

بازآفرینی هوشمند واژگان در فضای دیجیتال

گذری بر مراسم رونمایی پایگاه لغتنامه برخط قاموس نور



پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

به کوشش: هیئت تحریریه ره آورد نور

اشاره

مهرماه سال جاری، در مؤسسه لغتنامه دهخدا مراسم رونمایی از پایگاه لغتنامه برخط قاموس نور به شکل مجازی (پخش زنده) و با حضور حداقلی میهمانان، پژوهشگران زبان و ادبیات فارسی برگزار شد. سخنرانان این مراسم، دکتر محمود بیجن خان، رئیس مؤسسه لغتنامه دهخدا و مرکز بین المللی آموزش زبان فارسی، و دکتر محمدحسین بهرامی، رئیس مرکز تحقیقات کامپیوتری علوم اسلامی و نیز حجت الاسلام والمسلمین مسیح توحیدی، متکفل پروژه بودند که در این آیین رونمایی، مطالبی را در حوزه زبان فارسی و قابلیت‌های وبگاه قاموس نور بیان نمودند.

دکتر محمود بیجن خان

«بحث لغت‌نامه را با توجه به شرایطی که در جامعه علمی کشور داریم و از دیدگاه زبان فارسی، به عنوان زبان علم می‌توانیم مطرح کنیم. اساساً هر کشوری - مانند ایران - که تاریخ، ادبیات، فلسفه، فرهنگ و تمدن دارد، لاجرم باید یک فرهنگ لغت غنی داشته باشد و دسترسی به این فرهنگ غنی نیز باید برای همه بسیار ساده و آسان فراهم شود. این دسترسی ساده، از دو جهت بسیار مهم است: یکی از جهت واژه‌سازی، و دیگری از جهت رفع ابهام از معنای کلمات که ما از آن به ابهام‌زدایی معنایی یاد می‌کنیم.

زبان‌ها، به یکدیگر زیاد قرض می‌دهند؛ برای مثال، ما از زبان انگلیسی کلمات بسیاری قرض گرفته‌ایم و یا زبان فارسی کلمات قابل توجهی را به زبان عربی قرض داده و برعکس. اساساً چه بسا پیشینه صادرات و واردات کلمات، بیشتر از صادرات و واردات مرسوم در دنیا باشد و این، به معنای انتقال فکر است.

بنابراین، هرچه یک ملت یا کشور، واژه بیشتری در اختیار داشته و تصرف افزون‌تری در جهان خارج ذهن کرده باشد و بتواند آنها را با روش‌هایی به ملت‌های دیگر انتقال دهد، آن تمدن و فرهنگ و فکر، جایگاه علمی خود را در جهان نشان خواهد داد.

ما در حال حاضر، در واژه‌سازی بیشتر حالت انفعالی داریم و کلماتی را که وارد می‌شوند، معادل‌سازی می‌کنیم؛ نه اینکه یافته‌های علمی خود را تبدیل به واژه کنیم و تلاش نماییم دیگران از این واژه‌ها استفاده کنند و به تدریج استفاده زیاد از این واژه‌ها، به استفاده از عبارت و حتی ترکیبات معنایی تبدیل شود.

متأسفانه، از این نظر، در واژه‌سازی عقب هستیم. شاید به این دلیل که ما علم را اصیل تولید نمی‌کنیم و یا به عبارت دیگر، کمتر علم اصیل تولید می‌نماییم تا همه چیز آن، دست خودمان باشد. خوشبختانه، فرهنگستان زبان و ادب فارسی این واژه‌ها را تبدیل به فارسی کرده و ما از این جنبه که زبان فارسی، زبان علم باشد، عقب نمی‌افتیم؛ اما اگر قرار بود همین کار هم اتفاق نیفتد، آنگاه مجبور می‌شدیم مانند بسیاری از کشورها که اصلاً زبانشان را از دست داده‌اند، زبانمان را از دست بدهیم؛ برای مثال، در حال حاضر، انگلیسی سنگاپوری، انگلیسی هنگ‌کنگ و انواع دیگری، نظیر انگلیسی استرالیایی و انگلیسی آمریکایی داریم؛ که البته این آخری، خیلی قدیمی است؛ یعنی کشورها زبان اصلی خود را از دست می‌دهند و در عوض، انگلیسی آن کشور جایگزین شده و به گونه‌های زبانی انگلیسی تبدیل می‌شوند.

