینه مالی و زمانی، از بسیاری از اشکال‌های ناشی از خطای نیروی انسانی نیز اجتناب می‌شود؛ هرچند ممکن است قواعد در نظر گرفته‌شده در استخراج مداخل و توصیفات در مواردی جامع و مانع نبوده و یا علایم ویرایشی موجود در متون که مبنای کارآمدی یک قاعده بوده است، در مواردی درست اعمال نشده باشد. بنابراین، مقرر گردید تا اطلاعات استخراج‌شده، توسط پژوهشگران مورد بازبینی و احیاناً اصلاح یا ویرایش قرار گیرد. برای تحقق تمام مراحل ساماندهی اطلاعات موجود در لغت‌نامه‌ها و تولید مداخل و توصیفات، نیاز به ابزاری است تا امکانات لازم را در اختیار محققان بخش ادبیات و متن‌کاوی مرکز تحقیقات کامپیوتری نور جهت اصلاح یا ویرایش اطلاعات استخراج‌شده قرار دهد. این برنامه با نام «ابزار مداخل»، پس از طراحی و تولید اولیه، متناسب با نیازهای هوشمندسازی و پروژه‌های بازبینی، ارتقا می‌یابد. (تصویر شماره ۱)

۲. تعیین مدخل نماینده توسط ماشین با استفاده از نشان‌ها (فرمت‌ها) و ریشه

از آنجاکه «لغت‌نامه هوشمند» به دنبال ارائه اطلاعات یکپارچه و منسجم از منابع مختلف لغوی است، لازم است بین مداخل لغوی در لغت‌نامه‌های مختلف، هماهنگی ایجاد شود. در اکثر منابع لغوی متقدم، مؤلفان جهت تفهیم معنای یک مدخل، کلماتی را همراه با

برخی از لغت‌نامه‌های تایپ‌شده در مرکز نور، طی سالیان متمادی توسط محققان نشان‌گذاری شده و برخی از اطلاعات همچون: کلمه مدخل، ریشه و مشتقات ذیل آن، آیات قرآن، روایات و غیره در این لغت‌نامه‌ها مشخص شده‌اند. در حال حاضر، نمونه‌ای از این زحمات در نرم‌افزار رومیزی «قاموس نور ۲» بخش «پژوهش در لغت»، قابل دسترسی است. برای بسیاری از مداخل مرکب، با استفاده از برچسب ریشه و مشتق، مدخل نماینده به صورت ماشینی استخراج شده و در بانک مداخل، به ازای هر مدخل نماینده، چندین مدخل لغوی به همراه توصیف آنها از منابع لغت، قرار گرفته است



هدف از طراحی و تولید «قاموس نور»، به عنوان یک لغت‌نامه هوشمند الکترونیکی، گذر از شیوه‌های سنتی مراجعه به منابع لغوی متعدد و پیاده‌سازی حداکثری الگوهای ماشینی هوشمند جهت ارائه سریع پاسخ‌های صحیح به پژوهشگران است. ساماندهی لغت‌نامه‌های متقدم عربی و نظام‌بندی این منابع ارزشمند بر اساس یک ساختار واحد و همچنین تحلیل صرفی کلمات ورودی کاربران توسط دو موتور قدرتمند تحلیلگر صرفی نور و استمر هوشمند نور جهت اتصال ماشینی به نزدیک‌ترین مداخل لغوی، از دستاوردهای منحصر به فرد این واژه‌نامه هوشمند الکترونیکی است که در قالب پایگاه اینترنتی، نرم‌افزار رومیزی و تلفن‌های همراه قابل استفاده است



نهایی محققان، تعداد مداخل لغوی به حدود سیصد هزار مدخل تقلیل یافته است.

۳. نرمال‌سازی اعراب مداخل نماینده

اگرچه تعیین «مدخل نماینده» در مرحله قبل، موجب تجمیع حجم قابل توجهی از مداخل می‌گردد، اما با توجه به تنوع اعراب در مداخل مفرد و مرکب در کتب مختلف، همچنان امکان تقلیل تعداد «مدخل نماینده» با نرمال‌سازی اعراب آنها وجود دارد؛ مثلاً مدخل نماینده «السَّمْنُ»، در منابع مختلف لغوی موجود به هفت شکل نگارش شده است که تمامی آنها، اشاره به یک کلمه دارد: «السَّمْنُ»، «السَّمْنُ»، «السَّمْنُ»، «السَّمْنُ»، «السَّمْنُ» و «السَّمْنُ». با استفاده از قواعدی هوشمند در مورد اعراب که مشخص‌کننده مداخلی است که امکان تجمیع با سایر مداخل را دارند، اقدام به پردازش ماشینی و تجمیع بیشتر «مدخل نماینده» و انتخاب کامل‌ترین مورد از حیث اعراب به عنوان نماینده شده است. بر اساس این قواعد، دیگر کلمات «حَسَنٌ» و «حُسْنٌ» نرمال نشده و در یک دسته قرار نخواهد گرفت. با انجام این پردازش و پس از بازبینی دستی، تعداد «مدخل نماینده» غیرتکراری، از حدود سیصد هزار عدد به ۱۲۷ هزار مورد تقلیل یافته است؛ برای مثال، یکی از وظایف محققان در بازنگری نتایج، علاوه بر تکمیل اعراب، جداسازی افعال از اسماست؛ مثلاً دو کلمه «بِکْر»

اطلاعات نشان‌گذاری (فرمت‌گذاری) شده در لغت‌نامه‌ها استفاده گردیده است. برخی از لغت‌نامه‌های تایپ‌شده در مرکز نور، طی سالیان متمادی توسط محققان نشان‌گذاری شده و برخی از اطلاعات همچون: کلمه مدخل، ریشه و مشتقات ذیل آن، آیات قرآن، روایات و غیره در این لغت‌نامه‌ها مشخص شده‌اند. در حال حاضر، نمونه‌ای از این زحمات در نرم‌افزار رومیزی «قاموس نور ۲» بخش «پژوهش در لغت»، قابل دسترسی است. برای بسیاری از مداخل مرکب، با استفاده از برچسب ریشه و مشتق، مدخل نماینده به صورت ماشینی استخراج شده و در بانک مداخل، به ازای هر مدخل نماینده، چندین مدخل لغوی به همراه توصیف آنها از منابع لغت، قرار گرفته است. از جمله اقدامات مهم صورت‌گرفته جهت استفاده از برچسب ریشه و یافتن کلمات هم‌ریشه، انجام فرآیند نرمال‌سازی بین ریشه‌های اعلام‌شده در منابع لغوی است. تفاوت در شیوه نگارش همزه و همچنین انجام اعلال در حروف عله ریشه در برخی از منابع لغوی، از جمله مواردی است که با نرمال‌سازی در ریشه نیز اصلاح می‌گردد.

با انجام تمام فعالیت‌های فوق، تعداد قابل توجهی از مداخل تجمیع شدند؛ به عنوان نمونه، در فاز اول که کار استخراج مداخل و توصیفات و بازبینی پنج منبع لغوی با استفاده از علائم ویرایشی، به استخراج نزدیک به هفتصد هزار مدخل منجر گشت، با پردازش ماشینی همچون استفاده از نشان‌ها، شناسایی ریشه و نرمال‌سازی آنها و در نهایت، تعیین مدخل نماینده، پس از بازبینی

و «بکر» توسط نرمال‌سازی ماشینی در یک دسته قرار می‌گیرند که با دقت در توصیف آنها، یکی فعل و دیگری اسم بوده و باید توسط محققان از یکدیگر تفکیک شوند.

۴. پیراسته‌سازی مداخل توسط موتور هوشمند تشخیص میان‌وند

پس از پردازش ماشینی «مداخل نماینده» و تقلیل تعداد آنها، با استفاده از موتور هوشمند تشخیص میان‌وند که از آن به «استمر نور» تعبیر می‌شود، اقدام به پیراسته‌سازی مداخل نماینده می‌گردد. استم (Stem)، یک الگوریتم محاسباتی است که کلمات هم‌ریشه را با تهی نمودن آنها از پیشوندها و پسوندها، به یک شکل مشترک بازمی‌گرداند (Lovins; 1968: 22) که می‌توان به جای آن، از معادل فارسی: «پیراسته»، «میان‌وند» یا «هسته» نیز استفاده نمود. از ساخت استم در زبان عربی، با کلمه «التجذیع» تعبیر می‌شود؛ مثلاً کلمات: «یعلمون»، «سیعلمان»، «یعلمن»، «فیعلمونه» و «أیعلم؟» که همگی از ریشه «علم» هستند، دارای میان‌وند مشترک «یعلم» می‌باشند. با استفاده از استمر نور، مواردی همچون: ضمایر منصوبی و مجروری، نون تأکید، الف پایه کرسی تنوین (مانند تبدیل ءا به ؤ یا مذوماً به مذموم)، علائم مثنی و جمع سالم، پیشوندهایی مانند: حروف جاره، «ال»، «س» استقبال، انواع لام، انواع فاء و غیره حذف می‌گردد. با استفاده از فرآیند ساخت پیراسته، تمامی مداخل لغوی: «أسمنته»، «أسمنتها»، «أسمنه»، «أسمنها» و «أسمنوا»، ذیل

مدخل نماینده «أسمن» قرار خواهند گرفت. این مدخل پیراسته، نماینده ۵ مدخل دیگر است. شایان ذکر است در این مرحله، کلماتی که از حیث معنا کاملاً یکسان هستند، ولی دارای بیش از یک وجه استعمال‌اند، مانند باذنجان، باذنجان و بادنجان، به موردی که بیشترین استعمال و بیشترین تکرار را در بین لغت‌نامه‌ها داراست، ملحق می‌گردند؛ البته با تدوین یک آیین‌نامه اجرایی، پس از پردازش ماشینی توسط استمر نور و تجمیع مداخل نماینده، اقدام به بازبینی دستی نیز ضروری است؛ مثلاً ممکن است در این مداخل، کلماتی که «ال» در آنها موضوعیت داشته، با غیر آن خلط شود؛ برای مثال، کلمه «الجزائر» که علم مکان است، در کنار «جزائر» که جمع مکسر جزیره است، قرار گرفته باشد. این موارد، به یکدیگر ملحق نشده و به عنوان دو مدخل مستقل لحاظ می‌شوند. با پایان مرحله پیراسته‌سازی، تعداد مداخل نماینده از ۱۲۷ هزار به کمتر از ۱۰۰ هزار تقلیل یافته است.

