

کاربرد هوش مصنوعی در پژوهش‌های علوم اسلامی



مهندس احمد ربیعی زاده*
arabiei@noornet.net

اشاره

آنچه در این مقاله از نظر شما خوانندگان محترم می‌گذرد، گزیده‌ای از متن وینار مهندس احمد ربیعی زاده، مدیر محترم اداره پردازش هوشمند مرکز تحقیقات کامپیوتری علوم اسلامی است که در ۲۴ آذرماه ۱۳۹۹ مصادف با هفته پژوهش ارائه شده بود. از آنجا که این وینار، مشتمل بر نکات مفید در خصوص کاربرد هوش مصنوعی در پژوهش‌های علوم اسلامی است، مناسب دیده شد در فصلنامه ره آورد نور انعکاس یابد. طی این وینار، با مقوله هوش مصنوعی و کاربردهای آن در پژوهش‌های علوم اسلامی و نمونه‌هایی از این دستاوردها در تولیدات وبی و نرم‌افزارهای ویندوزی مرکز نور آشنا می‌شوید.

* مدیر اداره پردازش هوشمند مرکز نور.

درآمد

از مجموعه ویدئوهای هفته پژوهش، امروز با ویدئو با موضوع «کاربرد هوش مصنوعی در پژوهش‌های علوم اسلامی» در خدمت شما هستیم. ابتدا در مرحله اول، تعریف و تاریخچه‌ای از هوش مصنوعی را خدمت شما ارائه می‌دهم و سطوح مختلفی از هوش مصنوعی که می‌شود از آنها استفاده کرد را بیان می‌نمایم. در مرحله دوم، مراحل تولید محصولات هوشمند داده‌محور را به طور خلاصه تبیین می‌کنم و مروری نیز بر ملاحظات اجرایی که در عمل با آنها روبه‌رو هستیم، خواهیم داشت. در مرحله سوم ویدئو، سیر تکامل نرم‌افزارهای نور و سطوح مختلف کاربری هوش مصنوعی را در محصولات مرکز نور خدمت شما ارائه می‌دهم و به اهداف به‌کارگیری فناوری هوش مصنوعی اشاره می‌نمایم و در مرحله بعد، نمونه‌هایی از دستاوردهای متعدد مرکز را در این زمینه که طی هشت دسته مرتب شده است، معرفی می‌کنیم.

تاریخچه و تعریف پردازش هوشمند متن

قریب به ده سال است که در مرکز تحقیقات کامپیوتری اسلامی و معاونت فنی، بخشی تحت عنوان هوش مصنوعی و پردازش هوشمند تشکیل شده است. به طور دقیق‌تر، ابتدا بیشتر تمرکز ما روی مباحث متنی بود که از آن به عنوان «متن‌کاوی» یاد می‌کنیم؛ ولی کم‌کم به فراخور نیازهایی که پیش رو داشتیم، به سوی پردازش تصویر و صوت نیز حرکت کردیم و اسم و عنوان این بخش، به بخش پردازش هوشمند تغییر پیدا کرد.

هوش مصنوعی، عبارت است از توانایی ماشینی برای اینکه بتواند رفتاری معادل با رفتار انسان را در مسائل خاص از خودش نشان بدهد. از سال ۱۹۵۰م شخصی به نام «تورینگ» این اصطلاح را بر سر زبان‌ها انداخت. هوش مصنوعی موفق، کاربری تکنیک‌هایی است که شخص ناظر متوجه نشود که این خروجی را از سوی انسان دریافت می‌کند یا از جانب ماشین. اگر چنین چیزی محقق شود، می‌توانیم بگوییم هوش مصنوعی، کار خودش را به‌درستی انجام داده است.

سطوح مختلف هوش مصنوعی

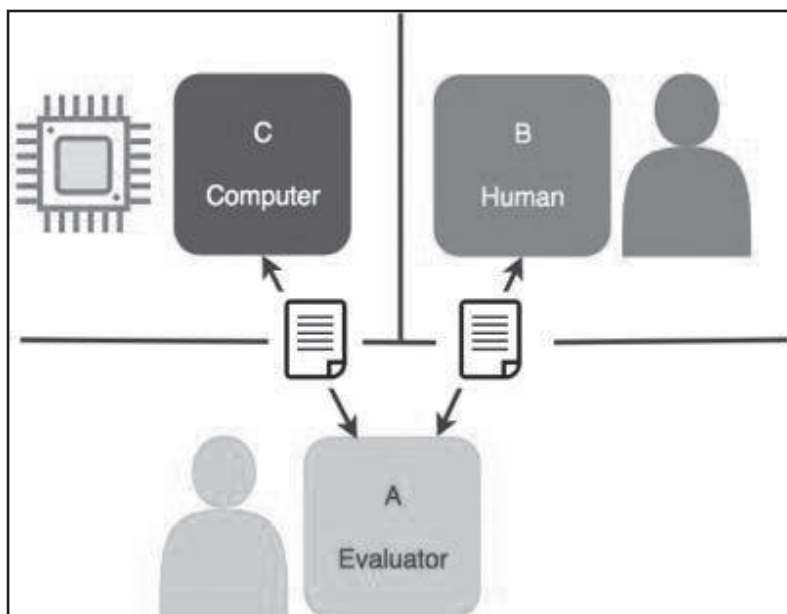
هوش مصنوعی، لایه‌های مختلفی دارد و در سطوح

گوناگونی از پیچیدگی قابل به‌کارگیری است. به طور کلی، مسائل هوش مصنوعی را به دو دسته اصلی تقسیم می‌کنند:

یکی، مسائل قاعده‌مند و قانون‌محور است؛ یعنی مسائلی که قوانین حل آن مسئله از قبل توسط خبره شناسایی شده، به ماشین داده می‌شود و برنامه نیز طبق همان عمل می‌کند؛ به بیان دیگر، قوانینی در این برنامه‌های پیاده‌سازی می‌شوند و سیستم طبق این قوانین از پیش تعریف‌شده، به نتیجه مورد نظر می‌رسد.