نکته دوم، بحث ابهام‌زدایی معنایی است که شاید خیلی از قسمت اول مهم‌تر باشد. در قسمت اول، اگر ما حالت فعال داشته باشیم، وضعیت خوبی خواهیم داشت؛ اما در قسمت دوم، ابهام‌زدایی معنایی، به این معنی که لازم است معنای واژه‌هایی را که در طول تاریخ در متون ما وجود دارند، مشخص کنیم. این کار، هم شناخت ما را از فرهنگمان بیشتر می‌کند و هم می‌توانیم از آنها در واژه‌سازی استفاده نماییم؛ یعنی در فرآیند واژه‌سازی که می‌خواهیم واژه‌های اصیل خود را به فارسی ترجمه کنیم و به سایر نقاط دنیا نشر دهیم، اهمیت بسیاری دارد.

چند روز پیش کلمه «بار» را بررسی می‌کردم. دیدم چقدر معناهای متفاوتی دارد. دیروز در قاموس نور نگاه می‌کردم، دیدم ۲۸۵ معنا برای «بار» آورده شده؛ اما آن معنایی را که ما در یک فرهنگ لغت قدیمی دیده بودم و «بار» را «اجاق» معنا کرده بود، در قاموس نور ندیدم. البته کلمات فراوانی

دکتر بیجن خان:

زبان فارسی به عنوان زبان علم، با واژه‌سازی و ابهام‌زدایی معنایی ارتباط مستقیم دارد و به همین جهت، جایگاه و ارزش کارهایی مثل «قاموس نور» اینجا معلوم می‌شود؛ چرا که گسترش زبان فارسی و گسترش یافته‌های علمی خودمان در قالب کپسول‌های واژگانی و صادرات آنها، از نظر توسعه فرهنگ و تمدن ایرانی اهمیت بسزایی دارد؛ به خصوص اینکه ما در سال‌های بعد از انقلاب، تحول و پیشرفت قابل توجهی نیز داشته‌ایم

بود که ترکیبات آنها را می‌شد گفت معنای «اجاق» را دارد؛ ولی اینکه صراحت داشته باشد معنای این کلمه، اجاق است، وجود نداشت؛ مانند اصطلاحاتی که ما الآن به کار می‌بریم و مثلاً می‌گوییم «آبگوشت را بار گذاشت»؛ یعنی اینکه یک وسیله‌ای گذاشتیم گرم شود؛ هر غذایی را می‌شود بار گذاشت.

یکی از کاربردهای دانستن معنای کلماتی که در اختیار داریم، استفاده در واژه‌سازی است؛ مثلاً فردا اگر چیزی را ساختیم که کارش گرم کردن یک وسیله بود، می‌توانیم از کلمه «بار» به عنوان یک کلمه اصیل فارسی استفاده کنیم و بعد، این کلمه می‌تواند توسعه پیدا کند و خیلی از جاها از آن استفاده شود و یا آن را به زبان‌های دیگر نیز قرض دهیم.

بنابراین، زبان فارسی به عنوان زبان علم، با واژه‌سازی و ابهام‌زدایی معنایی ارتباط مستقیم دارد و به همین جهت، جایگاه و ارزش کارهایی مثل «قاموس نور» اینجا معلوم می‌شود؛ چرا که گسترش زبان فارسی و گسترش یافته‌های علمی خودمان در قالب کپسول‌های واژگانی و صادرات آنها، از نظر توسعه فرهنگ و تمدن ایرانی اهمیت بسزایی دارد؛ به خصوص اینکه ما در سال‌های بعد از انقلاب، تحول و پیشرفت قابل توجهی نیز داشته‌ایم.

در اینجا لازم است درباره لغت‌نامه دهخدا نیز توضیحاتی عرض کنم. در لغت‌نامه ۱۵ جلدی دهخدا که یکی از منابع قاموس نور است، از سایر فرهنگ‌های فارسی موجود در «قاموس نور» نیز استفاده شده است؛ مانند: فرهنگ نفیسی، فرهنگ رشیدی و برهان قاطع که همگی در لغت‌نامه دهخدا وجود دارند؛ اما نسخه کامل این فرهنگ‌ها نیز در «قاموس نور» وجود دارد. در لغت‌نامه دهخدا، بسته به تشخیص مؤلف، از معادل فارسی موجود در این فرهنگ‌نامه‌ها استفاده شده است.

