

لغت‌نامه استاندارد نور

ویژگی‌ها و چشم‌اندازها



گفت‌وگو با مسیح توحیدی

مدیر گروه ادبیات متن‌کاوی مرکز تحقیقات کامپیوتری و علوم اسلامی

به کوشش: علی نعیم‌الدین خانی

ضمن تشکر از جنابعالی، لطفاً خودتان را برای خوانندگان فصلنامه معرفی کنید.

بنده مسیح توحیدی هستم که در سومین دوره همکاری خود با مرکز تحقیقات کامپیوتری علوم اسلامی، مدت چهار سال است در خدمت همکارانم در گروه ادبیات و متن‌کاوی از معاونت پژوهشی مرکز می‌باشم.

بحث لغت‌نامه، چه پیشینه‌ای را در مرکز نور پشت سر نهاده است؟

موضوع صحبت من در اینجا، قاعدتاً وضعیت فعلی لغت‌نامه نور است؛ اما برای رسیدن به نقطه فعلی، باید مقدماتی را طی می‌کردیم. همچنین، در ادامه، توسعه‌ای در این زمینه خواهیم داشت که چیزی فراتر از لغت‌نامه مصطلح خواهد بود.

اطلاع دارید که اولین نسخه‌ای که در مرکز نور به عنوان لغت‌نامه تولید شده، قاموس النور بوده است. نسخه دوم این محصول که آخرین نسخه آن باشد،

اشاره

امروزه با توسعه و رشد فناوری اطلاعات، ابزارهای گوناگونی در حوزه تحقیقات شکل گرفته و راه دستیابی به علوم و تولید دانش، هموارتر از گذشته گردیده است. مرکز تحقیقات کامپیوتری علوم اسلامی، از جمله مراکز پیشرو در زمینه داده‌پردازی علوم اسلامی است که تولیدات فاخر و ارزشمندی را تقدیم جامعه علمی نموده است.

یکی از اموری که پیوسته پژوهشگران در روند تحقیق به آن نیاز داشته و دارند، دسترسی به لغت‌نامه و فرهنگ واژگان یا اصطلاحات است که آنان را در خصوص فهم آسان کلمات و درک بهتر معانی عبارات کتب یاری رساند.

نظر به اهمیت این بحث، در این شماره از فصلنامه، به حضور مدیر گروه ادبیات متن‌کاوی نور، جناب آقای مسیح توحیدی رسیدیم و کارهای علمی و فنی صورت‌گرفته در این مرکز را در حوزه مربوط به لغت جويا شدیم؛ امید که مباحث مطرح‌شده در این گفت‌وگو برای خوانندگان و علاقه‌مندان عزیز قابل استفاده باشد.

چند سال پیش تولید شده است. قاموس النور نسخه ۱ و ۲، هر دو، نوعی کار کتابخانه‌ای بوده‌اند که در آنها علاوه بر در اختیار قرار دادن تعدادی از منابع لغوی، دسترسی سریع و آسان به محتوا نیز امکان داشت. در نسخه اول این محصول، تعداد منابع موجود در آن خیلی محدود و معدود بود؛ اما در نسخه بعدی قاموس، این منابع به چهل عنوان رسید و البته کار نسبتاً پژوهشی نیز در کنار آن انجام گرفته بود.

برای ساماندهی دیجیتالی متون منابع لغوی، پیش از این چه اقداماتی صورت گرفته بود؟

به طور کلی، در زمانی که کار لغت در مرکز شکل گرفت، دو دسته فرمت را روی متون لغت پیاده‌سازی کردند؛ یکی فرمت ریشه و دیگری فرمت مشتق.