۵. استفاده از تحلیلگر صرفی نور در ایجاد یک نظام اشتقاقی

آنچه سبک نگارش واژه‌نامه‌های جدید عربی را از لغت‌نامه‌های متقدم متمایز نموده، چینش نظام‌مند مداخل لغوی است. معمول لغت‌نامه‌های متأخر، بر اساس یک نظام اشتقاقی ابتدا فعل ماضی، مضارع و سپس مصادر ثلاثی مجرد و در ادامه، ثلاثی مزید و رباعی و اسم‌های مشتق قیاسی و سمعی را به همراه توصیفات آنها ذکر می‌کنند؛ ولی در لغت‌نامه‌های قدیمی، چنین چینشی



نماینده»، طبیعی است که ذیل آن مدخل، تعداد بسیاری عبارات مشابه در کنار هم قرار گیرد؛ مثلاً اگر برای واژه «ضَرْبٌ»، از منابع لغوی متعدد، توصیفات برای پژوهشگر نمایش داده شود، با تعداد فراوانی توضیح مشابه و تنها تفاوت در نشانی مواجه خواهد شد. در اینجا، لازم است نسبت به دسته‌بندی توصیفات متعدد به گونه‌ای اقدام شود که با انتخاب یک «توصیف نماینده»، از نمایش باقی توصیفات مشابه خودداری شده، تنها نشانی‌ها برای پژوهشگر در دسترس قرار گیرد. ذیل هر «توصیف نماینده»، توصیفات قرار می‌گیرند که در عین تفاوت در تعبیر، ناظر به مفهوم واحدی باشند. این دسته‌بندی ماشینی، در دو مرحله صورت می‌پذیرد:

۱. مشابهت‌یابی لفظی بین توصیفات هر مدخل نماینده؛
 ۲. مشابهت‌یابی معنایی بین توصیفات هر مدخل نماینده با استفاده از برخی الگوریتم‌های همنشینی و سیستم عصبی.
- در نهایت، با بازبینی دسته‌بندی‌های ماشینی توسط محققان خبره،

وجه تمایز لغت‌نامه الکترونیکی و هوشمند قاموس نور نسبت به دیگر نرم‌افزارها و پایگاه‌های برخط زبان عربی، در بهره از همین تحلیلگرهای صرفی است. دو تحلیلگر صرفی کلمات (موتور صرف نور و استمر نور)، بر اساس آخرین رویکردهای بهینه‌سازی اطلاعات متنی در مرکز تحقیقات کامپیوتری علوم اسلامی نور طراحی و تولید شده است که با به کارگیری آنها در لغت‌نامه هوشمند نور، مسیر دشوار دسترسی کاربران به معانی لغات زبان عربی، بسیار آسان و سریع گردیده است

رعایت نشده و گاهی فرهنگ‌نگاران توصیف را برای فعل ماضی، گاهی فعل مضارع یا امر و گاهی نیز برای مصدر و اسم فاعل ذکر کرده‌اند. رتبه‌بندی این مداخل با استفاده از یک تحلیلگر صرفی هوشمند، یعنی «موتور صرف نور» صورت می‌پذیرد. موتور صرف ابتدا کلمه را از لحاظ صرفی، تجزیه و تحلیل می‌نماید و سپس، این کلمات را بر مبنای یک نظام اشتقاقی مدون، در بانک مداخل مرتب می‌کند؛ مثلاً ابتدا افعال ماضی ثلاثی مجرد، بعد مضارع و مصادر و دیگر مشتقات قیاسی و غیره ذیل یکدیگر قرار می‌گیرند.

یکی از نتایج خروجی تحلیلگر صرفی نور، لَمای (Lemma) کلمات است که از آن، به «بن» یا «ستاک» کلمه نیز تعبیر می‌شود. مقصود از «لَمّا» در کلمات، یک فرم متعارف حداقلی از کلمه است که دارای معنا بوده و به عنوان مدخل ورودی در فرهنگ لغت نیز یافت می‌شود (Boudchiche and Maz-roui; 2018: 2)؛ مثلاً لَمّا از فعل ثلاثی مجرد «يَضْرِبُونَ»، فعل ماضی «ضَرْبٌ» و از فعل ثلاثی مزید «يُعْلِمُونَ»، فعل «أَعْلَمَ» است. اولین موتور هوشمند تولید لَمّا از کلمات عربی، در سال ۲۰۰۸ معرفی گردید (Al-Shammari, Lin; 2008). ثمره تشخیص «لَمّا» در کلمات، این است که می‌توان به واسطه آن، ارتباط کاملی میان مداخل لغوی ایجاد کرده و آنها را در دسته‌های مختلفی قرار داد. بدین ترتیب، هرگاه پژوهشگری در جست‌وجوی یک کلمه ماضی باشد، ولی مؤلف یک لغت‌نامه، آن را ذیل یک مضارع یا فعل امری توصیف کرده باشد، می‌توان تمامی آنها را بر اساس یک چپش نظام‌مند در اختیار پژوهشگر قرار داد؛ زیرا این کلمات، دارای لَمّا و بن مشترک بوده و طبق نتایج موتور صرف نیز رتبه‌بندی شده‌اند. با توجه به امکان اشتباه در تجزیه صرفی برخی کلمات توسط موتور صرف، به‌ویژه لغات معتل و معرب، نتایج خروجی نیازمند بازبینی محققان است؛ مثلاً با توجه به اختلاف لغت‌نامه‌ها در مورد ریشه کلمات، باید در الحاق کلمات معتل به صیغه اول ماضی، به حروف عله کاملاً دقت شود. اگر در یک لغت‌نامه، یک ماضی با ریشه «دعو» آمده و همان کلمه با همان معنا به صورت مضارع در لغت‌نامه‌ای دیگر با ریشه «دعی» آمده، در نظام اشتقاقی به یکدیگر مرتبط می‌شوند؛ برخلاف جایی که آن دو کلمه، دارای یک مفهوم مشترک نبوده و از هر ریشه، یک معنای متفاوت اراده شده باشد.

۶. دسته‌بندی توصیفات ذیل هر مدخل از حیث مفهوم با توجه به تشابه بسیاری از واژه‌نامه‌ها در ارائه توضیحات یکسان برای کلمات، در صورت تجمیع مداخل لغوی ذیل یک «مدخل

445: 2013; Abdulghani; na). یک لغت‌نامه هوشمند چندزبانه و منسجم از منابع غنی زبان‌های مختلف، می‌تواند بهترین راهکار در ارتقای دانش واژگان کاربران و فعالیت‌های هوشمند متن‌کاوانه باشد. هم‌اکنون استفاده از فرهنگ غنی دهخدا نیز جهت ارائه پاسخ برای کلمات فارسی و همچنین کلمات عربی وارد شده در این زبان آغاز گردیده است که با توسعه موتورهای تحلیلگر صرفی کلمات در زبان فارسی، دستیابی به نتایج مطلوب با رشد بیشتری همراه خواهد شد.

تحلیل صرفی کلمات ورودی توسط ماشین‌های هوشمند

گستره وندها (پیشوند و پسوند) در برخی زبان‌ها همچون انگلیسی و فرانسوی، محدود بوده و تغییرات چندانی در کلمات ایجاد نمی‌کند؛ مانند بازگشت کلمات «engineered» و «-engi» و «neering» به مدخل «engineer» یا بازگشت کلمه «-Ski ing» به مدخل «Sky» که با حذف پسوند «-ing»، نهایت با یک تغییر در حروف اصلی همراه خواهد بود. در بسیاری از لغت‌نامه‌های انگلیسی، معمول این کلمات در فهرست مداخل قرار گرفته و به چرخه تحلیل هوشمند صرفی کلمات وارد نمی‌شوند؛ اما در مورد زبان عربی که گستره استعمالی وندها دایره وسیع‌تری نسبت به دیگر زبان‌ها دارد و همچنین به سبب قواعد اعلال، ادغام و تخفیف، کلمات عربی دست‌خوش تغییرات بیشتری می‌شوند و عملاً احصای تمام کلمات عربی و اتصال آنها به یک مدخل لغوی توسط نیروهای انسانی، محال می‌نماید. از این‌رو، تنها

توصیف برگزیده برای هر دسته از مفاهیم یک مدخل مشخص می‌گردد. پس از دسته‌بندی توصیفات، می‌توان در هر دسته به یک توصیف برگزیده رسید که آن نیز می‌تواند از میان توصیفات موجود در آن دسته انتخاب شود و یا توسط پژوهشگران مرکز توصیفی که بیانگر مفاهیم موجود در دسته است، تدوین و عرضه گردد. این کار، به مخاطب در دسترسی سریع‌تر به مقصد خود کمک می‌نماید؛ هرچند امکان دسترسی به توصیفات متنوع ذکر شده در منابع مختلف نیز برای وی فراهم است.

