

چارچوبی برای تشخیص خودکار آیات قرآن در متون فارسی و عربی

محسن شاه‌محمدی^۱، تکتم 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@noor.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؛ چراکه 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.

۱ - مقدمه

• **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 هیچ‌گونه 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.

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!

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.

• **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 راه‌حل 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



دسته الگوریتم‌های مبتنی بر جدول انتقال^۱ و الگوریتم‌های مبتنی بر ماشین خودکار^۲ تقسیم می‌شوند [۱۰].

۳- تطابق چند الگویی در مورد آیات قرآن کریم

در موتور هوشمند تشخیص آیات قرآنی، الگوها کلمات قرآن هستند که به صورت تک کلمه‌ای در بانکی ذخیره شده‌اند. گرچه هسته اصلی الگوریتم این سیستم را الگوریتم‌های تطابق دقیق چند الگویی تشکیل می‌دهد، اما از جنبه‌های بسیاری این موتور با دیگر موتورهای جستجوی متنی که تطابق چند الگویی را انجام می‌دهند، تفاوت دارد. از جمله این تفاوت‌ها عبارتند از:

- **جستجوی الگوهای متوالی**
یکی از مسائلی منحصراً به فرد در این سیستم که الگوریتم آن را پیچیده‌تر از سایر روش‌های جستجوی تطابق چند الگویی می‌سازد، این است که علاوه بر یافتن الگوها، توالی الگوهای شناسایی شده نیز یکی از شروط جستجوی الگو است. بنابراین، مسئله از جستجوی الگو به جستجوی الگوهای متوالی تغییر می‌یابد که این رویه در اکثر الگوریتم‌های تطابق چند الگویی اهمیتی ندارد.

- **شکل‌های متفاوت برای یک الگو**
یک کلمه در قرآن ممکن است رسم‌الخط‌های مختلفی داشته باشد، یعنی ممکن است یک الگو به شکل‌های مختلفی ظاهر شود که تمامی آن‌ها باید به عنوان یک الگو تشخیص داده شوند.

- **حجم بسیار بزرگ مجموعه داده‌های الگو در مقایسه با دیگر الگوریتم‌ها**

قرآن کریم بیش از ۸۴۰۰۰ کلمه دارد که هر کدام خود یک الگو محسوب می‌شوند. علاوه بر آن، در این موتور جستجو از تمامی رسم‌الخط‌های قرآنی پشتیبانی می‌شود و این مسئله حجم مجموعه الگوها را حدوداً ۳ برابر افزایش می‌دهد. بدیهی است که الگوریتم به کار رفته برای تطابق چند الگویی در این سیستم، بسیار متفاوت از الگوریتم‌های به کار رفته در مجموعه‌های کوچک خواهد بود. این در حالی است که در بین الگوریتم‌های تطابق چند الگویی، مجموعه داده‌های بالاتر از ۱۰۰۰۰ الگو، به عنوان مجموعه داده‌های بسیار بزرگ شناسایی می‌شوند [۴] و معمولاً مجموعه داده‌هایی که به عنوان مجموعه داده‌های بسیار حجیم محسوب می‌شوند، تا حدود ۱۰۰۰۰۰ الگو را پشتیبانی می‌کنند؛ حال آن که مجموعه داده‌های (الگوها) این موتور جستجو چندین برابر این رقم است.

- **سیستم تطابق دقیق چند الگویی مشخصاً برای قرآن**
گرچه موتورهای جستجوی مختلفی برای قرآن پیاده‌سازی شده است که یک کلمه، آیه و یا بخشی از یک آیه را عیناً در قرآن جستجو و آدرس‌دهی می‌کند، اما هیچ کدام عملکرد سیستم مذکور را ندارد تا قادر باشد به طور خودکار آیات قرآنی موجود در متن را تشخیص دهد. مثلاً در [۱۱] یک موتور جستجوی قرآنی برای صفحات وب پیاده‌سازی شده است. این سیستم همانند جستجوهای معمول، یک آیه دریافت کرده و گزارشی مبنی بر وجود و یا عدم وجود آیه در آن صفحه وب ارائه می‌دهد. در سیستم پیاده‌سازی شده در [۱۲] نیز