می‌دانیم که عمده کتاب‌های لغت، ریشه‌محور هستند و تبویب این کتب، بر اساس ریشه می‌باشد و بر مبنای یک ریشه که معمولاً هم چینش الفبایی دارند، به توصیف مداخل می‌پردازند و از آنجا که بعضی از این کتاب‌ها، ریشه‌محور نبودند، مثل فرهنگ ابجدی، فرمت جداگانه‌ای در مرکز به عنوان «ریشه» در نظر گرفته شد. از سوی دیگر، برخی از کتب لغت نیز تفاوت مبنایی در تعیین ریشه برای یک مدخل داشتند و همچنین، ممکن بود بین محققان مرکز نور با مؤلف کتاب، تفاوت نظر در تعیین ریشه هر مدخل وجود داشته است. پس، فرمتی را به عنوان «ریشه» در کنار مشتقات ذکر کرده‌اند. فرمت دیگر، مشتق بود و بنا بر تعریف کاری که آن زمان صورت گرفت،

این فرمت به تمامی مشتقات ذکر شده در ذیل هر ریشه زده می‌شد که الزاماً همه مشتقات مدخلی که دارای توصیف باشند، نبودند. این مقدمه را عرض کردم که تفاوت مشتق با مدخل که بعداً توضیح بیشتری خواهم داد، روشن شود.

لغت‌نامه‌ای که در مرکز نور در حال تولید است، چه ویژگی‌هایی دارد؟

لغت‌نامه‌ای که الآن مد نظر ماست، یک لغت‌نامه استاندارد است. در این لغت‌نامه، ذیل هر ریشه، یک‌سری مدخل داریم. در مقابل این مداخل، توصیف و توضیح مشخصی که برای کاربر قابل فهم باشد، ارائه شده است. بین واژه مشتق و مدخل، رابطه عام و خاص مطلق وجود دارد؛ زیرا الزاماً همه مشتقاتی که ذیل یک ریشه می‌آید، در مقابل آن توضیح و توصیفی وجود ندارد؛ ولی تمام مداخل، دارای توضیح هستند. مشکلی که پیش از این وجود داشت، آن بود که همه مشتقات فرمت خورده بودند و در فهرست مشتقات شما، کلماتی وجود داشت که هیچ توضیحی نداشتند و عملاً برای محقق یا کاربر مفید نبود؛ به بیان دیگر، اگر شما به لغت‌نامه‌های قبلی مرکز مراجعه کنید، می‌بینید که در آنها دو نوع فهرست وجود دارد: یکی فهرست ریشه، و دیگری فهرست مشتقات. در قسمت مشتقات، وقتی واژگان آن را مرور کنید، ملاحظه خواهید کرد که بسیاری از آنها، معمولاً مدخل نبوده، فاقد توصیف مفید می‌باشند. این، نقیصه‌ای بود که در لغت‌نامه‌های قبلی نور وجود داشت و ما باید آن را اصلاح می‌کردیم.

لغت‌نامه هوشمند

(پیشنهاد مداخل لغوی متناسب با کلمات)

عدم نمایش کتب معانی

همه موارد ابجدی المنجد لسان الرائد الوسيط الغنوی

همه موارد اصل کلمه استم موتور صرف ریشه

کلمه	نوع	رتبه	کتاب	ریشه	پدر	مدخل	توصیف
مقاطع	استم	۲	لسان	قَطَعَ	مَقَطَعٌ	مَقَاتِعُ الْأَنْهَارِ	حَيْثُ يُعْبَرُ فِيهِ
مقاطع	استم	۲	لسان	قَطَعَ	مَقَطَعٌ	مَقَاتِعُ الْقُرْآنِ	مَوَاضِعُ الرَّقُوفِ وَ مَبَادِيهُ مَوَاضِعُ الْإِجْتِدَاءِ
مقطع	صرف	۲	لسان	قَطَعَ	مَقَطَعٌ	مَقَطَعٌ كُلُّ شَيْءٍ	أَخْرَجَ
مقطع	صرف	۲	لسان	قَطَعَ	مَقَطَعٌ	عَوْدٌ مَقَطَعٌ	إِذَا لَقِطَعَ عَنِ الضَّرَابِ
مقطع	صرف	۲	لسان	قَطَعَ	مَقَطَعٌ	رَجُلٌ مَقَطَعٌ	لَا دِيُونَ لَهُ
مقطع	صرف	۲	ابجدی	قَطَعَ	قَطَعٌ	أَيُّ أَقْطَعُ بِذَلِكَ...	مَنْ يَهِيَ أَنْ يَطُورَ قَطَعَ يَقِينُ دَارِمٍ
مقطع	صرف	۲	لسان	قَطَعَ	مَقَطَعٌ	رَجُلٌ مَقَطَعٌ	مُجْرِبٌ