۷. تفکیک زبانی

از آنجاکه میان برخی از لغت‌نامه‌های ساماندهی شده در بانک مداخل، واژه‌نامه‌های دوزبانه عربی فارسی همچون فرهنگ ابجدی نیز وجود دارد، امکان تفکیک توصیفات بر اساس نوع زبان نیز مهیا گردیده است. در واقع، استفاده از سه عنصر: مدخل لغوی، توصیف عربی و ترجمه فارسی و کنار هم قرار گرفتن آنها، الگویی را ایجاد خواهد کرد که می‌تواند کمک شایانی به سیستم‌های هوشمند متن‌کاوی همچون موتورهای ترجمه و ماشین‌های مشابهت‌یاب معنوی داشته باشد. دانش واژگان، مهم‌ترین مؤلفه هر زبان برای فراگیران و مترجمان است و بیشترین خطای پیش‌آمده برای پژوهشگران انسانی و ماشین‌های ترجمه نیز در همین ناحیه است (Al-sohbbani Yehia Ahmed and Muthan-)



تصویر شماره ۲: نمونه‌ای از ابزار واژگان استفاده شده در مرکز نور

راه حل پیش رو، استفاده از ماشین‌های هوشمند تحلیلگر صرفی کلمات عربی است.

وجه تمایز لغت‌نامه الکترونیکی و هوشمند قاموس نور نسبت به دیگر نرم‌افزارها و پایگاه‌های برخط زبان عربی، در بهره از همین تحلیلگرهای صرفی است. دو تحلیلگر صرفی کلمات (موتور صرف نور و استمر نور)، بر اساس آخرین رویکردهای بهینه‌سازی اطلاعات متنی در مرکز تحقیقات کامپیوتری علوم اسلامی نور طراحی و تولید شده است که با به‌کارگیری آنها در لغت‌نامه هوشمند نور، مسیر دشوار دسترسی کاربران به معانی لغات زبان عربی، بسیار آسان و سریع گردیده است. ویژگی منحصر به فرد قاموس نور، این است که می‌تواند ریشه، استم (میان‌وند) و لمای (بن) هر کلمه عربی را با استفاده از موتور قدرتمند استمر نور یا موتور صرف نور تشخیص دهد و پژوهشگر را به نزدیک‌ترین مدخل نماینده متصل نماید. مراحل تحلیل صرفی کلمات ورودی، بدین شرح است:

در مرکز علوم تحقیقات کامپیوتری
علوم اسلامی نیز طراحی و ساخت
یک موتور هوشمند برچسب‌گذار
کلمات عربی در لایه صرف، در
سال‌های پیشین در دستور کار
قرار گرفت (سریانی، مینایی،
۱۳۹۰) و با ارتقای آن در سال‌های
بعدی (دانش، ۱۳۹۳)، قدم‌های
مهمی را در تجزیه صرفی کلمات
عربی طی نمود

۱. تشخیص ریشه با استفاده از داده برچسب‌گذاری شده در مرکز نور

بنا بر آمار به‌دست‌آمده از مجموع متون عربی موجود در مرکز تحقیقات کامپیوتری علوم اسلامی (بیش از شش هزار عنوان کتاب)، نزدیک به دو میلیون و هفتصد هزار کلمه غیرتکراری در این متون استخراج شده است. تعداد تقریبی کل این دیتا - تا به امروز - یک میلیارد و پانصد میلیون کلمه بوده و بیش از ۹۹ درصد کلمات غیرتکراری و بدون اعراب این داده ارزشمند، توسط محققان مرکز نور تعیین ریشه و برچسب‌گذاری شده است. اگر کلمه دارای ریشه‌ای معتبر در لغت‌نامه‌های متقدم و متأخر باشد، آن ریشه متناسب با استعمال آن لغت انتخاب شده است؛ مانند کلمه «فتاه» که این کلمه در داده مرکز نور با در نظر گرفتن همه حالات اعرابی و نوشتاری، ۱۲۸۵ بار تکرار شده است. این کلمه در برخی متن‌ها، از ریشه «فتو» به معنای «جوان خدمتگزارش» و در برخی دیگر، از موارد استعمال به معنای «تکبر کرد/گمراه شد»، از ریشه «تیه» است. این دو ریشه، توسط محققان برای کلمه «فتاه» در نظر گرفته شده است. در صورت نداشتن ریشه، برچسب‌های دیگری مانند: برچسب، علم، دخیل و معرب، غیرعربی، دخیل در فارسی و... برای کلمات انتخاب شده است؛ مثلاً کلمه «النيسابوری» با ۳۴۸۱۹ بار تکرار، دارای برچسب معرب است که در زبان عربی ریشه‌ای ندارد. گفتنی است که انتخاب ریشه توسط محققان مرکز

تحقیقات نور، با نگرش استعمالی صورت گرفته؛ نه نگرش منطقی. برای مثال، در همان نمونه «فتاه» که در تصویر شماره ۲ آمده است، می‌توان با منطق صرفی، ریشه «فتت» را نیز در نظر گرفت؛ یعنی «آن دو نفر آن چیز را ریز ریز کردند»؛ اما در میان ۱۲۸۵ مورد از تکرار این کلمه، مواردی که از ریشه «فتت» باشد، یافت نشده و در نتیجه، پژوهشگر نیز به ریشه «فتت» در لغت‌نامه‌ها منتقل نخواهد شد. مثال دیگر، کلمه «لبناتک» که در تمام ۱۲۵ مورد تکرار آن، موردی که بر اساس ریشه «لبن» باشد نیز یافت نشده است.

بدیهی است با داشتن این داده ارزشمند، دیگر طراحی یک موتور هوشمند تشخیص ریشه همچون Khoja and (Khoja and Garside; 1999) که در بسیاری از سیستم‌های بازبایی اطلاعات مورد استفاده واقع شده، امری بیهوده خواهد بود. در واقع، به واسطه تلاش پژوهشگران مرکز تحقیقات نور و با بهره از فعالیت‌های ماشینی، برای غالب کلمات، ریشه صحیح بر اساس استعمال صحیح انتخاب شده و می‌توان برای کلمات ورودی کاربران، به ریشه دست یافت؛ البته با استفاده از موتورهای استمر و تحلیلگر صرف، همواره توانایی تشخیص ریشه برای کلمات جدید، به‌ویژه کلمات متون عربی معاصر که در آن داده یک‌ونیم

میلیاردی وجود ندارند، مهیاست.

۲. تشخیص میان‌وند کلمات توسط استمر هوشمند نور

نتایج تحقیقات نشان داده که تجزیه صرفی کلمات با استفاده از الگوی «استم»، بیش از الگوهای دیگر همچون تشخیص ریشه و لَمَّا (بن)، در فرآیندهای هوشمندسازی اثربخش بوده است (Sembok and Abuata; 2013: 1577). شکستن کلمات یک متن عربی به کلمات ساده و قابل تحلیل صرفی، از مهم‌ترین فواید ساخت استم در زبان پیچیده عربی است. تمامی سیستم‌های مورد استفاده در متن کاوی همچون سیستم جست‌وجو، طبقه‌بندی موضوعی متون، خلاصه‌ساز، خوشه‌بندی متون، مشابه‌یاب لفظی و معنایی و سیستم‌های دیگری که وابسته به تجزیه صرفی کلمات‌اند، می‌توانند از استم به عنوان یک الگوی کاربردی در پیش‌پردازش‌ها و ارتباط کلمات استفاده نمایند. از موتور استمر نور، جهت تشخیص میان‌وند کلمات و تطبیق آن بر نزدیک‌ترین مدخل لغوی مشابه استفاده می‌شود.

طراحی یک استمر عربی، با چالش‌های فراوانی روبه‌روست که جهت رفع هریک از آنها، از قواعد دقیق و کاربردی و داده‌های آماده‌شده توسط محققان، استفاده گردیده است؛ مثلاً تجزیه کلمه «والده» (پدر او) با استفاده از موتور استمر Light Stemmer (Larkey et al; 2007) که یک استمر جهانی و سبک برای حذف وندهاست، با در نظر گرفتن «و + ال» به عنوان پیشوند و «ه» به عنوان پسوند، استم نادرست «د» را به عنوان خروجی اعلام می‌کند؛ درحالی‌که «وال» جزئی از استم صحیح والد است. همچنین استم در مورد کلمه «الم»، «فتح» و دیگر موارد مشابه که با استفاده از قواعد دقیق در مراحل کدنویسی استمر نور، از حذف نادرست پیشوند و پسوندها جلوگیری شده است.