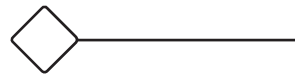
دوم، سیستم‌های هوش مصنوعی مبتنی بر یادگیری ماشین است. در این نوع، ماشین باید طبق ویژگی‌ها و پارامترهای خاص فضای مسئله، تصمیم بگیرد؛ یعنی مثلاً یکسری ویژگی‌های مشخص را در اختیار سیستم قرار می‌دهیم و سیستم مبتنی بر تحلیل این ویژگی‌ها و کشف رابطه آنها با خروجی مورد نظر، دانش خود را به نمونه‌های جدید تعمیم می‌دهد؛ به بیان دیگر، به جای اینکه ما بیاوریم قوانین خاصی را برای سیستم تعریف کنیم، خود سیستم از طریق پارامترهایی که از لایه‌های دیتای آموزشی در اختیارش قرار گرفته، قوانین لازم را استخراج کند و جواب دهد. در واقع، ماشین در اینجا نوعی فرایند یادگیری را طی می‌کند و به اصطلاح، هوشمند عمل می‌نماید.

برای دسته اول، به عنوان مثال، در پروژه‌های خودمان مانند پروژه تحلیل ساختار و صرفی کلمات که مبتنی بر یکسری قوانین زبان‌شناسی تعریف‌شده است، اطلاعاتی را به سیستم آموزش دادیم که خودش یک کلمه را از جنبه‌های مختلف زبانی تحلیل کند و خروجی مناسب ارائه دهد.





هوش مصنوعی، عبارت است از توانایی ماشین برای اینکه بتواند رفتاری معادل با رفتار انسان را در مسائل خاص از خودش نشان بدهد. از سال ۱۹۵۰م شخصی به نام «تورینگ» این اصطلاح را بر سر زبان‌ها انداخت. هوش مصنوعی موفق، کاربرست تکنیک‌هایی است که شخص ناظر متوجه نشود که این خروجی را از سوی انسان دریافت می‌کند یا از جانب ماشین



در قسمت یادگیری ماشین، یک نسل پیشرفته‌تری نیز هست تحت عنوان یادگیری عمیق یا Deep Learning. در این مبحث، ما حتی آن ویژگی‌ها را نیز به سیستم نمی‌دهیم؛ بلکه خود سیستم باید با تحلیل حجم انبوه نمونه‌ها، ویژگی‌های اثرگذار در خروجی را شناسایی کند. معمولاً تعیین این ویژگی‌های اثرگذار، در بسیاری از مسائل، کار دشوار و پیچیده‌ای است؛ به‌خصوص در مسائلی که با متن، محتوا، مباحث طبیعی و انسانی مواجه هستیم. با وجود این پیچیدگی‌ها، ابزارهای فناورانه ما به حدی پیشرفت داشته که می‌توانیم بدون ارائه پارامترهای از پیش تعریف شده به سیستم، به نتیجه مطلوب خود برسیم. در واقع، یک داده حجیم به سیستم می‌دهیم و در آن داده تعیین می‌کنیم که مثلاً موضوع این حدیث، خمس است و موضوع فلان حدیث، صلوات است؛ یعنی با ارائه نمونه‌هایی چند، خود سیستم پشت صحنه آنها را تحلیل می‌کند و متوجه می‌شود که مثلاً چه ویژگی‌هایی از متن می‌تواند در مورد یک خروجی خاص، اثرگذار باشد.

مراحل تولید نرم‌افزار هوشمند داده‌محور

به‌طور کلی، اگر بخواهیم یک محصول هوشمند داده‌محور داشته

باشیم، لازم است ابتدا مرحله داده‌سازی را برای مسئله خودمان پیاده کنیم. یکی از مهم‌ترین قسمت‌ها در تولید نرم‌افزار و حل یک مسئله هوشمند که مبتنی بر محتواسست، همین مرحله است.

بعد از آنکه وضعیت مسئله و داده‌های لازم را از نظر حجم و کیفیت مورد نظر آماده کردیم، نوبت بررسی الگوریتم‌های مختلف و انتخاب الگوریتم مناسب است. آنگاه باید الگوریتم مناسب را روی دیتای اصلی به دفعات متعدد اجرا کرد. پس از این، نوبت به ارزیابی نتایج می‌رسد که مبحث خیلی مهمی است. به منظور محک زدن خروجی کار، لازم است از پیش، دیتای شاهد داشته باشیم تا بتوان خروجی را از نظر دقت یا جامعیت و مانعیت بررسی نمود.

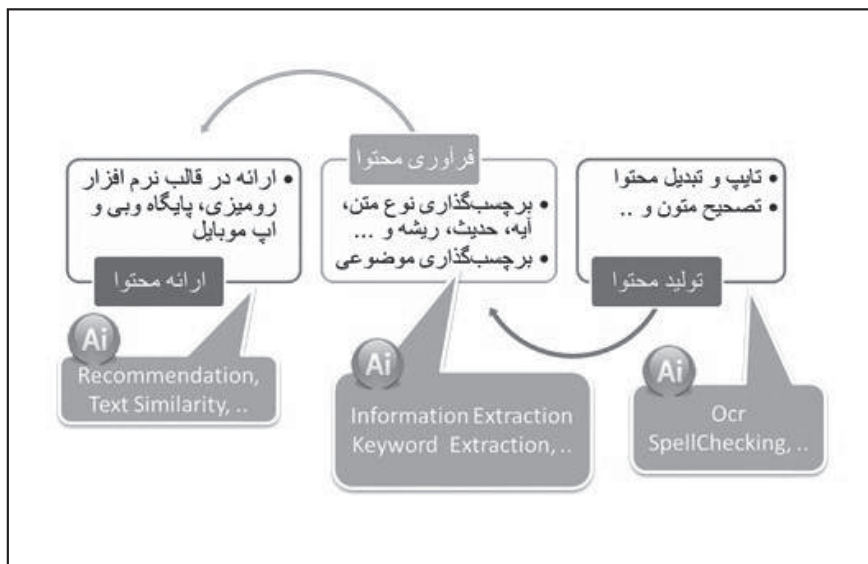
در کنار این، مباحث مربوط به بسترسازی زیرساخت پردازشی نیز می‌تواند به ما کمک کند که با پردازش خیلی سریع‌تر، الگوریتم بهتری را برگزینیم و کیفیت نتایج خودمان را اثبات کنیم. در نهایت، بعد از اینکه به یک خروجی مناسب دست یافتیم و ارزیابی‌های قابل قبولی هم روی آنها انجام دادیم، آنگاه مرحله تولید محصول را خواهیم داشت.