تکرار	eatNumber
بالمراقة	۲۵
بالمسافات	۲۵
بالمغلق	۲۵
بالمقاطع	۲۵
بالمكافئ	۲۵
بالوحد	۲۵
بالمفطور	۲۵
بانجائرا	۲۵
بانفرادك	۲۵
ببوز	۲۵
ببوزوا	۲۵
ببئلاج	۲۵
ببئجرك	۲۵
ببئجهمين	۲۵
ببئشكل	۲۵
ببئقلية	۲۵

چه کارها و اصلاحاتی انجام شد تا به وضعیت مطلوب دست پیدا کردید؟

کاری که باید انجام می‌دادیم، این بود که منابع لغوی را ساماندهی کرده، مداخل و توصیفات آنها را استخراج کنیم. برای این کار، مقدماتی را انجام دادیم تا به نقطه فعلی برسیم. نخست اینکه چون فرمتی برای این کار نداشتیم یا فرمت‌هایی که داشتیم، به درد این کار نمی‌خورد، کتاب‌هایی را انتخاب کردیم که از نظر علائم ویرایشی و نقطه‌گذاری دارای ساختار مناسبی باشند و بتوانیم با پردازش ماشینی، مداخل و توصیفاتش را حتی‌المقدور به شکل ماشینی استخراج کنیم. به همین جهت، نگاه اول ما به منابع از این نظر بود که آیا این منبع امکان آن را دارد که بتوانیم قاعده‌هایی را مشخص نماییم که ماشین بتواند تشخیص دهد که فلان کلمه، مدخل است و چیزی که مقابلش آمده نیز توصیف آن است؟ بحمدالله، توانستیم این کار را انجام دهیم و کتاب‌هایی با این ویژگی انتخاب شد و وارد چرخه کارمان قرار گرفت.

پس از این، شروع به بازیابی کار نمودیم؛ چون ممکن بود این قواعد در برخی مواضع، درست رعایت نشده باشد یا ماشین بعضی موارد را به‌درستی تشخیص نداده باشد. این بازیابی، در معاونت پژوهشی انجام شد؛ یعنی قسمت اول کار که استخراج قواعد بود، توسط همکارانمان در معاونت پژوهش به سامان رسید و بعد در معاونت فنی، دوستانمان به استخراج ماشینی این قواعد پرداختند و آنگاه دوباره توسط عزیزان ما در پژوهش، این کار ماشینی، به طور دستی بازیابی شد تا نواقص کار گرفته شود و اصلاح و تکمیل گردد. البته غالب موارد، صحیح بودند و بعضی موارد که اشکال داشتند، اصلاح شدند.

بنابراین، می‌توان گفت در نسخه جدید لغت‌نامه نور، اگرچه از نظر تعداد منابع در مقایسه با نسخه قبلی، شمار کمتری از کتب ارائه شده است، ولی از جهت کیفیت و کارآمدی برنامه، ارتقای قابل توجهی داشته است؛ زیرا بنای ما ساماندهی مطلوب و تجمیع بهینه اطلاعات و استانداردسازی کار بوده است؛ البته این ساماندهی، الزاماتی داشته که از ما نیز وقت و هزینه بسیاری گرفته و البته در جای خود، لازم و مفید بوده است.

با این توضیحات، در نسخه جدید، لغت‌نامه استاندارد و کارآمدی برای محققان طراحی کرده‌اید؟

بله؛ اگر بخواهم اجمالاً عرض کنم، باید بگویم که ما یک لغت‌نامه استاندارد را طراحی کرده‌ایم که مشتمل بر کتاب‌های

ویژگی جدیدی که لغت‌نامه استاندارد نور دارد، این است که بر خلاف نسخه‌های قبلی که با ارتقای نسخه، آمار کتب برنامه افزایش می‌یافت، در نسخه فعلی، حدود هفت منبع از میان پهل کتبی که در برنامه قبلی عرضه شده بود، کم شده است؛ زیرا بنای ما ساماندهی مطلوب و تجمیع بهینه اطلاعات و استانداردسازی کار بوده است؛ البته این ساماندهی، الزاماتی داشته که از ما وقت و هزینه بسیاری نیز گرفته و البته در جای خود، لازم و مفید بوده است

مرجع و مهم در حوزه لغت است و در مجموع، شامل تعدادی مدخل است و مقابل هر مدخل، حداقل یک توصیف یا توضیح وجود دارد. برخی مداخل، به تناسب نوع مدخل، توصیفات بیشتری دارند که ما همه آنها را ارائه کرده‌ایم.