نکته مهمی که در خصوص لغت‌نامه دهخدا لازم است عرض کنم، این است که به لحاظ حجم مطالبی که در لغت‌نامه درج شده، به اعتقاد بنده، یکی از آثار بی‌نظیر علوم انسانی در دوره فارسی معاصر است.

این لغت‌نامه، ۲۳۹۱۱ صفحه سه‌ستونی است که هر ستون آن ۵۲ سطر و در مجموع، حدود ۲۲ میلیون و ۴۰۰ هزار کلمه را دربر گرفته است و در واقع، از مجموعه‌ای از فرهنگ‌ها تهیه شده است. روش تولید این لغت‌نامه، از طریق نخبه‌سپاری و توسط ۷۰ نفر از نخبگان زبان و ادبیات فارسی

بوده است.

در تولید لغت‌نامه دهخدا، از فرهنگ برهان قاطع، فرهنگ رشیدی، آندراج و منتهی‌العرب و تعدادی بسیاری از فرهنگ‌هایی که در کتابخانه مرحوم دهخدا وجود داشتند و مؤلفان و نخبگان از آنها برای نوشتن و آوردن مواردی چون شاهد در لغت‌نامه، استفاده کرده‌اند.

خوشبختانه، ادامه این کار، در قالب لغت‌نامه بزرگ فارسی در جریان است و ان‌شاء‌الله بتوانیم بعد از تصویب در شورای مؤسسه دهخدا، این لغت‌نامه بزرگ را نیز به «قاموس نور» اضافه کنیم تا کاستی‌هایی که در لغت‌نامه دهخدا هست، جبران شود.

لغت‌نامه دهخدا، هم به لحاظ زمانی که تا قرن هشتم را دربرمی‌گیرد و هم به لحاظ شواهد و معادل‌های فارسی، باید کامل شود که این نواقص، در لغت‌نامه بزرگ فارسی رفع شده و تا فارسی معاصر امروز البته با یک محدودیت‌هایی، اضافه شده‌اند.

من خیلی خوشحالم که امروز در خدمت بزرگان مرکز تحقیقات کامپیوتری علوم اسلامی هستیم.»

سخنرانی دکتر محمدحسین بهرامی

«از حُسن میزبانی جناب آقای دکتر بیجن‌خان و مجموعه دست‌اندرکار مؤسسه لغت‌نامه دهخدا برای برگزاری مشترک مراسم رونمایی از «سامانه لغت‌نامه نور» با عنوان قاموس، تشکر می‌کنم. خوشحال هستم که در خدمت شما عزیزان و بزرگواران هستم و تشکر و خیر مقدم دارم خدمت عزیزان حاضر در جلسه و همین‌طور بزرگوارانی که به صورت مجازی این مراسم را دنبال می‌کنند. امیدواریم با برطرف شدن مشکلات، این جلسات پُررونق‌تر برگزار شود؛ هرچند گفته می‌شود دیگر به آن جایگاه قبلی باز نخواهد گشت؛ اما امیدواریم که حتی الامکان، این امکان فراهم شود.

ما در مرکز تحقیقات کامپیوتری علوم اسلامی، موظف به فراهم کردن زیرساخت‌های پژوهشی برای کسانی هستیم که در حوزه علوم اسلامی، مشغول تحقیقات‌اند. در همین راستا، طی حدود سه دهه گذشته، موفق شده‌ایم حجم قابل توجهی - که حتی می‌شود گفت حجم قریب به اتفاق - از منابع اسلامی در رشته‌های مختلف علوم اسلامی را به نسخه دیجیتال و نرم‌افزارهایی در قالب‌ها و حامل‌های مختلف نظیر محصولات رومیزی، وبگاه‌ها و اپ‌های موبایل تبدیل نماییم و در اختیار پژوهشگران قرار دهیم.

در راستای عمل به این وظیفه، یکی از حوزه‌هایی که به صورت جدی از ابتدا مد نظر مرکز تحقیقات کامپیوتری علوم اسلامی بوده، مباحث مربوط به زبان، ادب و لغت فارسی است. همان‌طور که جناب آقای دکتر بیجن‌خان به‌درستی اشاره نمودند، به عنوان بستر و حامل، بخش عمده‌ای از علوم اسلامی در این فضا شکل گرفته و پژوهش‌های علوم اسلامی، به‌خصوص در قالب دیجیتال آن، به جز با تحقیق و فراهم کردن زیرساخت‌های مرتبط با لغت و ادبیات، قطعاً نتیجه مطلوب را ارائه نخواهد داد.