مسائل هوش مصنوعی

برخی نکات در زمینه مفاهیم و مسائل مربوط به هوش مصنوعی وجود دارند که نتیجه تجربه چندین ساله ما در مرکز نورند و خیلی مهم هستند. در اینجا به بخشی از آنها اشاره می‌کنیم. برای تولید محصولات تخصصی مثل علوم اسلامی، با دو گروه از متخصصان مواجه هستیم. یکی، متخصصان توسعه فنی و پردازش هوشمند و دوم، متخصصان یک دانش خاص، همچون: قرآن، حدیث، فقه، تاریخ و... . به‌طور کلی، پروژه‌هایی موفق هستند که در مراحل مختلف کار، ارتباط خیلی نزدیک و تنگاتنگی بین هر دو گروه از متخصصان وجود داشته باشد و این مهم، در مسائل هوشمند که با حل مسئله محتوامحور روبه‌رو هستیم، به‌شدت خودش را نشان می‌دهد و چه بسیار پروژه‌هایی که به جهت فقدان این موضوع، به شکست انجامیده‌اند.

ممکن است تولید یک پایگاه یا نرم‌افزارهای موبایلی با رعایت یکسری چهارچوب‌ها و ویژگی‌ها و داشتن حداقل ارتباط اهل خیره با یکدیگر، امکان‌پذیر باشد؛ اما تحقق پروژه‌های هوشمند در زمینه تحلیل محتوا و تهیه محصولات در این راستا، نیازمند ارتباط قوی و نزدیک بین متخصصان فنی و علمی است.

نکته مهم دیگر اینکه هدف‌گذاری ما باید واقع‌گرایانه باشد. در



سال‌های اخیر، با توجه به موفقیت‌هایی که در هوش مصنوعی با آن روبه‌رو بودیم، یک فضای تبلیغاتی شکل گرفت که باعث شد انتظارات دیگران از متخصصان این حوزه، فراتر از حد متعارف بالا برود و ما از واقع‌گرایی فاصله بگیریم و در عمل، نتوانیم به نتیجه دلخواه برسیم و همین موجب شود در این حوزه، گاهی روح ناامیدی حاکم شود و به همین دلیل، با تصمیماتی هیجانی، خود و جامعه را از کاربردهای هوش مصنوعی محروم کنیم و حتی جلوی پیشرفت‌های بعدی را بگیریم. بنابراین، در مسیر توسعه هوش مصنوعی باید معقول و واقع‌گرایانه هدف‌گذاری کرد.

این روند، ادامه پیدا کرد تا حدود ده سال پیش که ما با یک موضوع جدید و سومی روبه‌رو شدیم و آن، حوزه به‌کارگیری تکنیک‌های محاسبات ریاضی و محاسبات آماری و پردازش‌های زبانی در کنار آن دو حوزه قبلی بود. با استفاده از این حوزه سوم، توانستیم مباحث مربوط به دستاوردهایی در زمینه Text Mining یا داده‌کاوی و متن‌کاوی و پردازش زبان‌های طبیعی یا Natural Language Processing و به طور کلی هوشمندسازی فرایندها، قابلیت‌هایی را برای کاربران و محققان به ارمغان بیاوریم که قبل از آن وجود نداشته است؛ یعنی نسل جدیدی از نرم‌افزارهای نور را تولید کردیم که از قابلیت‌های هوشمند بهره‌مند هستند.

به طور کلی، مراحل تولید نرم‌افزار در مرکز نور، در سه مرحله خلاصه می‌شود:

مرحله اول، تولید محتواست. تایپ و تبدیل محتوا و تصحیح متون

ابتدا باید دید هوش مصنوعی با توجه به سرعتی که شاهد آنیم، چه آورده و دستاوردهایی را می‌تواند برای ما به ارمغان بیاورد. بعد از آن، به طراحی پروژه خودمان بپردازیم تا بتوانیم به صورت حداکثری از این وضعیت‌ها استفاده کنیم و نیاز کاربران را به شکل مطلوبی پاسخ دهیم. بنابراین، در همان دو طراحی یک پروژه باید به مقوله هوش مصنوعی توجه ویژه داشت.

سیر تکاملی نرم‌افزارهای نور

مرکز تحقیقات کامپیوتری علوم اسلامی، از سال ۱۳۶۸ش تأسیس شد و از آن سال‌ها تاکنون، یکی از دستاوردهای نور این بوده است که توانسته فناوری اطلاعات و محتوای علوم اسلامی را با هم گره بزند و از تلفیق این دو دانش با همدیگر، محصولات نرم‌افزاری منحصر به فردی را در حوزه علوم اسلامی تولید کند.