الگوریتم اصلی موتور هوشمند تشخیص آیات قرآنی در حیطه الگوریتم‌هایی مانند تطابق چند الگویی^۱ قرار دارد که در آن‌ها مسئله اصلی «پیدا کردن یک مجموعه الگوی خاص در یک متن به طور همزمان» است [۱]. الگوریتم‌های تطابق رشته کاربردهای فراوانی در مباحث بازیابی اطلاعات^۲ دارد [۱،۲]. تاکنون الگوریتم‌های متعددی برای این مسئله به کار رفته است، اما تمامی این الگوریتم‌ها تفاوت‌هایی با الگوریتم و شرایط موجود در سیستم تشخیص هوشمند آیات قرآنی دارند که در ادامه آمده است. به علاوه تاکنون چنین سیستمی، مشخصاً برای متن قرآن کریم در جهان پیاده‌سازی نشده است. الگوریتم‌های تطابق چند الگویی را می‌توان به دو گروه کلی تقسیم‌بندی کرد [3,4,5]:

- **الگوریتم‌های مبتنی بر درخت^۲**
شامل الگوریتم‌هایی مانند Aho-Corasick [۶]، Commentz-Walter [۲]، SBOM [۷،۸] و ... است. رامحل اصلی این الگوریتم‌ها، ساخت درختی از الگوها و جستجوی متن با کمک این درخت است. مشکل اصلی این الگوریتم‌ها این است که به همان سرعت که مجموعه الگوها بزرگ می‌شود، درخت نیز به سرعت رشد می‌کند و نیاز به فضای بیشتری دارد؛ به همین دلیل برای مجموعه الگوهای بزرگ مانند قرآن که ۸۴۰۰۰ کلمه دارد، این روش چندان مؤثر نیست.

- **الگوریتم‌های مبتنی بر فیلتر کردن**
ایده اصلی این الگوریتم‌ها این است که یک موقعیت در متن در نظر گرفته می‌شود و توسط یک فیلتر مشخص می‌شود که آیا در این مکان یک تطابق با الگو وجود دارد یا نه. در [۴] از مقادیر hash به عنوان فیلتر پیدا کردن مکان تطابق استفاده شده است و در الگوریتمی در [۳] از یک فیلتر به نام feed-forward Bloom استفاده شده است. یک فیلتر خوب سریع است و جواب‌های غلط کمی تولید می‌کند. به علاوه یک بازبین مورد نیاز است تا جواب‌های درست را از جواب‌های غلطی که به اشتباه درست تشخیص داده شده‌اند، تشخیص دهد.

در نگاهی جزئی‌تر و در یک دسته‌بندی دقیق‌تر، الگوریتم‌های تطابق رشته به دو دسته «تطابق دقیق رشته» و «تطابق نسبی رشته» تقسیم می‌شوند.

- الگوریتم‌های تطابق دقیق رشته، سعی در پیدا کردن تمام مکان‌های تکرار یک الگو در یک متن ورودی دارند. برخی از الگوریتم‌های تطابق دقیق رشته عبارتند از الگوریتم‌های Naïve، Boyer-Moore [۱]، Brute force، Horspool، KMP [۹]، Rabin-Karp [۹]، Zhu-Takaoka و جستجوی سریع [۱].

بنابراین الگوریتم مورد نظر ما، در حیطه الگوریتم‌های تطابق دقیق چند الگو^۲ قرار می‌گیرد. این الگوریتم‌ها به طور کلی به دو



۴-۲- پیش‌پردازش

به منظور پردازش اولیه بر روی متن ورودی، کاراکترهای اضافه از قبیل علایم ویرایشی، اعراب و ... از متن ورودی حذف می‌شوند و برای هر یک از کلمات باقی مانده، اندیسی در نظر گرفته می‌شود تا جایگاهشان در متن مشخص باشد.

۴-۳- شناسایی کلمات قرآنی

در این مرحله، تک تک کلمات به ترتیب چینش در متن با الگوهای موجود در بانک الگو مقایسه می‌شود و اندیس الگوهایی را که با کلمه مورد نظر تطابق دارند، در آرایه‌ای که به آن کلمه اختصاص داده شده، ذخیره می‌کنیم (ممکن است بیش از یک الگو با کلمه مورد نظر تطابق داشته باشد؛ چرا که مثلاً کلمه «الله» در نقاط مختلفی از قرآن آمده، اما اندیس آن در هر مکان، متفاوت از مکان دیگر است). پس هر کلمه از متن به یک آرایه از اعداد نگاشته می‌شود. نکته‌ی مهم در این مرحله این است که ممکن است کلمه ورودی با هیچ الگویی مطابق نباشد، ولی با رسم‌الخط خاصی از یک الگو مطابق باشد؛ به همین دلیل در پیاده‌سازی این موتور جستجو فیلتری تعبیه شده که تمامی کلماتی که رسم‌الخط‌های متفاوتی دارند نیز در این مرحله یافت شده و اندیس‌گذاری می‌شوند.