از این جهت، در کنار سایر رشته‌های علوم اسلامی، بحث ادبیات عرب و ادبیات فارسی به صورت خاص مد نظر بوده و اقداماتی در این زمینه انجام شده است.

در مرکز تحقیقات کامپیوتری علوم اسلامی، نرم‌افزارهای فاخری در حوزه ادب فارسی تولید شده است؛ از جمله نرم‌افزارهایی برای برخی از بزرگان و بعضی آثار تولید گردیده است؛ مانند نرم‌افزار:

حافظ، شاهنامه فردوسی، مثنوی معنوی و پروین اعتصامی. در واقع، ما نسخه‌های متفاوت این آثار را گردآوری کرده، به نسخه دیجیتال تبدیل کرده‌ایم. علاوه بر این، حواشی، شروح و اطلاعات پیرامونی قابل توجهی گردآوری شده و هم نسخه‌های صوتی بسیار ارزشمندی با صدای اساتید و صدایشگان بزرگوار و بنام، تهیه شده است؛ مثل مرحوم آهی که به صورت کامل، حافظ را قرائت نموده‌اند و یا جناب آقای ساعد باقری که زحمت قرائت اشعار عطار را کشیده‌اند و یا سرکار خانم ژاله علو و سرکار خانم بی‌نیاز که کلیات پروین اعتصامی را قرائت کرده‌اند. همه اینها با کیفیت بسیار بالایی فراهم شده است.

نرم‌افزار «کتابخانه نظم و نثر فارسی»، از دیگر کارهای جامع و فاخری است که در حال شکل گرفتن است. بخش عمده‌ای از آن، به اتمام رسیده و البته مشکلات مالکیت معنوی، مثل هر حوزه دیگری، به‌خصوص در حوزه زبان فارسی، برای ما وجود دارد. متأسفانه، هنوز این فرهنگ در برخی از عزیزان غلبه پیدا نکرده که مجوز نشر دیجیتال آثار خود را به ما بدهند و ما نیز به رعایت این حقوق، بسیار حساس هستیم. امیدواریم به‌زودی این نرم‌افزار را عرضه کنیم.

پیکره‌های تخصصی که در حوزه زبان فارسی توسط معاونت تهران مرکز نور به صورت ویژه برچسب‌گذاری و تولید شده، شامل: پیکره صرفی، پیکره نحوی، پیکره گفتمانی یا سایر حوزه‌هایی است که دوستان زحمت کشیده‌اند و امروز در دسترس پژوهشگران است.

در مجموع، به برکت آثار فارسی که در مرکز تحقیقات کامپیوتری علوم اسلامی گردآوری شده، امروزه یک پیکره دو میلیارد کلمه‌ای از کتاب‌ها و مجلات متقدم و متأخر داریم که با دقت بالا و قابل قبول و مورد اعتماد فرهیختگان این حوزه فراهم آمده و با شیوه خاص و ابداعی مرکز، به نسخه دیجیتال تبدیل شده‌اند. این پیکره، قریب به دو میلیون شکل کلمه فارسی را به صورت غیرتکراری دارد که اینها به یک نسبتی، برچسب‌های قابل توجهی نظیر برچسب زمانی و برچسب حوزه دانشی به همراه دارند و می‌توان برچسب‌های دیگری مانند برچسب جغرافیا را نیز بدان اضافه کرد.

پیکره متناظر، حدود ۶۰۰ کتاب فارسی و عربی را مشتمل می‌گردد که برخی از اینها در سطح کلمه،

دکتر محمدحسین بهرامی:

به برکت آثار فارسی که در مرکز تحقیقات کامپیوتری علوم اسلامی گردآوری شده، امروزه یک پیکره دو میلیارد کلمه‌ای از کتاب‌ها و مجلات متقدم و متأخر داریم که با دقت بالا و قابل قبول و مورد اعتماد فرهیختگان این حوزه فراهم آمده و با شیوه خاص و ابداعی مرکز، به نسخه دیجیتال تبدیل شده‌اند. این پیکره، قریب به دو میلیون شکل کلمه فارسی را به صورت غیرتکراری دارد که اینها به یک نسبتی، برچسب‌های قابل توجهی نظیر برچسب زمانی و برچسب حوزه دانشی به همراه دارند و می‌توان برچسب‌های دیگری مانند برچسب جغرافیا را نیز بدان اضافه کرد

بعضی در سطح فراز و برخی در سطح جمله یا حداکثر در سطح پاراگراف به هم متصل شده‌اند و در بسیاری از کارهای ماشینی مانند بحث ترجمه ماشینی و مانند آن، می‌تواند قابل استفاده باشد.