اگر بخواهیم یک محصول هوشمند داده‌محور داشته باشیم، لازم است ابتدا مرحله داده‌سازی را برای مسئله خودمان پیاده کنیم. یکی از مهم‌ترین قسمت‌ها در تولید نرم‌افزار و حل یک مسئله هوشمند که مبتنی بر محتواست، همین مرحله است. بعد از آنکه وضعیت مسئله و داده‌های لازم را از نظر حجم و کیفیت مورد نظر آماده کردیم، نوبت بررسی الگوریتم‌های مختلف و انتخاب الگوریتم مناسب است





برخی پروژه‌ها به دلیل همین دستی بودن خیلی از مراحل، بیش از ده سال در صف توسعه محتوا باقی می‌ماندند؛ مانند پروژه موضوع‌گذاری روی کتاب تفسیر المیزان یا بحار الأنوار و الغدير. بدیهی است که ورود هوش مصنوعی به این حوزه، می‌تواند نقش مهمی در تسریع و پشبرد کمیّت کار داشته باشد.

هدف بعدی، ارتقای سطح کیفی محتوا و خدمات ارائه‌شده است. با تکیه بر پردازش‌های هوشمند و استفاده از تکنیک‌های جدید می‌توان سطح کیفی خدمات خود را ارتقا داد و اشکال‌هایی را که ممکن بود در کار انسانی پیش بیاید، حل کنیم.

هدف دیگر اینکه با قابلیت‌های برآمده از هوش مصنوعی، نیازهای پیچیده‌ای را می‌توانیم برطرف کنیم که تا قبل از آن، اصلاً امکان آنها وجود نداشت؛ مثلاً در پروژه سمیم فرایند مشابه‌یابی مقالات در سطح خیلی حجیم و وسیعی انجام می‌شود؛ یعنی این سامانه، یک مقاله را با تمام محتواهای موجود در پایگاه نورمگز و نورلاب و یا کتاب‌ها و منابع موجود، مشابه‌یابی می‌کند و بخش‌های مشابه را کشف یا تقلب‌یابی می‌نماید. بدیهی است

و فرایندهایی از این دست، در این مرحله انجام می‌شود.

در مرحله بعد، فرآوری محتوا را داریم که برچسب‌گذاری محتوا، مثلاً برچسب: آیه، حدیث، ریشه و انواع و اقسام برچسب‌گذاری‌های موضوعی، روی آن محتوای دیجیتال‌شده انجام می‌شود تا متن، خوانایی لازم را برای ماشین پیدا کند. در مرحله سوم، محتوای فرآوری‌شده، تصحیح‌شده و برچسب‌خورده و غنی‌سازی‌شده، در قالب نرم‌افزارهای رومیزی یا وبی به انتشار می‌رسد.

در مورد اینکه ما در کدام‌یک از این مراحل توانستیم هوش مصنوعی را به طور مناسب به کار بگیریم، باید بگوییم که ما در هر سه مرحله، توانستیم از هوش مصنوعی به‌رمنده شویم و کارهای خودمان را توسعه داده و پیش ببریم.

در مرحله اول، به طور نمونه، با تکنیک‌هایی مثل OCR دیتاهای تصویری خودمان را تحلیل می‌کنیم و متن را از آنها استخراج می‌نماییم و محتوا را توسعه می‌دهیم. همچنین، با تکنیک‌هایی تصحیح لازم را در متون انجام می‌دهیم و به صورت ماشینی، برچسب‌گذاری‌ها و غنی‌سازی‌هایی را روی متن اعمال می‌کنیم. در مرحله سوم هم قابلیت‌های هوشمند و جدیدی را به کار گرفته‌ایم که از جمله می‌توانیم به انواع پیشنهاددهنده‌ها در حوزه عرضه محتوا اشاره نماییم.

اهداف به‌کارگیری هوش مصنوعی در محصولات نور

یکی از این اهداف، چابک‌سازی فرایندهای جاری تولید و فرآوری محتوای خوب است که قبل از ورود هوش مصنوعی به این عرصه، بسیاری از پردازش‌ها به صورت دستی انجام می‌شد که قاعدتاً زمان، هزینه و نیروی بسیاری صرف آن می‌شد؛ چنان‌که

برای تولید محصولات تخصصی

مثل علوم اسلامی، با دو گروه از

متخصصان مواجه هستیم. یکی،

متخصصان توسعه فنی و پردازش

هوشمند و دوم، متخصصان

یک دانش خاص، همچون:

قرآن، حدیث، فقه، تاریخ و...

به‌طور کلی، پروژه‌هایی موفق

هستند که در مراحل مختلف کار،

ارتباط خیلی نزدیک و تنگاتنگی

بین هر دو گروه از متخصصان

وجود داشته باشد

که این کار، اصلاً به شکل دستی و سنتی، قابلیت انجام ندارد.

دستاوردهای نور در هوشمندسازی علوم اسلامی

مرکز تحقیقات کامپیوتری علوم اسلامی، تاکنون دستاوردهای بسیاری در زمینه هوشمندسازی فرایندها و ارائه قابلیت‌های جدید در زمینه تحلیل هوشمند محتوا داشته است که در این مجال، به برخی از آنها اشاره می‌کنم.

- تجزیه صرفی و نحوی:

اولین دستاورد ما، ابزار تجزیه صرفی بود؛ بدین ترتیب که یک کلمه عربی را به سیستم می‌دهیم و سیستم آن را از حیث صرفی و ساختمان کلمه، تحلیل می‌کند که چه پیشوندها یا پسوندهایی می‌تواند داشته باشد. اصل یا هسته اصلی کلمه و یا پسوند و پیشوند آن، چه ویژگی‌هایی صرفی دارند؛ مثلاً ریشه این کلمه چیست، اعراب آن چیست، جمع است یا مفرد، اسم است یا فعل، چه وزنی دارد و امثال آن. در واقع، اقسام ویژگی‌هایی که در تحلیل صرفی کلمه می‌تواند برای ما مهم باشند، ماشین همه حالات ممکن در آن کلمه را تحلیل و بررسی می‌کند و حالتی را که به نظرش می‌تواند صحیح باشد، به عنوان خروجی ارائه می‌دهد. بدیهی است که چنین چیزی، کاربردهای خیلی متعددی برای خواهد داشت. دور از واقعیت نیست اگر بگوییم که هزاران قانون و قاعده در پشت صحنه این موتور پیاده‌سازی شده تا بتواند رفتار مناسب و صحیحی از خودش بروز دهد.