چالش دیگر موتورهای استمر، این است که یا تنها یک پاسخ به عنوان خروجی اعلام می‌دارند و یا اینکه چندین جواب غیرمستعمل نیز برای کلمات اعلام می‌کنند؛ مثلاً برای کلمه قرآنی «لمسنا» (الجن/۸)، در حالت بدون اِعراب می‌توان سه استم منطقی: «لمس»، «مس» (از ریشه مسس) و «مسن» (از ریشه سنن) را در نظر گرفت؛ درحالی‌که بسیاری از موتورهای استمر عربی موجود، تنها یک راه حل را در خروجی استم خود ارائه می‌دهند و یا مانند برخی استمرهای پیشرفته‌تر (Jaafar et al; 2017)، هر سه استم را به عنوان پاسخ صحیح اعلام می‌کنند؛ حال آنکه کلمه «لمسنا»، بر پایه استم «مسن» به هیچ‌وجه استعمال ندارد. با استفاده از برچسب‌های ریشه که با نگرش استعمالی (نه نگرش منطقی صرفی) انجام شده و استفاده از الگوریتم‌های بررسی

استعمال و کاربرد کلمات، تا حد بسیار مطلوبی این چالش در موتور استمر نور مرتفع گردیده است. از این‌رو، پژوهشگر با مراجعه به قاموس هوشمند نور، ابتدا با پاسخ‌های مربوط به «لَمَس» و سپس پاسخ‌های مربوط به «مَس» در منابع لغوی مواجه شده و دیگر به ریشه «مسن» و مشتقات آن منتقل نخواهد شد. مسئله تغییر کلمات به واسطه قواعد اعلال و ادغام و تشخیص مفرد برای جمع‌های مکسر، از دیگر مسائلی است که در استمر نور بررسی و برای آن راهکاری ارائه شده است تا پژوهشگر در صورت مطالعه بر یک جمع مکسر و معنای لغوی آن، به‌راحتی به واژه مفرد آن نیز در منابع لغوی دست یابد.

۳. تشخیص نوع کلمه توسط موتور تحلیلگر صرفی نور

در کنار استمرها، دسته‌ای از ماشین‌های هوشمند تحلیل‌گر صرفی نیز طراحی شده‌اند که علاوه بر تشخیص ریشه و استم، تمرکز اصلی خود را بر برچسب‌گذاری صرفی کلمات عربی همچون تعیین وضعیت نوع کلمه از لحاظ اسم، فعل و حرف و نوع هریک از آنها نهاده‌اند؛ برای مثال، ماشین هوشمند تحلیل صرفی BAMA (Buckwalter 2007)، از پیشگامان این عرصه

بررسی نتایج تحلیلگر صرفی
«نور» و مقایسه آن با برخی دیگر
از تحلیلگرهای صرفی، حاکی از
روند رو به رشد پاسخ‌های مطلوب
در این موتور هوشمند داشت
(الهی منش، ۱۳۹۴: ۱۷). شیوه
عملکرد موتور صرفی، جهت تعیین
نوع یک کلمه مبتنی بر تشخیص
هوشمند تمامی حروف اصلی و
زاید بوده و تنها به تشخیص پیشوند
و پسوند کلمه اکتفا نمی‌شود

شناخته می‌شود. از شاخص‌ترین ماشین‌های تحلیل‌گر هوشمند متون عربی، می‌توان به الخلیل ۲ (Boudchiche; 2016) اشاره کرد. این موتور، در راستای ارتقای الخلیل ۱ (Boudlal et al; 2010) که یک تحلیلگر صرفی منبع باز بوده و توسط تیم NLP دانشگاه حضرت محمد(ص) مراکش طراحی گشت، ساخته شد. این موتور، تمام حالات صرفی متصور برای یک کلمه عربی را به همراه ریشه، اِستم، لِمَا و برخی برچسب‌های صرفی مانند نوع کلمه، اعلام می‌داشت. موتور تحلیلگر صرفی -MAD AMIRA که توسط محققان بخش NLP دانشگاه کلمبیا در سال‌های اخیر طراحی و معرفی گردید (Pasha; 2014)، به همراه موتور هوشمند دیگری که در حال حاضر به صورت برخط و آفلاین در اختیار محققان قرار گرفته، یعنی تحلیلگر هوشمند SAFAR از فعالان این عرصه به شمار می‌روند.

در مرکز علوم تحقیقات کامپیوتری علوم اسلامی نیز طراحی و ساخت یک موتور هوشمند برچسب‌گذار کلمات عربی در لایه صرف، در سال‌های پیشین در دستور کار قرار گرفت (سربانی، مینایی، ۱۳۹۰) و با ارتقای آن در سال‌های بعدی (دانش، ۱۳۹۳)، قدم‌های مهمی را در تجزیه صرفی کلمات عربی طی نمود. بررسی نتایج تحلیلگر صرفی «نور» و مقایسه آن با برخی دیگر از تحلیلگرهای صرفی، حاکی از روند رو به رشد پاسخ‌های مطلوب در این موتور هوشمند داشت (الهی‌منش، ۱۳۹۴: ۱۷). شیوه عملکرد موتور صرف، جهت تعیین نوع یک کلمه مبتنی بر تشخیص هوشمند تمامی حروف اصلی و زاید بوده و تنها به تشخیص پیشوند و پسوند کلمه اکتفا نمی‌شود. مجموعه حروف زاید، عبارت از ۱۰ حرفی است که در کلمه «سألتومنیها» جمع آمده و به واسطه برخی قواعد ماشینی در کلمات عربی قابل شناسایی هستند. تشخیص حروف اصلی در کلماتی که در معرض قواعد اعلال همچون حذف و قلب واقع شده‌اند، از نقاط قوت موتور صرف نور است. در کنار کدنویسی پیشرفته‌ای که در راستای شناسایی و تحلیل صرفی کلمات معتل، مضاعف و مهموز صورت گرفته، از تعداد داده آماده که توسط محققان بخش متن‌کاوی مرکز نور فراهم آمده نیز استفاده شده است. موتور صرف نور، برای تحلیل کلماتی که سماعی هستند، از تعدادی بانک جانبی استفاده می‌کند که از آن جمله، می‌توان به بانک‌های اطلاعاتی مجموعه افعال غیرمتصرف (حدود ۷۰ رکورد)، حروف (حدود ۱۳۰ رکورد)، جوامد (حدود ۱۵۰۰۰ رکورد)، جمع‌های مکسر (بیش از ۱۸۰۰۰ رکورد)، صفات مشابهه (بیش از ۷۵۰۰ رکورد) و مصادر ثلاثی مجرد (بیش از ۱۰۰۰۰ رکورد) اشاره نمود.

با استفاده از موتور صرف نور، علاوه بر ریشه، اِستم و لِمَا (بن کلمه)، نوع صرفی آن نیز از لحاظ اسم، فعل یا حرف مشخص شده و به همراه دیگر جزئیات صرفی، در فرآیند هوشمند اتصال به کلمات بانک مداخل مورد استفاده قرار می‌گیرد.

فرآیند اتصال کلمات ورودی به مداخل لغت‌نامه‌های عربی
از یک سو، با ایجاد «بانک مداخل» که داده‌ای منسجم از مداخل لغوی و توصیفات مندرج در واژه‌نامه‌های معتبر بوده و با چینیسی نظام‌مند بر اساس «مدخل نماینده» و «توصیف نماینده» مرتب شده و از سوی دیگر، با تحلیل هوشمند صرفی کلمات ورودی پژوهشگران که به تشخیص ریشه، اِستم (میان‌وند) و لِمَا (بن) کلمات منجر شده، تنها فرآیند باقی‌مانده، اتصال ماشینی کلمه مورد نظر پژوهشگران به نزدیک‌ترین مدخل نماینده از حیث لفظ و معنا است. با دقت نظر در اقدامات ذهنی و عملی یک پژوهشگر در هنگام مراجعه به منابع لغوی، می‌توان مراحل دستیابی به پاسخ مطلوب را در فرآیندهای ماشینی چنین رتبه‌بندی نمود:

۱. نمایش مداخلی که با کلمه یا عبارت ورودی کاربر تطابق داشته یا اختلاف آنها یک «ال» باشد

مداخل لغوی که بر اساس فرآیندهای پیراسته‌سازی و تحلیل صرفی ذیل یک مدخل نماینده و یک لِمَا مشترک قرار گرفته‌اند، گاهی با همان حالت صرفی خود با کلمه ورودی کاربر تطابق می‌یابند؛ مثلاً کلمه «مُؤَلَّها» - که اشاره به آیه ۱۴۸ سوره بقره داشته و قرائت دیگری از «مولیها» است - در تاج العروس (همان؛ ج ۲۰: ۳۱۳) به صورت مدخل وارد شده و در توصیف آن آمده است: «أی الله تعالی یولی أهل كل ملة القبله التي تُربد.» با توجه به فرآیند پیراسته‌سازی در بانک مداخل، این کلمه ذیل مدخل نماینده «مولى» قرار می‌گیرد و با توجه به تجزیه صرفی و تشخیص لِمَا آن، در دسته «ولی / یولی / تولیة» قرار خواهد گرفت. اگر پژوهشگر کلمه «مُؤَلَّها» را جست‌وجو نماید، دقیقاً عبارت تاج العروس به عنوان اولین پاسخ برای او نمایش داده خواهد شد. پس از آن، دیگر مداخل لغوی که ذیل مدخل نماینده «مولى» هستند و پس از آن، تمامی مداخلی که در دسته «ولی / یولی / تولیة» قرار گرفته و دارای لِمَا مشترک با کلمه «مُؤَلَّها» هستند و در مرحله آخر، مداخل دارای ریشه «ولی» در یک نظام اشتقاقی مرتب‌شده در اختیار کاربر قرار خواهد گرفت.