در استخراج ماشینی مداخل و توصیفات از منابع انتخاب‌شده، در کتبی مثل: العین، المحيط فی اللغة و تاج العروس، وجود فرمت‌های ریشه و مشتق می‌توانست کمکی برای استخراج بهتر ماشینی باشد؛ ولی در خصوص برخی کتب که فاقد این فرمت‌ها بودند، فقط با استفاده از علائم ویرایشی موجود در متن، اقدام به استخراج ریشه، مداخل و توصیف آنها شد.

به صورت میانگین، چیزی حدود صد هزار مدخل در همه کتاب‌های لغت مطرح شده است. کاری که در قاموس النور صورت گرفته بود، این بود که حدود یک میلیون مشتق غیرتکراری داشتیم؛ یعنی چیزی حدود نه دهم آن، اضافه بوده است. به همین جهت، لازم می‌نمود که این فهرست تکراری را ساماندهی کنیم. به هر حال، یک تعداد ریشه و نیز مشتق، در کتاب‌های لغت هست که مشترک‌اند؛ با این تفاوت که توضیحات مداخل آنها، با یکدیگر تفاوت‌هایی دارند.

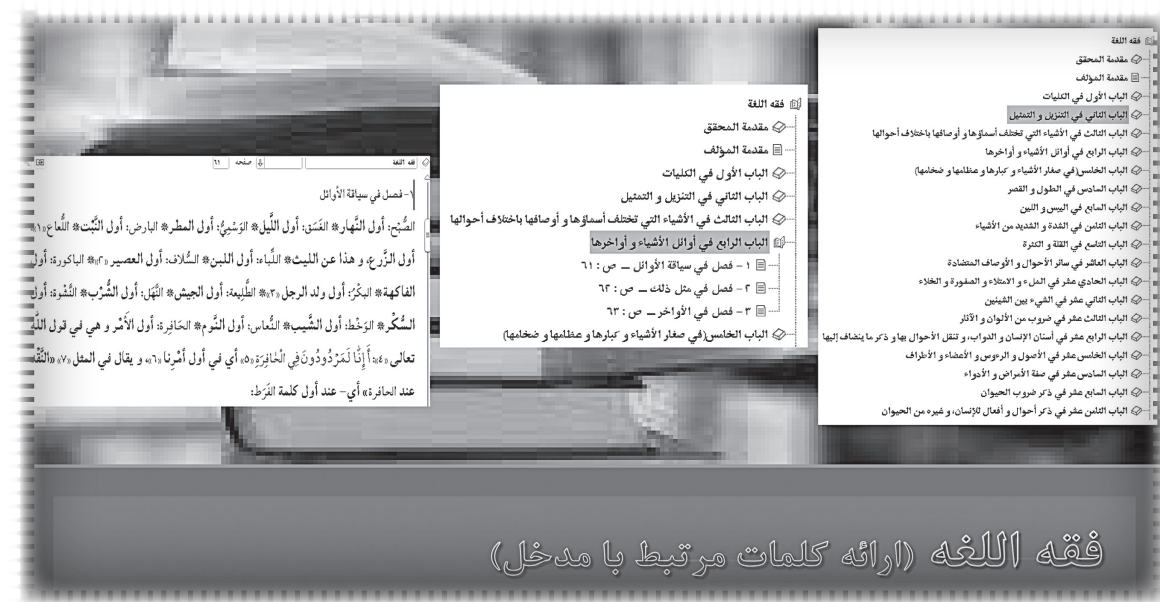
دلیل درست شدن این حجم انبوه از مداخل چه بود؟