علاوه بر این، در لایه نحو هم وارد شده‌ایم و مثلاً در محتوای قرآنی، داده استاندارد آماده شده که در آن، ساختار یک جمله

که مبتنی بر قواعد نحوی زبان عربی است، در قالبی قابل فهم برای ماشین آماده گردید؛ مثلاً این کلمه، مبتدأست یا خبر، فاعل است یا مفعول و مانند آن. متن کامل قرآن کریم، در حال حاضر از صرفی و نحوی، برچسب‌گذاری شده و در پایگاه جامع قرآنی مرکز در اختیار مخاطبان قرار داده شده است.

همچنین، زیرساخت و رابط کاربری خاصی آماده شده که کاربران قادر باشند در این دیتای ارزشمند صرفی و نحوی قرآن به جست‌وجو بپردازند و به مطلب مورد نظرشان دست یابند.

- مدخل‌یابی کلمات:

استفاده دیگری که از این موتور صرفی هوشمند نموده‌ایم، مدخل‌یابی هوشمند کلمات است؛ یعنی ماشین، به‌ازای جست‌وجوی یک کلمه توسط کاربر، از میان انبوه مداخلی که وجود دارد، بر اساس تحلیل صرفی کلمه، نتایج را به صورت هوشمندانه توسعه داده و بهترین مدخل‌ها را به کاربر پیشنهاد می‌دهد و پاسخ‌های غیرمرتبط و دور از هدف کاربر را از چرخه پاسخ‌های جست‌وجو خارج می‌کند و نتایج مطلوب را ارائه می‌دهد.

- مشابه‌یابی متون:

دستاورد دیگر ما، به بحث مشابه‌یابی متون مربوط می‌شود؛ به طور مثال، در متون عربی مثل احادیث، این قابلیت وجود دارد که روایات مشابه حدیث انتخابی کاربر را با درصد تشابه مورد نظر کاربر نمایش دهیم. برای اولین بار در نرم‌افزار جامع الاحادیث سه‌ونیم از این قابلیت مفید رونمایی کردیم که برای کاربران، به‌خصوص حدیث‌پژوهان بسیار کاربردی است. اینکه کسی بخواهد برای یافتن روایات مشابه به

شکل سنتی عمل کند و در منابع حدیثی جست‌وجو نماید، وقت و هزینه بسیاری از او می‌گیرد؛ ولی شما اکنون با یک کلیک می‌توانید به فهرستی جامعی از روایات مشابه حدیث انتخابی خودتان دسترسی داشته باشید. گفتنی است که علاوه بر نرم‌افزار جامع الاحادیث سه‌ونیم که به احادیث شیعه اختصاص دارد، در کتابخانه احادیث فریقین هم این ابزار به کار گرفته شده است تا علاقه‌مندان بتوانند در منابع اهل سنت هم به مشابه‌یابی احادیث مبادرت ورزند.

The screenshot shows the 'Rahmat' website interface. At the top, there is a header with the site's logo and navigation tabs: 'مشابه‌یابی', 'تازه‌به‌سرقت علمی', 'قابلیت‌ها', 'راهنمای کاربری', 'اخبار', 'عضویت نشریات ویژه'. Below the header, the search results for 'مشابه‌یابی' are displayed. A table lists search results with columns for 'درصد کل مشابهت' (Overall Similarity) and 'بست اسناد مشابه' (Similar Documents). The first result is 'مقایسه تأثیر تدریس زبان انگلیسی با نرم افزار آ...' with a similarity of 71.30%. Other results include 'بررسی تأثیر آموزش به کمک شبکه های اطلاع...', 'مقایسه تأثیر آموزش به کمک نرم افزارهای آمو...', 'آموزش به شیوه حرف هجا کلمه با کمک چند ر...', 'بررسی رابطه میزان استفاده از فن آوری اطلاع...', and 'نرم افزار آموزشی ریاضی ابتدایی بر اساس رویک...'. Each result has a magnifying glass icon and a percentage. The right side of the screenshot shows a preview of the selected document, which is a PDF titled 'توسعه فناوری اطلاعات و ارتباطات در برنامه‌های آموزش و پرورش'. The preview text discusses the importance of information technology in education and training.



هدف گذاری ما باید واقع گرایانه باشد. در سال های اخیر، با توجه به موفقیت هایی که در هوش مصنوعی با آن روبه رو بودیم، یک فضای تبلیغاتی شکل گرفت که باعث شد انتظارات دیگران از متخصصان این حوزه، فراتر از حد متعارف بالا برود و ما از واقع گرایی فاصله بگیریم و در عمل، نتوانیم به نتیجه دلخواه برسیم و همین موجب شود در این حوزه، گاهی روح ناامیدی حاکم شود و به همین دلیل، با تصمیماتی هیجانی، خود و جامعه را از کاربردهای هوش مصنوعی محروم کنیم و حتی جلوی پیشرفت های بعدی را بگیریم



همچنین، از این قابلیت، در پایگاه جامع الأحادیث نسخه جدید هم استفاده شده است. قبلاً اگر کاربر جست و جویی در پایگاه حدیث نور انجام می داد، معمولاً به نتایج متعدد و تکراری منجر می شد؛ ولی در نسخه جدید پایگاه، احادیث را گروه بندی کرده ایم و پاسخ های جست و جو بسیار دقیق و نزدیک به خواسته کاربر است و تنها با یک تیک، جلوی تکرار نتایج زاید گرفته می شود.

همان طور که اشاره کردم، قابلیت مشابه یابی در سامانه سمیم نور هم به کار گرفته شده است. محقق می تواند تمام اجزای مقاله خود را در میان انبوهی از مقالات، کتابها و پایان نامه های حوزه علوم اسلامی و انسانی کاوش نماید و اصالت یک متن و میزان یا درصد

مشابهت آن را با سایر متون بررسی نماید.