گفتنی است، اگر عبارت ورودی کاربر بیش از یک کلمه باشد،



اگر پژوهشگری در جست‌وجوی مفهوم «ولیت‌موه» در عبارت «لَمَنْ وَلِيْتُمُوهُ أَمْرُكُمْ» از نهج البلاغه (رضی، همان: ۱۳۶) باشد، از آنجاکه این کلمه به همین شکل در بانک مداخل وجود ندارد، این کلمه پیراسته شده و به صورت «ولی» میان مداخل نماینده جست‌وجو می‌گردد. مدخل نماینده «ولی» در بانک مداخل، دارای مداخلی چون: «ولیت»، «ولیتک»، «ولیتم» و «ولیتن» بوده و تمامی این موارد، همراه با توصیفات به عنوان پاسخ نخست برای کاربر ارائه می‌گردد.

نکته قابل توجه اینکه در صورت بهره بیشتر از نتایج خروجی ابزارهای هوشمند تحلیل صرفی، می‌توان مداخل را متناسب با کلمه ورودی کاربر به شکلی هوشمند مرتب نمود؛ مثلاً اگر کلمه کاربر دارای پسوند مذکر (مانند همان کلمه «ولیت‌موه») است، ابتدا مداخل مذکر (یعنی «ولیت»، «ولیتک» و «ولیتم») و سپس مداخل مؤنث (یعنی «ولیتن») را، برای او نمایش داده شود.

۳. نمایش مداخلی که «مدخل نماینده» آنها با بِن (لَمَّا) کلمه ورودی کاربر تطابق دارد

اگر کلمه ورودی و استم آن با هیچ‌یک از مداخل پسر یا مداخل نماینده تطابق نداشته باشد، از لَمَّا کلمه که توسط موتور تحلیلگر صرفی نور شناسایی گردیده، استفاده شده و نزدیک‌ترین مدخل نماینده مرتبط به آن کلمه، در بانک مداخل جست‌وجو می‌شود. سپس، مداخل دیگری که ذیل یک لَمَّا مشترک قرار دارند، با حفظ نظام رتبه‌بندی خود برای کاربر نمایش داده می‌شوند؛ برای نمونه، اگر کاربری جهت فهم معنای عبارت مستدرک الوسائل

همین فرآیند جاری خواهد بود؛ یعنی اگر عبارت جست‌وجوشده با مدخل مرکبی از میان مداخل لغوی موجود در واژه‌نامه‌ها، تطابق داشته باشد، همان مدخل مرکب به عنوان پاسخ نخست در اختیار کاربر قرار خواهد گرفت؛ برای مثال، در تاج العروس (همان)، برای عبارت «ولی عنه» آمده است: «ولی عنه: أَيْ أُعْرَضُ أَوْ نَأَى». اگر عبارت جست‌وجوشده توسط کاربر نیز «ولی عنه» باشد، دقیقاً همین توصیف از تاج العروس پاسخ نخست خواهد بود؛ در غیر این صورت، هریک از کلمات آن عبارت، به صورت مجزا در فرآیندهای تحلیل صرفی هوشمند قرار گرفته و برای تمامی آنها ریشه، استم و لَمَّا تولید می‌شود. در مرحله آخر نیز از همنشینی و ترکیب چند استم یا چند لَمَّا نیز می‌توان به نزدیک‌ترین پاسخ دست یافت؛ مثلاً اگر عبارت مورد پژوهش کاربر، عبارت «ولیت عَنِّهَا» از نهج البلاغه باشد (رضی، ۱۴۱۴: ۴۱۸) باشد، پیراسته کلمات آن به صورت «ولی عن» بوده و به همان مدخل تاج العروس، یعنی «ولی عنه» مرتبط خواهد شد.

۲. نمایش مداخلی که «مدخل نماینده» آنها با پیراسته (استم) کلمه ورودی کاربر تطابق دارد

اگر مطابق کلمه یا عبارت ورودی کاربر، مداخلی در میان بانک مداخل یافت نشود، به سراغ تطابق میان استم‌ها خواهیم رفت. استم مطلوب، آن میان‌وندی خواهد بود که در خروجی موتور استم و موتور صرف، هر دو یکسان باشد. در صورت تطابق پیراسته کلمه ورودی کاربر با یکی از مداخل نماینده، تمام مداخل ذیل آن (یعنی مداخل پسر)، در اختیار پژوهشگر قرار خواهد گرفت؛ برای مثال،

در «هَلْ يَنْحَابُونَ وَ هَلْ يَنْتَاصِحُونَ» (نوری، ۱۴۰۸: ج ۸، ۳۱۰) بر کلمه «یتناصحون» کلیک نماید، از آنجا که عین این کلمه و پیراسته آن (یتناصح)، در میان مداخل لغوی و مداخل نماینده وجود ندارد، با استفاده از بُن کلمه، مدخل نماینده «تناصح» شناسایی شده و مداخل ذیل آن، مانند «تَنَاصِحُوا: با یکدیگر نصیحت کردند»، از لغت‌نامه مقدمه الأدب (زمخشری، ۱۳۸۶: ۲۶۴) به عنوان پاسخ‌های نخست برای کاربر عرضه شده و پس از آن، دیگر مداخل مشترک در یک دسته از لَمَّا عرضه خواهد شد.

۴. نمایش مداخل هم‌ریشه

در صورت عدم تطابق مداخل نماینده با پیراسته یا بُن کلمات ورودی یا فقدان تشخیص پیراسته یا بُن کلمات، تمامی مداخلی که دارای ریشه مشترک با کلمه ورودی کاربر هستند، با همان چینش نظام اشتقاقی مدون در بانک مداخل برای کاربر عرضه خواهد شد. بیان این نکته خالی از لطف نخواهد که کلماتی که تنها به واسطه ریشه به مداخل متصل شده‌اند، به عنوان بازخوردی از عملکرد ابزارهای هوشمند تحلیلگر صرفی لحاظ شده و جهت رفع نقایص، به بخش متن کاوی گزارش می‌گردد.

غنی‌سازی محتوای مداخل لغوی

با فراهم شدن زیرساخت مناسب در لغت‌نامه هوشمند قاموس نور و محوریت یافتن «مداخل نماینده»، این امکان فراهم شده است تا شبکه‌ای از ارتباطات بین این مداخل بر اساس رویکردهای متفاوت شکل بگیرد. بی‌تردید، یک گام مهم در استفاده از پایگاه داده‌ها و استنتاج متون جهت استخراج ارتباطات و اطلاعات، بهره‌گیری از شبکه واژگانی (وردنت) است (Rajman and

Besancon; 1998: 3). استفاده از شبکه واژگان، نه تنها در متن کاوی و تحلیل داده‌ها، بلکه در یادگیری و آموزش نیز کاربرد فراوان دارد. به‌طور کلی، استفاده از دانش واژگان، در یادگیری و خواندن متون زبان‌های خارجه ضروری است (Chen and Hsu; 2008: 153-180) معلمان زبان، از دو مهارت جهت درک معنای کلمات ناآشنا برای آموزش شاگردان خود استفاده می‌کنند: ۱. مهارت حدس زدن (به عنوان مثال، توانایی استفاده از قرائن متنی و ساختاری برای تحصیل معنای صحیح کلمه همچون هم‌خانواده‌ها)؛ ۲. توانایی استفاده مناسب از فرهنگ لغات (به عنوان مثال، ارجاع سریع به صفحه‌ای که شامل کلمه مورد نظر بوده و خواندن توضیحات مربوط) (See Sun et al.; 2011). با پیشرفت و گسترش فرهنگ لغات الکترونیکی و نرم‌افزارهای مبتنی بر داده، زبان آموزان نیز با کاهش زمان جست‌وجو، دسترسی کاملی به دانش واژگان پیدا کرده‌اند. فراتر از فرهنگ لغت، با استفاده از فرآیند ماشینی، امکان اطلاعات زبانی بیشتری مانند ترجمه ماشینی که در گذشته مهم‌ترین انگیزه طراحی این شبکه واژگانی بوده است، ابهام‌زدایی واژگان (WSD)، طبقه‌بندی اسناد، طراحی هستان‌شناسی در وب و همچنین، غنی‌سازی محتوای لغت‌نامه‌ها مانند: تعداد تکرار کلمات، تنوع استعمال و چگونگی ساختارهای گرامری و بسیاری دیگر از فواید مرتبط با متن کاوی در علوم اسلامی، امکان‌پذیر خواهد شد (سریانی، ۱۳۹۵: ۴۸-۴۹).

استفاده از برچسب‌های مختلف در شناسایی معنای یک کلمه، می‌تواند آن را در حلقه‌های ارتباطی مختلف از کلمات قرار دهد. در مباحث وردنت (شبکه واژگان)، اولین و اصلی‌ترین ارتباط معنایی کلمات، بین مترادفات است. از کلمه‌ای که محور قرار گرفته و باقی مترادفات و کلمات، با آن حلقه ارتباطی تشکیل می‌دهند، به

با فراهم شدن زیرساخت مناسب در لغت‌نامه هوشمند قاموس نور و محوریت یافتن «مداخل نماینده»، این امکان فراهم شده است تا شبکه‌ای از ارتباطات بین این مداخل بر اساس رویکردهای متفاوت شکل بگیرد. بی‌تردید، یک گام مهم در استفاده از پایگاه داده‌ها و استنتاج متون جهت استخراج ارتباطات و اطلاعات، بهره‌گیری از شبکه واژگانی (وردنت) است

دارند. از جمله این کتاب‌ها، می‌توان به کتاب المعجم المفصل فی المتضادات (طراد، ۲۰۱۰) اشاره نمود که به سبب نگارش متأخر آن، دارای نگارش منظم و بهتری است. اطلاعات این کتب نیز پس از پردازش ماشینی، در بانک متضادات لغت‌نامه قرار گرفته و متناسب با برخی از پاسخ‌های لغت‌نامه، موارد متضاد با پاسخ نیز در اختیار کاربر قرار می‌گیرد. با نظر به گستره وسیع مفهوم متضاد که شامل هر سه نوع ارتباط تلازم (مانند مفهوم بدهکار و طلبکار) تتمیم و تکمیل (تناقض) و تضاد منطقی است، به تضاد در معنای عام آن اکتفا شده و باید در صورت ارتقا و بررسی محققان خبره، امکان ایجاد شبکه‌ای برای تفکیک تمامی انواع متضاد و مناسب با زبان عربی مهیا شود.