وقتی مداخل حدود هفت کتاب معروف لغت را کار کردیم و همه را با هم تجمیع نمودیم، با حدود پانصد هزار مدخل مواجه شدیم که هر کدام دارای توصیف هم بود. دلیل این کثرت در مداخل، چند چیز بود؛ از جمله، تنوع در نگارش این مداخلها؛ مثلاً در یک کتاب از نظر اعراب گذاری هیچ اعرابی روی مدخل وجود نداشت؛ ولی در کتاب دیگر، اعرابش گویا بود و به راحتی می توانستیم تشخیص دهیم این کلمه چگونه تلفظ می شود و یا چه صیغه ای است. از این رو، وقتی این مداخل را تجمیع نمودیم، ماشین این تفاوت های جزئی را شناسایی کرد و بر اساس قواعدی که مشخص شد، مواردی که قابل جمع بودند، در یک گروه قرار گرفتند.

دلیل دیگر، تفاوت تعبیر مؤلف در بیان مقصود خود بود؛ به

بودند، ولی تنوع در نگارش داشتند، قواعدی را مشخص کردیم که ماشین بتواند به واسطه آنها، کلمات متنوع از حیث اعراب، اما مشابه را شناسایی کند و همه را یکجا جمع نماید؛ چه بدون اعراب و چه با اعراب؛ به عنوان نمونه، «استکرم» در این منابع لغوی به چهار صورت نگارش یافته بود که به شکل ماشینی تجمیع شدند (اَسْتَكْرَمَ، اسْتَكْرَمَ، اسْتَكْرَمَ) و از میان آنها، مورد اول که کامل ترین اعراب را دارا بود، به عنوان مدخل نماینده انتخاب شد.

اما آنجا که مداخل ما چند کلمه ای بودند، با توجه به ضرورت تجمیع این مداخل، اصطلاحی به عنوان «مدخل پدر» جعل کردیم و برای هر مدخل، یک مدخل پدر که همان کلمه محوری بود، در نظر گرفتیم؛ برای مثال، اگر هم اکنون مدخل نماینده «العامة» را ملاحظه کنیم، مشاهده می کنیم که ذیل



فقه اللغة (ارائه کلمات مرتبط با مدخل)

عنوان مثال، مدخل «استکرم»، ذیل ریشه «کرم» آمده است و مدخل دیگر، «استکرم الشيء» است که در کتابی دیگر، ذیل همان ریشه آمده است. در این گونه موارد، اگرچه توصیف هر مدخل، متفاوت از توصیف دیگر است، اما در دو مدخل، واژه محوری که ذیل ریشه «کرم» بیان شده، «استکرم» می باشد.

چگونه این مشکل را حل کردید؟

برای حل این مشکل، برای مواردی که به صورت تک کلمه

آن، ۲۳ مدخل همچون: «المشروطة العامة»، «الأمور العامة» و «الممكنة العامة» وجود دارند که تمامی این مداخل، در داشتن کلمه «العامة» که ما از آن به «مدخل پدر» تعبیر می کنیم، مشترک هستند.

وقتی به سوی ایجاد مدخل پدر رفتیم، تعداد مداخل ما به چیزی حدود سیصد هزار عدد رسید و بعد از آن، زمانی که مداخل اعراب دار را تجمیع کردیم و با مداخل بدون اعراب سازگاری دادیم، به حدود صدویست و هفت هزار مدخل دست یافتیم.

آیا ساختار ارتباطی که هم‌اکنون در بین مداخل و توصیفات کتب لغوی در نظر گرفته شده، قابل تعمیم به سایر اطلاعات غیر لغوی نیز هست؟

بله؛ فارغ از بحث لغت، در ساختار، به یک‌سری مداخل رسیده‌ایم که بسیار محوری هستند و همه لغات، ذیل این مداخل پدر قابل توصیف هستند؛ به بیان دیگر، از موضوع مداخل محوری در لغت‌نامه، به سوی ایجاد یک فرهنگ‌نامه واژگان پیش می‌رویم و همه معانی یک واژه را که در منابع مختلف آمده، به صورت دسته‌بندی شده و یکجا در دسترس قرار می‌دهیم. این مسئله، در مرکز شدنی است و به توفیق الهی و مدد دوستان و همکارانمان، قدم‌های اولیه آن را برداشته‌ایم و پذیرش خوبی هم برای این قضیه ایجاد شده است؛ چون خروجی این کار، برای کاربران مختلف با علایق متنوع، کاربردهای بسیار ارزشمندی دارد. نظیر این قابلیت را شما در ویکی‌ها مشاهده می‌کنید؛ برای مثال، وقتی در ویکی پدیا واژه‌ای را جست‌وجو می‌نمایید، معانی و توضیحات متعددی را در حوزه‌های مختلف به صورت دسته‌بندی شده، درباره موضوع مورد نظرتان مشاهده می‌کنید؛ یعنی جست‌وجوگر ویکی پدیا، اطلاعات مفیدی را به طور یکجا برایتان جمع می‌کند.