علاوه بر این، در زمینه مشابه یابی معنایی و فرالفظی نیز کارهای ارزشمندی انجام شده است؛ برای نمونه، در قسمت احادیث مرتبط پایگاه حدیث، روایات مرتبط با حدیث مد نظر کاربر، با استفاده از تکنیک های تحلیل معنایی متون و مشابه یابی فرالفظی، پیاده سازی شده است.

در محتوای قرآنی نیز از تکنیک های مشابه یابی استفاده شده و کاربر می تواند به آیات مشابه لفظی یا معنایی آیه مورد نظر خودش دسترسی داشته باشد؛ برای مثال، وقتی کاربر آیه مربوط به نیکی به پدر و مادر را به سیستم بدهد، ماشین تمام آیات مرتبط با این موضوع را برایش فهرست می کند؛ حتی آیاتی که در آنها هیچ اشاره لفظی به این موضوع نشده باشد؛ اما مفهوم و محتوای این موضوع مورد توجه قرار گرفته است.

- برچسب گذاری آیات و روایات:

کار دیگر ما، برچسب گذاری آیات در متون است؛ به بیان دیگر، سیستم ما، محتوای خام را می گیرد و بعد مشخص می کند کجای این محتوا، آیه است و نشانی آن چیست. این سیستم، رسم الخط های مختلف قرآن را نیز پوشش می دهد. همین کار، برای متون حدیثی هم پیاده سازی شده است و به شکل خودکار، عملیات برچسب گذاری احادیث در متون را انجام می دهیم.

- برچسب گذاری پاورقی ها:

کار خوب دیگری که انجام شد، برچسب گذاری پاورقی های متون است؛ اینکه این پاورقی، به چه کتاب، جلد یا صفحه ای مربوط است و آن را به محتوای مربوطه لینک بدهد.

- برچسب گذاری اعلام اشخاص:

در زمینه استخراج خودکار اعلام اشخاص و به اصطلاح

```
<Parag ParagNumber="12">
  وفي السنة السادسة أخرجته أمه إلى أخواله زائرة
  فتوفيت بين مكة و المدينة، و نمي ذلك إلى أم أيمن
  <Event Name="Die Event">on event
    <Trg>
      توفيت Trigger
    </Trg>
    <Arg>
      أمه Victim:
      - Agent:
      بين مكة و المدينة Place:
      السنة السادسة Time:
    </Arg>
  </Event>
</Parag>
```

بیتار مجلات تخصصی نور
 مقالات | بیداری اسلامی

۱۰. سیاست خارجی جمهوری اسلامی ایران در قبال بیداری اسلامی
 نویسنده: جعفری، سید اصغر؛ قاسمی، علیرضا؛
 مجله: مطالعات سیاسی جهان اسلام « تابستان ۱۳۹۳ - شماره ۱۰ علمی-پژوهشی/ISC (۲۲ صفحه - از ۵۳ تا ۷۴)
 ... سیاست خارجی جمهوری اسلامی ایران در قبال بیداری اسلامی سیداصفرجعفری استادیار حقوق بین الملل دانشگاه ... جنبش های
 اجتماعی خاورمیانه و شمال افریقا که از آنها به عنوان بیداری اسلامی یاد می شود از مهم ترین و بحث ...

کلیدواژه ها | چکیده | مقالات مرتبط | پیشنهاد دیگران

عبارات پیشنهادی برای جستجو

+ بهار عربی	+ مصر	+ بحرین
+ انقلاب مصر	+ سوریه	+ تونس
+ اخوان المسلمین	+ خاورمیانه	+ بهار عربی
+ بیداری	+ انقلاب اسلامی	+ بیداری اسلامی مصر
+ بیداری اسلامی در مصر	+ تحولات خاورمیانه	+ یمن

۶ ۵ ۴ ۳ ۲ ۱

موجودیت‌های نامدار هم کار خوبی انجام شده است. ان‌شاءالله در آینده بتوانیم محصولات مبتنی بر این استخراج‌های ماشینی داشته باشیم.

– استخراج هوشمند رویدادهای تاریخی:

در این زمینه، متون خام تاریخی را به سیستم می‌دهیم و آنگاه سیستم، رویدادهای تاریخی موجود در متن را شناسایی می‌نماید و تمامی اطلاعاتی را که به آن مربوط هست، گزارش می‌کند. امیدواریم بتوانیم از این قابلیت، در پایگاه تاریخ رونمایی کنیم.

– ردبندی متون:

در برخی متون تخصصی مثل فقه، ما از قابلیت ردبندی متون استفاده کرده‌ایم؛ برای مثال، اگر کاربر یک متن فقهی یا حدیثی را به سیستم ارائه دهد، سیستم تعیین می‌کند که برچسب موضوعی آن متن چیست. بدیهی است که تعیین موضوع یک متن، کارایی بسیاری برای کاربر و محقق دارد و از این طریق به‌آسانی می‌تواند پژوهش‌های خود را ساماندهی کند.

– با هم‌آبی متون:

فیلد دیگری که به برکت هوش مصنوعی، آن را آماده کردیم، قابلیت‌هایی مبتنی بر تحلیل باهم‌آبی متون است؛ برای مثال، از این تکنیک در سیستم استخراج آیات مرتبط استفاده کرده‌ایم؛

یعنی علاوه بر ارتباط لفظی و معنایی و موضوعی، می‌توانیم گونه دیگری از آیات مرتبط با آیه کاربر را نیز برای او فهرست کنیم. این قابلیت، مبتنی بر این نظریه است که آیاتی که مفسران متعدد در کنار هم ذکر کرده‌اند، با یکدیگر ارتباط خواهد داشت. این تکنیک، در انبوه تفاسیر قرآنی اجرا شد و بدین ترتیب، کاربر می‌تواند بفهمد آیات مرتبط با آیه انتخابی او چه آیاتی هستند و دیدگاه مفسران را هم ملاحظه کند.

