

چارچوبی برای تشخیص خودکار آیات قرآن در متون فارسی و عربی

محسن شاه‌محمدی^۱، تکتام علیزاده لوشابی^۲ و بهروز مینایی بیدگلی^۳

^۱ دانشگاه آزاد اسلامی تهران-شمال، تهران، ایران؛ مرکز تحقیقات کامپیوتری علوم اسلامی نور، قم، ایران

^۲ دانشگاه پیام نور قم، ایران؛ مرکز تحقیقات کامپیوتری علوم اسلامی نور، قم، ایران

^۳ دانشگاه علم و صنعت ایران، تهران، ایران؛ مرکز تحقیقات کامپیوتری علوم اسلامی نور، قم، ایران

^۱mshahmohammadi@noornet.net ^۲T_alizadeh20@yahoo.com ^۳b_minaei@iust.ac.ir

چکیده

در این مقاله چگونگی طراحی و پیاده‌سازی موتور هوشمند تشخیص آیات قرآن مطرح شده که می‌تواند به طور خودکار آیات قرآنی موجود در متون دیجیتال مختلف را تشخیص داده و برجسته کند. حیطه پردازشی این سیستم جزو سیستم‌های متن‌کاوی قرار دارد و عملیات انجام شده در آن به دلایلی که در مقاله آرایه شده، فراتر از حد سیستم‌های تطابق دقیق چند الگویی معمول است؛ چرا که در آن با مسایلی از قبیل تشخیص توالی الگوها و تشخیص چندین الگوی متفاوت به عنوان یک الگو روبرو هستیم. در پیاده‌سازی این سیستم، الگوریتمی نوین مبتنی بر اندیس‌گذاری متن و الگوها طراحی شده که ایده اصلی آن نگاشت الگوها و متن مورد پردازش به یک سری آرایه‌های عددی و پردازش بر روی این اعداد به جای پردازش متن است. نتایج آزمایش‌ها نشان می‌دهد که این ایده تأثیر بسزایی در کارایی (دقت و سرعت) الگوریتم داشته است.

کلمات کلیدی

بازیابی اطلاعات، متن‌کاوی، تشخیص هوشمند آیات قرآن، تطبیق الگو، تطابق دقیق چند الگویی.

پنجم). بنابراین هرگونه تغییر در آن با هدف مؤلف ناهمخوان تلقی شده و باید چاره‌ای دیگر اندیشید.

• سرعت، دقت و جامعیت قابل قبول برای جستجو

سیستمی که طراحی می‌شود باید سرعت قابل قبولی داشته باشد تا بتواند جایگزین انسان شود. همچنین، معیارهای دقت و صحت تشخیص باید سطح اطمینان بالایی در اعلام کردن یا اعلام نکردن بخشی از متن به عنوان آیه قرآنی برآورده سازند. این معیارها در موتور هوشمند تشخیص آیات قرآنی کاملاً در نظر گرفته و ارزیابی شده‌اند. جامعیت این سیستم نیز به گونه‌ای است که هیچ گونه فرضیه محدودکننده‌ای در آن وجود ندارد. برای نمونه، متن مورد جستجو می‌تواند شامل عبارات قرآنی باشد یا نباشد، نشان‌گذاری‌های مختلف ویرایشی داشته باشد یا نداشته نباشد و یا دارای یک یا چندین رسم‌الخط مختلف باشد. حجم متن مورد جستجو می‌تواند کوچک و یا بسیار بزرگ (یک و یا چند جلد از یک کتاب) باشد.

ادامه مقاله در پنج بخش تدوین شده است. در بخش ۲ مختصری از کارها و الگوریتم‌های مشابه این موضوع سخن رفته است. بخش ۳ درباره سیستم تطابق دقیق چند الگویی، مشخصاً برای قرآن است. در بخش ۴ الگوریتم پیشنهادیمان آرایه شده؛ ارزیابی کارایی الگوریتم پیشنهادی و نتایج آزمایش‌ها در بخش ۵ و نتیجه‌گیری در بخش ۶ آمده است.

۱- مقدمه

برجسته کردن آیات قرآنی در متون مکتوب نیازی است که مدت‌هاست مورد توجه پژوهشگران، مؤلفان و نسخه‌برداران قرار گرفته است. شاید در نگاه اول، جستجو و نشان‌گذاری متون قرآنی، آن هم با استفاده از ماشین، موضوعی ساده به نظر آید؛ اما فقط بعد از شناخت تفاوت‌ها، دشواری‌ها ظاهر می‌شود. برای نمونه، ویرایش هر یک از موارد متفاوت ذیل و یا صرف نظر کردن از هر کدامشان، می‌تواند در نتایج آماری برگرفته از نشان‌گذاری‌ها به شدت تأثیرگذار باشد.

• متنی که با آیه‌ای از قرآن شباهت فراوان دارد، ولی از برخی جهات مانند اعراب و ضمیر، تفاوت‌هایی با آن آیه دارد. این متن ممکن است یک آیه از قرآن نباشد!

• وجود رسم‌الخط‌های مختلف از آیات قرآن در یک متن برای کتابت قرآن، رسم‌الخط‌های مختلفی وجود دارد. ممکن است در کتابی از چند رسم‌الخط (املا) مختلف برای نگارش قرآن استفاده شود و یا مؤلفی برای عبارات قرآنی در کتاب خود، از رسم‌الخط خاص خود استفاده کند که با رسم‌الخط‌های مرسوم متفاوت باشد. یک راحل این مشکل این است که تمام رسم‌الخط‌های کتاب را به رسم‌الخطی واحد تبدیل کرده و جستجو را بر روی آن رسم‌الخط انجام داد. اما این مسئله زمانی مشکل‌ساز می‌شود که هدف از تألیف کتاب و یا تدوین سرفصل خاصی از آن، بررسی اختلاف کتابت‌ها باشد (مثلاً کتاب «المقنع فی رسم مصاحف الامصار» از مؤلفی در قرن