آیا در خصوص لغت‌نامه علاوه بر ارائه مداخل و توصیفات، اطلاعات دیگری نیز در اختیار کاربر قرار می‌گیرد؟

برخی از کتب، به بیان نحوه ارتباط بین مداخل مبادرت ورزیده‌اند؛ مثلاً کتبی همچون «المکنز العربی» و «المعجم المفصل فی المترادفات» به کلمات مترادف پرداخته‌اند و در مقابل یک مداخل، چند کلمه آورده و می‌گویند اینها مترادف آن واژه است. در خصوص این گونه کتب، پس از استخراج ماشینی کلمات مترادف و حفظ رابطه ترادف بین کلمات ذکر شده، اقدام به یافتن این کلمات از میان مداخل نماینده خود کرده‌ایم. از این‌رو، در ذیل برخی مداخل نماینده، علاوه بر ارائه توصیفات لغوی، می‌توان مترادفات آن مداخل را هم عرضه کرد.

این قابلیت، در حال حاضر در وبگاه قاموس که در دست تکمیل و راه‌اندازی است، پیاده‌سازی شده است و زمانی که شما کلمه‌ای را جست‌وجو می‌کنید، علاوه بر معنا یا توصیف آن، مترادفات آن را نیز گزارش می‌دهد. در بخش مترادفات، بین دو یا چند مداخل ارتباط معنایی صورت گرفته و

در این روند، ما مداخلی که درای «ال» تعریف بودند را با مداخل فاقد «ال»، جمع نمودیم و همچنین، مداخل فاقد اعراب را به واسطه توضیحاتی که داشت، به مداخل نماینده اعراب‌دار متصل کردیم؛ مثلاً «ضرب» بدون اعراب، به دو شکل ممکن است تلفظ شود؛ مصدر و فعل ماضی. به کمک توضیحات جلوی هر کدام، متوجه شدیم که این مداخل، باید مصدر یا فعل باشد و از این راه، آنها را در جای خودشان جمع نمودیم. بدیهی است که برخی از مراحل و بازیابی آنها، نیاز به کار دستی داشت که در نهایت، به هشتادونه هزار مداخل دست یافتیم؛ یعنی از پانصد هزار مداخل، به حدود نود هزار مداخل پدر رسیدیم که البته جلوی مداخل پدر ما، انواع توضیحات مرتبط و نیز ارجاعات به منابع مختلف وجود داشت که مثلاً فلان مداخل، در منابع گوناگون چه توضیحاتی دارد و در چه منابعی از کتب لغت آمده است.

ردیف	پدر	مداخل پدر	منبع	مداخل	توصیف
۵	افق	الأفق	السنه	الأفق	الجماعة
۶	لذ	الأذ	السنه	الأذ	من أفعال بالذ و السند
۷	احس	أحس	السنه	أحس	
۱۱	اد	أد	السنه	أد	وصلة و ارضی و حیوانات
۱۲	دی	أدی	السنه	أدی	لما د
۱۳	سب	أسب	السنه	أسب	سب و اسب
۱۴	هم	أهم	السنه	أهم	من أو رأيت
۱۵	عق	أعق	السنه	أعق	عق و عاق
۱۶	فر	أفر	السنه	أفر	أف و فر
۱۷	من	أمن	السنه	أمن	أمن و امن
۱۸	نن	أنن	السنه	أنن	نن و نن
۱۹	نوب	أنوب	السنه	أنوب	نوب و نوب