مراحل تولید نرم‌افزار در مرکز نور، در سه مرحله خلاصه می‌شود:

مرحله اول، تولید محتواست. تایپ و تبدیل محتوا و تصحیح متون و فرایندهایی از این دست، در این مرحله انجام می‌شود.

در مرحله بعد، فرآوری محتوا را داریم که برچسب‌گذاری محتوا، مثلاً برچسب: آیه، حدیث، ریشه و انواع و اقسام برچسب‌گذاری‌های موضوعی، روی آن محتوای دیجیتال شده انجام می‌شود تا متن، خوانایی لازم را برای ماشین پیدا کند. در مرحله سوم، محتوای فرآوری شده، تصحیح شده و برچسب‌خورده و غنی‌سازی شده، در قالب نرم‌افزارهای رومیزی یا وبی به انتشار می‌رسد

از این قابلیت، در پایگاه نورمگز استفاده شده است. بدین ترتیب، صدها هزار صفحه از مقالاتی را که به دلیل نبود متن، امکان جستجو در آنها وجود نداشته و صرفاً تصویرشان در دسترس ما بوده است، به متن تبدیل کرده‌ایم. بدیهی است که این نوع پردازش، دستاورد خیلی پُررنگ و چشمگیری به شمار می‌رود.

– پردازش صوت:

در جستجوی صوتی هم گام‌های خوبی برداشته‌ایم. البته باید بگوییم به جهت پُرهزینه بودن فیلد پردازش صوت و منابع کمی که داشتیم، بیشتر از منابع متن‌باز که در این حوزه موجود بودند، استفاده کردیم و با سفارشی‌سازی آنها توانستیم به یک سامانه‌ای مطلوب برسیم که به وسیله آن، کاربر می‌تواند در محتوای صوتی مورد نظر خودش جستجو کند. از این قابلیت، برای اولین بار به صورت آزمایشی در نرم‌افزار «کتابخانه مقتل الحسین(ع)» استفاده شد. این امکان نیز خیلی کاربردی و مفید به نظر می‌رسد.

افزون بر به‌کارگیری این تکنیک در تحلیل آیات و آرائه آیات مرتبط در پایگاه جامع قرآنی، در خود موتور جستجوی نور هم آن را به کار گرفته‌ایم؛ مثلاً در پایگاه مجلات تخصصی نور (نورمگز)، قابلیت پیشنهاد عبارت جستجو راه‌اندازی شد؛ به طوری که مثلاً وقتی کاربر عبارت «بیداری اسلامی» را جستجو می‌کند، سیستم علاوه بر عرضه پاسخ‌های مربوط به این واژه، عباراتی همچون «بهار عربی» و «اخوان المسلمین» را هم به کاربر پیشنهاد می‌دهد که کاملاً یک پیشنهاد هوشمندانه است؛ یعنی سیستم به شکل خودکار، عباراتی را که با متن مورد نظر کاربر، هماهنگی و ارتباط لفظی یا معنایی دارد، ارائه می‌کند.

– ترجمه ماشینی متون:

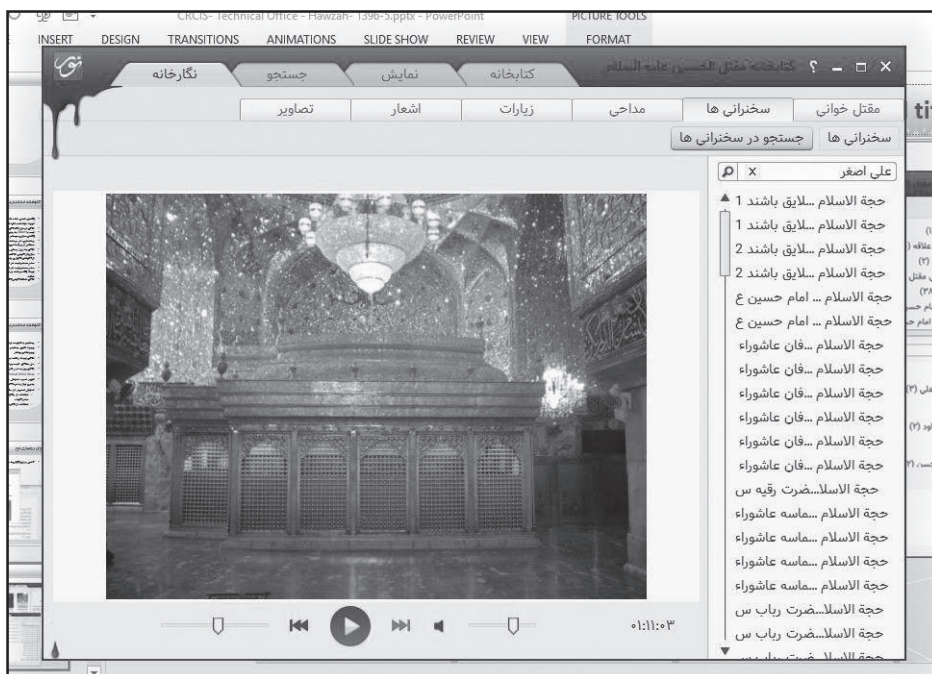
فیلد دیگری که خیلی جذاب است و در سال‌های اخیر به آن ورود پیدا کرده‌ایم، بحث ترجمه ماشینی است. از این امکان، در دو جای اصلی استفاده نموده‌ایم. یکی از آنها، در بحث ترازبندی خودکار متن ترجمه است. خوب می‌دانید که خیلی از جاها متون منبع و مرجع در علوم اسلامی، دارای ترجمه‌های متعددی هستند. تاکنون ارتباط بین متن و ترجمه آن، توسط انسان انجام می‌شد؛ اما حالا توسط تکنیک ترجمه ماشینی، صفحه یا پاراگراف متون را به متن ترجمه آن متصل کرده‌ایم و خود ماشین، به صورت هوشمند این فرایند را برای کاربر انجام می‌دهد.