۳. اضداد

رابطه اضداد، یعنی اینکه یک کلمه دارای دو معنای متضاد یا مخالف به معنای اعم باشد. این تقابل دو معنا، گاهی از اختلاف برخی لغویون نیز ناشی می‌گردد؛ مثلاً ابن‌انباری (۱۹۶۰: ۲۲) برای فعل «عسی» دو معنای متضاد ذکر کرده است؛ یکی به معنای شک و طمع، و دیگری به معنای یقین. کتاب المعجم المفصل فی الأضداد (بطرس، ۲۰۰۳)، از جمله کتاب‌های نگارش متأخر است که زحمات گذشتگان را به شکلی مطلوب ساماندهی نموده و در فرآیندهای شبکه‌وارگانی مرکز نور مورد استفاده قرار گرفته است.

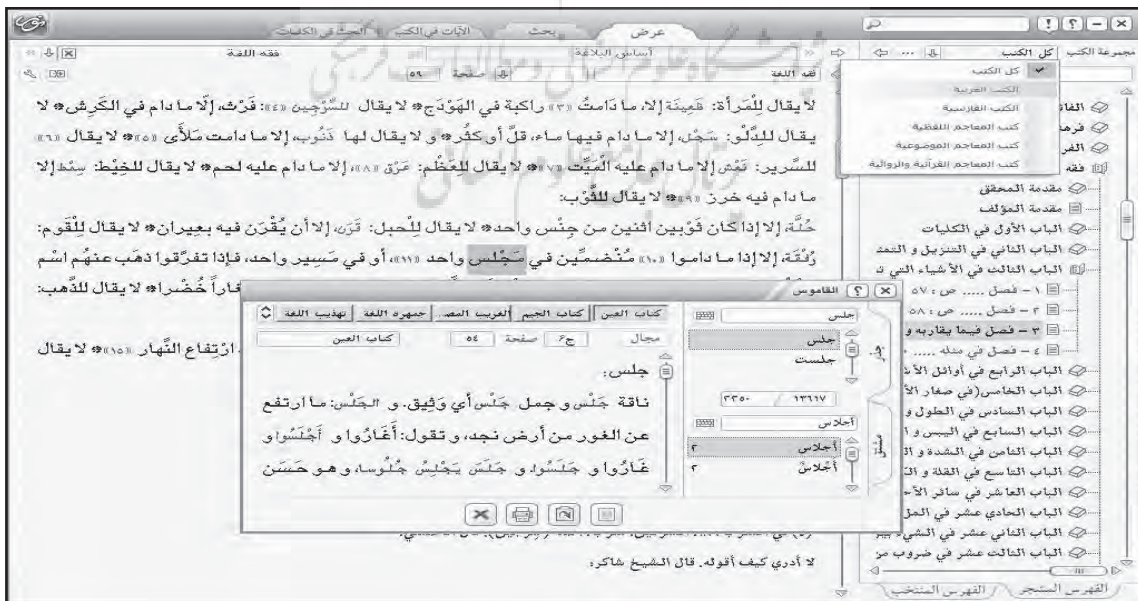
عنوان Lemma یاد شده و حلقه‌های ارتباطی حول یک کلمه را Synset می‌نامند. قاموس نور نیز به عنوان یک لغت‌نامه هوشمند، با استفاده از داده‌هایی که در اختیار داشته، اقدام به تولید حلقه‌های ارتباطی و ارائه برخی اطلاعات در حوزه شبکه‌وارگانی در قالب مترادفات، متضادات و مرتبطات نموده و از سوی دیگر، با غنی‌سازی محتوی به واسطه ارائه توصیفات مربوط به ریشه، غریب الحدیث و برخی فرهنگ اصطلاحات، اطلاعات مفیدی را در اختیار پژوهشگران قرار می‌دهد.

۱. مترادفات

ترادف، رابطه بین دو مفهوم معادل یا نزدیک است که یکی قابلیت جانمایی دیگری را داشته باشد. از این رو، رابطه ترادف، یک رابطه دوسویه و متقارن است؛ مانند سه واژه: فهم، درک و علم. در برخی از منابع همچون المکنز (صینی، ۱۴۱۴) و المعجم المفصل فی المترادفات (طراد، ۲۰۰۹)، مؤلف به بیان مداخل مترادف پرداخته است. با پردازش محتوای این کتب و برقراری ارتباط ترادف بین مداخل ذکرشده، هنگام ارائه پاسخ به کاربر، مداخل مترادف با درخواست کاربر ارائه گشته و امکان انتقال بین این مداخل نیز فراهم می‌گردد. باید توجه داشت، در فرآیندهای ماشینی، برخلاف مراجعه به کتاب فیزیکی، امکان ایجاد رابطه دوسویه کاملاً محقق است.

۲. متضادات

تضاد، رابطه بین دو مفهوم مخالف است. تعدادی از منابع به جمع‌آوری و ارائه مداخلی پرداخته‌اند که رابطه تضاد با یکدیگر



۴. مرتبطات

«مرتبطات»، عنوان عامی است که حتی شامل «مترادفات» و «متضادات» هم می‌شود؛ ولی آنچه در اینجا مقصود است، مداخلی هستند که حول محور یک موضوع با هم ارتباط دارند. استفاده از فرهنگ‌های طیفی، امروزه نقش مهمی را در فعالیت‌های هوش مصنوعی ایفا نموده و جای خالی آن در متون علوم اسلامی نیز کاملاً مشهود است. در معاجم موضوعی همچون «فقه اللغة» ثعالبی (۴۲۹ق)، به دسته‌بندی مداخل حول موضوعات - و نه حول ریشه - پرداخته شده است. اگر مداخلی که به عنوان پاسخ لغت‌نامه ارائه می‌شود، دارای این نوع ارتباط موضوعی با سایر مداخل باشد، کاربر می‌تواند علاوه بر مشاهده توصیف مدخل درخواستی خود، مداخل مرتبط با آن را ذیل موضوع مشخص شده ببیند. در زمینه ارتباطات شمول (رابطه کلی و جزئی) یا جزءاژگی (رابطه جزء و کل) و حتی سببیت، کتاب‌هایی همچون المخصص (قرن پنجم) نوشته ابن‌سبیده، در لغت‌نامه هوشمند مرکز نور مورد استفاده قرار گرفته است. المخصص، لغت‌نامه‌ای ۱۷ جلدی از گسترده‌ترین معاجمی است که بر اساس معنای کلمات و به صورت موضوعی تدوین یافته است. مؤلف، اولین موضوعی که در کتاب خود بعد از مقدمه‌اش ذکر می‌کند، انسان است و سعی نموده یک سیر منطقی و عرفی را در ترتیب موضوعات و به تبع آن، در ترتیب باب‌ها رعایت نماید؛ مثلاً ذیل موضوع «خلق انسان»، بابی به نام «الحمل والولادة» قرار گرفته است که کلمات آن، بر اساس اولین روزهای انعقاد نطفه تا ایام بزرگسالی مرتب و شرح داده شده است. تلخیص ۲ جلدی از کتاب المخصص نیز با نام الإفصاح فی الفقه اللغة منتشر گردیده که می‌تواند جهت پرهیز از تطویل کتاب اصلی و چینش منسجم آن، در استخراج اطلاعات مفید باشد. استخراج جمع‌های مکسر جمع‌آوری شده در کتاب ارزشمند المعجم المفصل فی الجموع نوشته دکتر بدیع یعقوب و تکمیل آن با داده برچسب‌گذاری شده در مرکز نور که منجر به تجمیع بیش از ۲۰ هزار جمع مکسر شده، به عنوان مرتبطات در جهت اتصال مفرد به جمع مکسر و برعکس و ارائه اطلاعات جانبی به پژوهشگران نیز مورد استفاده قرار گرفته است.

۵. شاهد مثال؛ به‌خصوص از آیات و روایات

در کنار بیان مفهوم مداخل، آوردن مثال‌های متعدد نیز بسیار ضروری است. آوردن مثال و شیوه کاربرد لغات در جایگاه‌های مختلف زبانی، یکی از نقاط عطف لغت‌نامه‌های بزرگ است؛ حتی در مواردی که مثال مناسبی بیان نشده باشد، به‌راحتی می‌توان

مثال‌های ساختگی، ولی صحیح را استخراج نمود. نمونه‌هایی در قالب مثال‌های کلی همچون «ذهب بشیء» (معادل some-thing, someone, ... در زبان انگلیسی) و مثال‌هایی با مصداق مشخص و جزئی. با استفاده از فرآیندهای ماشینی، می‌توان آیات و روایاتی را که مشتمل بر کلمه یا عبارت مد نظر پژوهشگر است نیز برای او نمایش داد و بر اساس تعداد کلمات یا کاراکترهای از پیش تعیین‌شده، علایم ویرایشی و برچسب‌ها، محدوده مشخصی از آیات و روایات معتبر را به صورت گزینشی برای کاربران ارائه نمود.