شمهال سازی مداخل

قدم‌های بعدی در تکمیل روند لغت‌نامه نور چه خواهد بود؟ در نسخه‌های گذشته لغت‌نامه نور، ما متن هر کتاب را داشتیم و فهرستی از تمام ریشه‌ها و مشتقات آن را عرضه نموده بودیم؛ ولی در حال حاضر، مداخل هفت منبع اصلی لغت، معیار قرار گرفته است و همه ارجاعات به این مداخل اصلی انجام شده است. به همین جهت، ممکن است برای هر مداخل، چندین توصیف و معنا از منابع مختلف داشته باشیم. ما می‌کوشیم که محققان، همه معانی یک لغت را یکجا ملاحظه کنند. کاری که ما باید در آینده انجام دهیم، این است که این توصیفات متعدد مداخل را دسته‌بندی معنایی کنیم تا استفاده بهتری برای پژوهشگران داشته باشد. این، کاری است که در مرحله بعد انجام می‌دهیم و الآن مقدماتش را انجام داده‌ایم؛ ولی هنوز به نتیجه دلخواه نرسیده‌ایم.

دایره تحقیق کاربر را توسعه داده است؛ یعنی مداخل همسنخ با هم در یکجا جمع شده‌اند. این هم یکی از کارهایی است که در حوزه مترادفات انجام شده است.

اما در خصوص مرتبطات باید عرض کنم که برخی از کتب لغت، مانند الإفصاح، مداخل را به صورت موضوعی و دسته‌بندی شده ارائه کرده‌اند و نه بر اساس ریشه؛ یعنی ممکن است دو مدخل با یکدیگر مترادف نباشند، ولی از نظر موضوعی به هم مرتبط باشند. در این گونه موارد، مداخل مرتبط با مدخل مورد نظر را جمع کرده‌ایم تا کاربر به همه واژگان مرتبط با کلمه مورد نظر خویش دسترسی داشته باشد؛ به عنوان مثال، در کتاب «الإفصاح» ذیل موضوع «تغذیة الولد»، مداخل: «الغذاء»، «البناء»، «العائلة»، «اللذود»، «التَّرف»، «السَّرهقة»، «القرم» و «المعاجة» آمده است که هر مدخل، توصیف خاص خود را دارد و متفاوت از سایرین است؛ ولی تمامی این مداخل که هم‌ریشه هم نیستند، در مفهوم «تغذیة الولد» با یکدیگر دارای «ارتباط» هستند و چه بسا چنین ارتباطی، گره از کار محقق باز کند و او را به هدفش نایل گرداند.

در مورد کلمات متضاد نیز وضع به همین منوال است و کتبی در این باره تألیف شده است و اگر شما مدخلی را انتخاب کنید، مداخل و واژگانی که با مدخل انتخابی شما متضاد باشد، ارائه خواهد شد. روشن است، کلمات متضاد، جایگاه مهمی در تحقیقات دارد و به شفاف‌سازی هرچه بیشتر مفاهیم و اصطلاحات مورد پژوهش کاربر کمک می‌کند.

آیا علاوه بر ارائه مترادفات، مرتبطات و متضادات، اطلاعات جانبی دیگری نیز در اختیار کاربر قرار می‌گیرد؟ در برخی از منابع لغوی، تمرکز اصلی بر روی آیات قرآن کریم و احادیث بوده و در ذیل آنها اقدام به توضیح و توصیف واژگان می‌کنند؛ مثلاً مجمع البحرین، مفردات راغب و نه‌هایه ابن‌اثیر، این خصوصیت را دارند؛ یعنی به واژگان غریبی پرداخته‌اند که در قرآن کریم یا منابع روایی آمده است و نیازمند توضیح هستند. برای اهل لغت یا اهل عرب، مفهوم برخی لغات که در متن قرآن و روایات آمده، مشخص نبوده است. این نوع کتب، به توصیف چنین واژگانی پرداخته‌اند؛ یعنی محوریت در آنها، کلمات قرآن و حدیث است.

در خصوص این کتب، ما آیات و روایات را در حکم مدخل گرفتیم و با استخراج «مدخل پدر» که کلمه محوری در عبارت

است، سعی کرده‌ایم این آیات و روایات را با مداخل نماینده ارتباط دهیم. با این کار، پس از درخواست کاربر از توصیف یک مدخل، ضمن ارائه مترادفات، متضادات و مرتبطات، می‌توانیم آیات و روایاتی را که مشتمل بر آن مدخل هستند نیز عرضه کنیم.