علاوه بر این، از تکنیک ترجمه ماشینی در ترجمه‌یابی احادیث هم استفاده نموده‌ایم. خیلی از احادیث هستند که ممکن است تاکنون ترجمه مشخصی برای آنها توسط انسان تحریر نشده باشد؛ اما سیستم، از طریق بازیابی اطلاعات ترجمه ماشینی آن، ترجمه‌یابی کرده و پس از یافتن ترجمه مناسب، آن را پیشنهاد می‌دهد. ان‌شاءالله، به‌زودی این قابلیت را در پایگاه جامع‌الاحادیث به کار خواهیم گرفت.

– پردازش تصویر:

در گروه پردازش هوشمند نور، علاوه بر پردازش متن، به مقوله پردازش تصویر هم ورود پیدا کرده‌ایم. در زمینه OCR و استخراج متن از تصاویر، خوشبختانه توانستیم محصولی را تولید کنیم و از آن در فرایندهای ورود اطلاعات و دیجیتالی کردن متون استفاده کنیم. یکی از پُرهزینه‌ترین مراحل تولید نرم‌افزارهای نور، بحث دیجیتالی کردن اطلاعات است. بدیهی است که با داشتن یک چنین سیستم‌هایی، صرفاً با داشتن تصویر آن کتاب، می‌توانیم به‌راحتی از مرحله دیجیتالی کردن اطلاعات گذر نماییم و دیگر نیازی به تایپ تمامی متون نخواهیم داشت و با حداقل هزینه، این فرایند به انجام خواهد رسید.

با قابلیت‌های برآمده از هوش مصنوعی، نیازهای پیچیده‌ای را می‌توانیم برطرف کنیم که تا قبل از آن، اصلاً امکان آنها وجود نداشت؛ مثلاً در پروژه سمیم، فرایند مشابه‌یابی مقالات در سطح خیلی حجیم و وسیعی انجام می‌شود؛ یعنی این سامانه، یک مقاله را با تمام محتواهای موجود در پایگاه نورمگز و نورلایب و یا کتاب‌ها و منابع موجود، مشابه‌یابی می‌کند و بخش‌های مشابه را کشف یا تقلب‌یابی می‌نماید



پرسش و پاسخ

در پایان، کاربران حاضر در وبینار، به ارائه سؤالات خود پرداختند. در ادامه، برخی از این پرسش و پاسخ‌ها را جهت استفاده خوانندگان گرامی می‌آوریم.

۱. آیا مرکز برنامه‌ای برای ترجمه ماشینی متون اسلامی به زبان‌های غیرفارسی دارد؟

فعالاً در چشم‌انداز کوتاه‌مدت، برنامه‌ای در این راستا در نظر گرفته نشده است؛ اما با استفاده از ابزارهایی که در این حوزه وجود دارد، این قابلیت در فضای متون اسلامی قابل دستیابی است.

۲. برای شناسایی و تعیین مرجع ضمائر موجود در قرآن یا احادیث، چه برنامه‌ای دارید؟

در زمینه تعیین مرجع ضمیر، کارهای خوبی در مرکز به صورت دستی شده است. در سال‌های گذشته، یک کار دستی در این زمینه انجام شده است؛ یعنی در نرم‌افزار قرآنی جامع تفاسیر. البته نمونه و مشابه این قابلیت، در نرم‌افزار نور السیره هم به صورت دستی انجام شده است که با استفاده از این دادگان و با پیشرفت‌هایی که در زمینه شناسایی خودکار مرجع ضمیر در سال‌های اخیر وجود داشته، در صورت لزوم، می‌توان ابزارهای مناسبی در این زمینه تولید کرد و به فراخور حال، در محصولات نرم‌افزاری از آنها استفاده نمود.

۳. آیا در نرم‌افزارهای نور به امکان ترجمه هوشمند هم توجه شده است؟

فعالاً در نرم‌افزارهای نور، این قابلیت عرضه نشده است؛ اما در خصوص ترجمه ماشینی متون، ما به نتایجی رسیده‌ایم که ان‌شاءالله بتوانیم در نرم‌افزار جامع‌الاحادیث، از این ابزار یک خروجی بگیریم. اما اینکه مستقیماً ابزار ترجمه هوشمند را برای کاربران و علاقه‌مندان ارائه کنیم، باید بگوییم که احتمالاً در ماه‌های آتی، این قابلیت را در وبگاه متن‌کاوی مرکز نور قرار بدهیم.

۴. آیا امکان استفاده از برنامه ترازبندی برای کاربران وجود دارد؟

اگر منظور شما ترازبندی متن و ترجمه یا متن و صوت است،

فعالاً هیچ‌کدام از آنها رونمایی نشده و بیشتر در قالب یک ابزار، به عنوان دستیار پژوهشگران داخلی مرکز، از آن استفاده می‌شود. اگر بدانیم که استفاده‌کننده از این ابزار زیاد است، آن را در وبگاه متن‌کاوی قرار خواهیم داد.

۵. آیا در تولیدات نور، برای معرفی مقاله مرتبط با متن، امکانی وجود دارد یا خیر؟

در پایگاه نورمگز، ذیل هر مقاله، مقالات مرتبطی پیشنهاد می‌دهیم. اگر مقاله مد نظر شما جزء مقالات پایگاه نور مگز باشد، از آن مقالات مرتبطی که توسط ماشین پیشنهاد داده می‌شود، می‌توانید استفاده کنید.

البته باید بگوییم که در پایگاه نوریباب نیز می‌توانید متن یک جمله یا عبارت مدنظر خودتان را مشابه‌یابی کرده و به کتاب‌های حاوی عبارات مشابه دسترسی داشته باشید. ■