۶. غریب الحدیث

تعدادی از معاجم لغوی، فقط مداخلی را جمع‌آوری و توصیف

فرهنگ اصطلاحات در حوزه‌های

متنوع علمی نیز از آنجا که مشتمل

بر ساختار مدخل محور هستند،

می‌توانند به عنوان اطلاعات

جانبی، در کنار توصیف معنایی

مداخل عرضه گردد. با انتخاب

فرهنگ‌های اصطلاحات در حوزه

علوم متنوع و پردازش اطلاعات

آنها در لغت‌نامه، امکان آشنایی

کاربر با مفهوم اصطلاحی مداخل

نیز فراهم شده است. در فاز

نخستین پروژه لغت‌نامه هوشمند،

استفاده از اصطلاح‌نامه‌های

موجود در نرم‌افزارهای رومیزی

مرکز تحقیقات کامپیوتری علوم

اسلامی، در اولویت قرار گرفت



کرده‌اند که در ضمن حدیثی آمده و فهم معنایی آن، نیاز به توضیح داشته باشد. احادیث مشتمل بر این مداخل به همراه توصیف آنها، در سرویس لغت‌نامه در کنار ارائه توصیفات لغوی در اختیار کاربر قرار خواهد گرفت. از جمله این کتاب‌ها می‌توان به الفائق نوشته زمخشری (۵۳۸ق) و کتاب النهایة نوشته ابن‌اثیر (۶۰۶ق) اشاره نمود.

بحث ریشه‌شناسی می‌تواند تأثیری عمیق نیز در مباحث شبکه واژگانی عربی داشته باشد. در مباحث دستور زبان عربی، مبحثی به عنوان اشتقاق مطرح است که برخی از اندیشمندان این عرصه، آن را به سه نوع: صغیر، کبیر و اکبر تقسیم نموده‌اند (کبیر، ۱۳۸۹: ج ۱، ۱۷۴). نوع اول، مانند اشتقاق ضارب و مضروب از مصدر «الضرب» که در آن، ترتیب حروف اصلی و ریشه کاملاً رعایت شده است. نوع دوم، مانند کلماتی که دارای همان حروف هستند؛ ولی ترتیب رعایت نشده است؛ مانند «کلم» (به معنای جرح) و «لکم» (به معنای ضربت با دست) که دارای معنایی نزدیک به یکدیگرند و همچنین است کلماتی مانند «کمل» و «ملک».

۷. اصطلاحات

هر زبانی، اصطلاحات و عبارات ویژه‌ای دارد که به صورت تحت‌اللفظی قابل درک و ترجمه نیستند؛ حتی اگر زبان‌آموز تمام کلمات یک عبارت را بداند و به‌خوبی به دستور زبان مسلط باشد، باز هم در فهم معنای عبارت موفق نخواهد بود. بسیاری از ضرب‌المثل‌ها، عبارات غیررسمی و گفتارهای محاوره، ممکن است چنین مشکلی را برای زبان‌آموز ایجاد کنند. این اصطلاحات، در واژه‌نامه‌های ویژه‌ای با نام «Idioms dictionary» مورد بررسی قرار می‌گیرند. فرهنگ اصطلاحات در حوزه‌های متنوع علمی نیز از آنجاکه مشتمل بر ساختار مدخل‌محور هستند، می‌توانند به عنوان اطلاعات جانبی، در کنار توصیف معنایی مداخل عرضه گردد. با انتخاب فرهنگ‌های اصطلاحات در حوزه علوم متنوع و پردازش اطلاعات آنها در لغت‌نامه، امکان آشنایی کاربر با مفهوم اصطلاحی مداخل نیز فراهم شده است. در فاز نخست پروژه لغت‌نامه هوشمند، استفاده از اصطلاح‌نامه‌های موجود در نرم‌افزارهای رومیزی مرکز تحقیقات کامپیوتری علوم اسلامی، در اولویت قرار گرفت.

۸. ریشه‌شناسی

یکی از مهم‌ترین مباحث زبان‌شناسی، بحث ریشه‌شناسی است که نقش اصلی را نیز در سیر تحولات تاریخی یک زبان ایفا می‌کند. فهم درست متون، به‌خصوص متون تاریخی، منوط به فهم صحیح معنای واژگان آن است و در این بین، ریشه‌شناسی با مشخص نمودن تطورات واژه، رابطه لفظ و معنا را در هر دوره مشخص می‌نماید (شجاعی و ثقفیان، ۱۳۹۶: ۸-۹). در کتاب ارزشمند «معجم مقاییس اللغة» نوشته ابن‌فارس (۳۹۵ق)، در ابتدای هر بخش و قبل از پرداختن به مداخل، مؤلف ذیل هر ریشه، اقدام به توصیف معنایی خود آن ریشه نموده است. این اطلاعات نیز پس از استخراج ماشینی و بازمینی دستی، به عنوان یک سرویس مستقل در بیان معنای ریشه و ذیل هر مدخل، در اختیار کاربر قرار می‌گیرد. از کتاب‌های متأخرین نیز می‌توان به کتاب «التحقیق فی

اگرچه در لغت‌نامه‌های چاپی سنتی، در مورد مهارت‌های مرتبط با استفاده از فرهنگ لغات، بسیار صحبت شده است، اما سخن پیرامون مهارت‌های مورد نیاز در زمینه استفاده از فرهنگ لغات الکترونیکی، بسیار اندک بوده و به یکی از دغدغه‌های جدی طراحان این برنامه‌ها مبدل شده است. از این رو، در کنار طراحی یک پوسته و قالب مطلوب مطابق با نیازهای کاربران و خدمات ارائه‌شده به پژوهشگران، نیاز به آموزش‌هایی در کسب مهارت در زمینه استفاده از لغت‌نامه هوشمند نور نیز امری انکارناپذیر است.

گرفت و «قاموس نور» به عنوان یک لغت‌نامه هوشمند که در رفع آن چالش‌ها دارای عملکرد مطلوبی بوده است، معرفی گشت. در این لغت‌نامه، با ایجاد «بانک مداخل» که داده‌ای منسجم از مداخل لغوی و توصیفات مندرج در واژه‌نامه‌های معتبر بوده و با چینیسی نظام‌مند بر اساس «مدخل نماینده» و «توصیف نماینده» مرتب شده، لغت‌نامه‌های متقدم عربی بر اساس یک ساختار واحد ساماندهی شدند و از سوی دیگر، با تحلیل هوشمند صرفی کلمات ورودی پژوهشگران توسط دو موتور قدرتمند تحلیلگر صرفی نور و استمر هوشمند نور که به تشخیص ریشه، استم (میان‌وند) و لمای (بن) کلمات منجر شده، زمینه اتصال ماشینی به نزدیک‌ترین مداخل لغوی برای پژوهشگران مهیا گشت. تحلیل صرفی کلمات، از دستاوردهای منحصر به فرد این واژه‌نامه هوشمند است که در قالب پایگاه اینترنتی، نرم‌افزار رومیزی و تلفن‌های همراه قابل استفاده است. همچنین، با ایجاد شبکه واژگانی بین مداخل مشتمل بر مترادفات، متضادات، اضداد، الحاق اصطلاحات، بحث‌های ریشه‌شناسی و تطبیق بر آیات قرآن و روایات اهل بیت (ع)، اطلاعات متنوع و ارزشمند دیگری نیز در اختیار پژوهشگر قرار خواهد گرفت که این واژه‌نامه هوشمند را از دیگر رقبای خود متمایز می‌سازد.

۲. احیای لغت‌نامه‌های عربی مبتنی بر رویکرد قرآنی و روایی.

نوع سوم، اشتقاقی است که برخی حروف اصلی تغییر نموده و با حروف هجایی مشابه جایگزین شده است. مانند دو کلمه «فصم» و «قصم» که جامع معنای آن دو، شکستن و قطع کردن است و همچنین است کلمات «ثلب و ثلم» و «فطر و فطم». با بهره از ابزارهای هوشمند متن‌کاوی، می‌توان کلمات عربی را بر اساس نوع و چینش حروف تحلیل ماشینی نمود. اگر شبکه واژگانی به‌گونه‌ای ترسیم شود که اغلب کلماتی که دارای اشتقاق کبیر و اکبر هستند، رابطه شبکه‌ای و معنایی با دیگر کلمات داشته باشند، چه‌بسا بتوان چنین نتیجه گرفت که در زبان عربی، نه تنها کلمات، بلکه حروف نیز دارای ماهیتی استقلالی بوده و هر حرف در بار معنایی کلمات و هم‌معناها دخیل است؛ مطلبی که در هیچ‌یک از دیگر زبان‌ها به آن اشاره‌ای نشده و هنوز مورد پردازش هوشمند تحلیلگران زبان عربی واقع نشده است.

۹. استفاده از اطلاعات مربوط به حوزه مدیا

استفاده از تصاویر و صوت‌هایی جهت تلفظ صحیح مداخل عربی در کنار نوشتار فونتیک و بین‌المللی کلمات عربی، امری ضروری است.