با توجه به تعریفی که برای مدخل نماینده داشتید، آیا می‌توان کلیه اطلاعاتی را که مدخل محور هستند نیز به این مجموعه اضافه کرد؛ هرچند دارای توصیف لغوی نباشند؟

بله؛ به نظر می‌رسد تمامی اطلاعات مدخل محور، همچون: فرهنگ اصطلاحات، فرهنگ اعلام اشخاص، اماکن و... را که دارای چنین ساختاری هستند، می‌توان به این مجموعه اضافه نمود و بعد از تولید «لغت‌نامه»، آرام‌آرام به سوی تولید «فرهنگ‌نامه» حرکت کرد.

نمونه این کار را به صورت آزمایشی روی کتاب «التعريفات» جرجانی انجام دادیم. این کتاب، مشتمل بر اصطلاحاتی است که توضیح هر کدام در مقابلش آمده است.

در این اصطلاح‌نامه، ما مدخل پدر برگرفته از اصطلاح را با فهرست نودهزار تایی خود مرتبط کردیم و مداخل مربوط به این اصطلاح را که عمدتاً تک‌کلمه‌ای بودند نیز در دسترس قرار دادیم.

بعد از گذشت حدود سی سال
از کار مرکز و زحماتی که در
جمع‌آوری و دیجیتال‌کردن منابع
صورت گرفته، باید این منابع
ساماندهی مناسب پیدا کند و
به تسریع و تسهیل دسترسی
کاربران و محققان به متون
اسلامی و انسانی بینجامد

هماهنگ سازی ریشه‌ها در منابع لغوی:

بحمدالله، تاکنون توانسته‌ایم بحث هوشمندسازی لغات و ارائه لغت‌نامه استاندارد نور را به این نقطه برسانیم و نیازمند آن هستیم که به تقویت بیشتر جست‌وجوگرهای خود پردازیم تا بهترین و نزدیک‌ترین تحلیل را به درخواست کاربر ارائه دهد. در این زمینه، کارهای اساسی و خوبی در دو حوزه پژوهشی و فنی در دست انجام است که ان‌شاءالله هرچه زودتر ثمره آن در معرض استفاده همگان قرار بگیرد. ■

لغت‌نامه استاندارد نور در چه پروژه‌هایی از مرکز قابل استفاده است؟

آنچه عرض کردم، دورنمایی از کار ما در این قسمت بود و به مرور زمان، می‌توانیم گروه‌های پژوهشی دیگر مرکز را هم در تکمیل این اطلاعات به کمک بگیریم. به طور کلی، هر پروژه‌ای در مرکز که مرتبط با حوزه واژگان و نیازمند ارائه توصیفات و معانی کلمات است، به این لغت‌نامه نیاز خواهد داشت. عرض بنده این است که با کاری که در حوزه لغت شروع شده، امیدواریم گروه‌های مختلف پژوهشی مرکز، متناسب با تخصص خودشان، به این بانک لغوی مرتبط شوند و به مرور این بانک را توسعه دهند و ما کم‌کم به سوی تشکیل یک فرهنگ‌نامه جامع لغت برویم که مشتمل بر انواع اطلاعات و محتوای لازم در زمینه‌های مختلف علمی باشد.

در پایان اگر سخنی باقی مانده، بیان نمایید.

بعد از گذشت حدود سی سال از کار مرکز و زحماتی که در جمع‌آوری و دیجیتالی‌کردن منابع صورت گرفته، باید این منابع ساماندهی مناسب پیدا کند و به تسریع و تسهیل دسترسی کاربران و محققان به متون اسلامی و انسانی بینجامد؛ البته این موضوع، چیزی نیست که به گروه لغت و ادبیات منحصر باشد؛ بلکه مربوط به کارهای زیرساختی و کلان مرکز است و این هدف، تنها به برکت هوشمندسازی واژگان و ارائه آنها در یک بانک اطلاعاتی منسجم و یکپارچه دست‌یافتنی خواهد بود.