پیکره اشعار درون‌متنی که ما در اختیار داریم، با توجه به اینکه نگارش حجم وسیعی از کتاب‌ها مربوط به قرون ابتدایی است و در لابه‌لای متن کتب، اشعاری مورد استناد قرار گرفته‌اند و یا نویسنده خود تولید یا انشا کرده و در کتاب خودش آورده، در مجموع، گنجینه ارزشمند و بی‌نظیری را فراهم آورده است.

امروز، وبگاه «قاموس نور»، کار جدیدی است که در اختیار علاقه‌مندان به حوزه لغت فارسی قرار می‌گیرد. در این سامانه، تعداد ۷ لغت‌نامه فارسی به اضافه تعدادی از دائرةالمعارف‌ها و اصطلاح‌نامه‌های تخصصی ارائه شده است.

مرکز تحقیقات کامپیوتری علوم اسلامی، بعد از دوره‌ای که وظیفه اولیه خود را گردآوری منابع و عرضه آنها در قالب نرم‌افزارها می‌دید، وارد فاز جدیدی در حوزه علوم اسلامی شده که از آن با عنوان «علوم انسانی دیجیتال» یاد می‌شود. با این نگاه که با آمدن فضای دیجیتال، تنها در اختیار قرار دادن نسخه دیجیتالی، اتفاقی نیست که در علوم رخ می‌دهد؛ بلکه فناوری اطلاعات بعضی اوقات، ابزار، روش و حتی تا حدی ماهیت و غایات علوم را تحت تأثیر قرار می‌دهد. علوم انسانی و علوم اسلامی هم به عنوان زیرمجموعه آن، همین تأثیر را از فضای فناوری اطلاعات می‌پذیرند.

مثال مشهود آن، بحث دائرةالمعارف‌ها و دائرةالمعارف‌نویسی است. دائرةالمعارف‌های ما، زمانی ایرانیکا، بریتانیکا و دائرةالمعارف بزرگ اسلامی و مانند اینها بود؛ ولی مسیری که طی شده، امروز مفهوم جدیدی به نام ویکی‌پدیا با یک روش جدید، نگاه تازه و یک دامنه تأثیرگذاری نوین را پدید آورده که لازم است در همه علوم اسلامی و انسانی، این بازتعریف و بازآفرینی علوم را در فضای دیجیتال داشته باشیم.

در حوزه علوم اسلامی، تلاش می‌کنیم محتوای علوم را دیجیتالی نماییم و حتماً این اتفاق، باید در حوزه لغت و ادب فارسی هم روی دهد. ما کارهایی را آغاز کرده‌ایم و ان‌شاءالله در هفته پژوهش سال جاری، «آزمایشگاه هوش مصنوعی و علوم انسانی - اسلامی دیجیتال» مرکز نور را افتتاح خواهیم کرد.

فازهای آینده کار قاموسی که در حال انجام هست - چه در حوزه لغت فارسی و چه در حوزه لغت عربی - به سوی تبدیل شدن به عنوان یک بازآفرینی از لغت‌نامه‌ها پیش می‌رود. شاید خروجی قاموس در نگاه اول، کار جدیدی باشد؛ اما در واقع، بهره‌گیری ماشینی از لغت‌نامه‌های پیشین است.

امیدواریم با همکاری خوبی که با مجموعه دهخدا داریم و لطفی که این مجموعه به مرکز نور دارند و کارهایی که تاکنون انجام شده و اعلام آمادگی هر دو طرف برای اجرای کارهای مشترک، شاهد پیشرفت‌های خوب و رویدادهای قابل توجه، مفید و تأثیرگذاری در این حوزه باشیم.

در پایان عرایضم، مجدداً از جناب آقای دکتر بیجن خان، جناب آقای دکتر ویسی، جناب آقای تجلی و نیز از تمامی عزیزانی که زمینه برگزاری این مراسم را فراهم کردند، تشکر می‌کنم.

گفتنی است که در بخش پایانی، حجت‌الاسلام والمسلمین مسیح توحیدی به توضیح امکانات و ویژگی‌های پایگاه قاموس پرداخت و در انتهای مراسم نیز از پایگاه لغت‌نامه برخط قاموس نور رونمایی گردید. ■