نتیجه و پیشنهاد

پس از بررسی اجمالی انواع زبان عربی و سیر تحولات تاریخی آنها، به ساختار و ویژگی لغت‌نامه‌های مربوط به هر کدام از زبان‌ها پرداخته شد و بر ضرورت طراحی و تولید یک لغت‌نامه هوشمند رقمی، از دو جهت تأکید گشت:

۱. دسترسی سهل و آسان پژوهشگران به منابع لغوی با گذر از شیوه‌های سنتی مراجعه به منابع لغوی متعدد و پیاده‌سازی حداکثری الگوهای ماشینی هوشمند جهت ارائه پاسخ‌های صحیح و جامع به پژوهشگران.

در این بخش، چالش‌های مطرح در لغت‌نامه‌های هوشمند مورد تحلیل و بررسی قرار



انواع موتورهای تحلیلی کلمات فارسی و انگلیسی و دیگر زبان‌ها که طراحی و تولید آنها به مراتب از زبان عربی راحت‌تر است، می‌توان به لغت‌نامه معتبر چندزبانه مبدل گشت. با تحقق این اهداف، امید است «قاموس نور» به مرجعی در حوزه لغت عربی مبدل گردد و بستری مناسب جهت ارائه بحث‌های مربوط به ریشه‌شناسی، زبان‌شناسی و تمامی مسائل مهم مربوط به لغت توسط محققان و اندیشمندان باشد.

همان‌گونه که ریشه‌شناسی در گروهی فهم تغییر و تحولات تاریخی یک زبان است (Campbell; 1998: 1) و می‌توان به واسطه ریشه‌شناسی سیر تغییر کلمات را در طول تاریخ مشاهده نمود، عکس آن نیز صادق بوده و بحث زبان‌شناسی و ریشه‌شناسی و مباحث مربوط به فقه اللغة نیز می‌تواند در مباحث تاریخی مربوط به اشخاص و مکتوبات به‌جای‌مانده راهگشا باشد؛ یعنی بین زبان‌شناسی تاریخی و مباحث فقه اللغة و ریشه‌شناسی، یک پیوند دوسویه برقرار گردد. تحلیل هوشمند متن‌های برجای‌مانده از اعصار گذشته، به واسطه ابزارهای متن‌کاوی و بررسی تعداد، نوع و شیوه استعمال کلمات در یک نگاشته، می‌تواند به عنوان شناسنامه‌ای برای یک اثر تلقی شده و زمان تألیف، اثرپذیری آن از نویسندگان متقدم از خود، اثربخشی آن بر آثار پسینی و خصوصیات سبک نگارش نویسنده مشخص گردد.

باید توجه داشت ایجاد بانک مداخل به همراه تمامی توصیفات و اطلاعاتی که حول این مداخل شکل گرفته، می‌تواند به عنوان یک داده جدید در عرصه زبان‌شناسی قلمداد شود و در سیستم‌های بازیابی اطلاعات نیز نقش مهمی را ایفا نماید. استفاده از عناصری چون: مدخل لغوی، توصیف عربی، ترجمه فارسی و دیگر زبان‌ها و شبکه واژگانی، الگویی از هم‌نشینی کلمات را ایجاد خواهد کرد که می‌تواند کمک شایانی به سیستم‌های هوشمند متن‌کاوی همچون موتورهای ترجمه و ماشین‌های مشابهت‌یاب معنوی، طبقه‌بندی موضوعی متون، خلاصه‌ساز، خوشه‌بندی متون و سیستم‌های دیگری که وابسته به مداخل لغوی و تجزیه صرفی کلمات هستند، داشته باشد. نه تنها بهره از این داده ارزشمند می‌تواند راهگشای بسیاری از فعالیت‌های متن‌کاوانه باشد، مراحل طراحی و شیوه دستیابی به این داده نیز می‌تواند به عنوان یک الگوی کاربردی در دیگر علوم اسلامی مورد استفاده قرار گیرد.

در این طرح، تمامی گروه‌های پژوهشی متناسب با حوزه فعالیت‌های خود درگیر بوده و هر توضیح و توصیفی (توصیف به معنای عام آن، یعنی عبارتی را که دارای «از» و «تا» باشد) را که جهت اتصال به مداخل این پیکره دادگان آماده باشد، به

طراحی لغت‌نامه هوشمند «قاموس نور» می‌تواند حرکتی پویا در نشر معارف اهل‌بیت (ع) بوده و به عنوان خدمتی به علوم اسلامی، پژوهشگران را به بحث‌های زبان‌شناسی و لغت‌شناسی مربوط به قرآن و روایات نیز ترغیب نماید.

آنچه تاکنون در فاز عملی «قاموس نور» صورت پذیرفته، تنها محدود به برخی واژه‌نامه‌های معتبر عربی بوده و در برخی فعالیت‌های هوشمندسازی، هنوز بازبینی نهایی توسط محققان به سرانجام نرسیده است. از این‌رو، توسعه منابع لغوی، تقویت ابزارهای تحلیل صرفی، بازبینی دقیق مداخل و توصیفات آنها جهت تجمیع و ساماندهی بهتر، گسترش شبکه واژگانی، اصطلاح‌نامه‌ها، توضیحات مربوط به اعلام از کتاب‌های تخصصی و ارتباط بهتر با آیات و روایات، از اهداف پیش‌رو در فازهای بعدی این لغت‌نامه هوشمند است. هدف نهایی و مطلوب از انجام تمامی این مراحل، ارائه یک لغت‌نامه متقن بر اساس مداخلی است که برای هر یک از آنها، توصیفات مندرج در تمامی لغت‌نامه‌های معتبر بر اساس مفاهیم، درصد استعمال و حقیقت و مجاز ترتیبی منسجم یافته و تمامی کلمات عربی نیز با تحلیل صرفی هوشمند به آنها متصل گردیده است. توسعه زبانی از عربی به دیگر زبان‌ها و برعکس، هدفی است که با پالایش مداخل نضج‌یافته در واژه‌نامه‌های متأخر، به‌راحتی قابل دستیابی بوده و با به‌کارگیری

**توسعه منابع لغوی، تقویت ابزارهای
تحلیل صرفی، بازبینی دقیق مداخل
و توصیفات آنها جهت تجمیع و
ساماندهی بهتر، گسترش شبکه
واژگانی، اصطلاح‌نامه‌ها، توضیحات
مربوط به اعلام از کتاب‌های
تخصصی و ارتباط بهتر با آیات
و روایات، از اهداف پیش‌رو در
فازهای بعدی این لغت‌نامه هوشمند
است**



16: 2013). از این رو، در کنار طراحی یک پوسته و قالب مطلوب مطابق با نیازهای کاربران و خدمات ارائه شده به پژوهشگران، نیاز به آموزش‌هایی در کسب مهارت در زمینه استفاده از لغت‌نامه هوشمند نور نیز امری انکارناپذیر است. فرآیندهای تحلیل صرفی و استفاده از ابزارهای متن‌کاوی به همراه گستره اطلاعات ارائه شده در مورد کلمات عربی، چه بسا عرضه الکترونیکی لغت‌نامه هوشمند را با چالش‌هایی مواجه سازد که در گذر زمان و بازخورد نظرات پژوهشگران، در مسیر رشد و ارتقا قرار گیرد. ■

منابع:

پی‌نوشت‌ها و منابع این مقاله، به دلیل فراوانی ارجاعات، در پایان نوشتار ذکر نگردید. تمامی مستندات مقاله، در دفتر فصلنامه موجود است.

مداخل متصل می‌نمایند. استفاده از متون کتاب‌های موجود، اعم از فهرست‌های متنوع موجود در هر کتاب و اطلاعات پیرامون یک مدخل یا موضوع در متن اصلی یا مستندسازی و حاشیه‌نگاری پاورقی‌ها نیز از جمله اطلاعاتی است توسط محققان هر علوم به مداخل متصل می‌گردد. نمونه روشن آن، در کتاب‌های نرم‌افزار تراجم مانند اعیان الشیعه، الذریعة و یا نرم‌افزار طب، جغرافیا و موارد مشابه جهت توصیف یک شخص، یک گیاه، یک منطقه و غیره خواهد بود. یکی از مهم‌ترین منابع موجود برای این امر نیز مجلات علوم اسلامی منتشرشده در پایگاه نورمگز است. کلیدواژه، عناوین، چکیده، پاورقی‌ها، منابع انتهایی، لاگ کاربران، پیشنهاد مقالات مشابه، لینک ارتباطی با کتابخانه آنالاین و بسیاری اطلاعات دیگر، همگی گنجی غنی هستند که می‌توانند در ایجاد یک پیکره دادگانی منسجم از علوم مختلف مؤثر باشد.

گفتنی است، در مرحله تحلیل صرفی کلمات ورودی و فرآیند اتصال عبارات به پیکره دادگان، یکبار این مسیر در لغت‌نامه هوشمند طی شده است و پژوهشگران با جست‌وجوی موضوعات پژوهشی خود، با استفاده از این ابزارهای هوشمند می‌توانند به گنجینه اطلاعات مدون در پیکره دادگان آن علوم دست یابند.

نکته پایانی، اینکه اگرچه در لغت‌نامه‌های چاپی سنتی، در مورد مهارت‌های مرتبط با استفاده از فرهنگ لغات، بسیار صحبت شده است، اما سخن پیرامون مهارت‌های مورد نیاز در زمینه استفاده از فرهنگ لغات الکترونیکی، بسیار اندک بوده و به یکی از دغدغه‌های جدی طراحان این برنامه‌ها مبدل شده است (Lew;