این روند همچنان ادامه می‌یابد تا زمانی که برای کلمه ورودی، تطبیقی در بانک الگو پیدا نشود؛ در این صورت به تعداد کلمات قبلی (مقایسه شده)، آرایه‌هایی داریم که هر کدام مشخص‌کننده مکان‌های وجود آن کلمه در قرآن است. مثلاً اگر کلمات ابتدایی متن ورودی شامل «الله الذی هو الحیّ القیوم»، لا تأخذه سنة و لا نوم ...» باشد، کلمات به ترتیب پردازش شده، تا جایی که به کلمه‌ای غیر قرآنی برسیم؛ در اینجا فرض شده بعد از کلمه «نوم»، کلمه‌ای غیر قرآنی آمده است، بنابراین از ابتدای متن تا آخر کلمه «نوم» پردازش شده و آرایه‌هایی برای آن تشکیل می‌شود. به شکل ۱ توجه کنید.

۴-۴- جستجوی توالی، تشخیص و برجسته‌سازی آیه

بعد از پیمودن مراحل قبلی اکنون متنی در اختیار داریم که تمامی کلمات آن جزو قرآن هستند، اما ممکن است الزاماً یک آیه از قرآن نباشد؛ برای تشخیص این امر با استفاده از جستجو بر روی اندیس‌ها، به دنبال یک توالی اندیسی در بین این کلمات می‌گردیم؛ چرا که در مرحله اول بانک قرآنی را به گونه‌ای تشکیل دادیم که تمامی کلمات یک آیه، دارای اندیس‌های پشت سر هم هستند. بنابراین اگر حتی یک توالی اندیسی در بین این کلمات

جستجوی آیات قرآن بدون وابستگی به حروف صدادار و تفکیک‌کننده‌ها انجام می‌شود. در هر صورت، در هر دوی این سیستم‌ها و دیگر سیستم‌های جستجوی عبارات قرآنی موجود (با اندکی اختلاف در قابلیت‌ها)، یک آیه و یا بخشی از آن به عنوان ورودی به سیستم داده شده و جستجو انجام می‌شود؛ در حالی که در سیستم مورد نظر ما هیچ آیه‌ای به عنوان ورودی به سیستم داده نمی‌شود، بلکه تمامی آیات موجود در متن به طور خودکار تشخیص داده می‌شوند.

با توجه به مطالب ذکر شده و با در نظر گرفتن مزایا و معایب روش‌های موجود، الگوریتمی نوین مبتنی بر اندیس‌گذاری متن و الگوها طراحی کرده‌ایم که در دو مرحله آیات قرآنی موجود در متن را تشخیص می‌دهد. در اولین مرحله با استفاده از الگوریتم تطابق دقیق چند الگویی اندیسی، کلمات قرآنی موجود در متن تشخیص داده شده و به آرایه‌های اندیسی تبدیل می‌شوند و در مرحله دوم فیلتری طراحی شده که با استفاده از جستجو بر روی اندیس‌ها، قادر است توالی اندیسی در بین کلمات را تشخیص داده و آیات قرآن را برجسته سازد.

۴- الگوریتم پیشنهادی

الگوریتم سیستم تشخیص هوشمند آیات قرآنی موجود در متن شامل چهار مرحله اصلی به شرح زیر است:

۱. ساخت بانک‌های قرآنی
۲. پیش‌پردازش
۳. شناسایی کلمات قرآنی
۴. جستجوی توالی الگوها، تشخیص و برجسته‌سازی آیات قرآنی درون متن اصلی.

۴-۱- ساخت بانک‌های قرآنی

ساختن بانک‌های قرآنی یکی از مهم‌ترین وظایف در چارچوب پیشنهادی ماست. برای ساختن بانک اصلی قرآنی، متن کامل قرآن را به صورت کلمه به کلمه و پشت سرهم در آرایه‌ای ذخیره می‌کنیم تا به هر کلمه از قرآن به صورت منحصر به فرد، اندیسی تعلق یابد. هدف از این اندیس‌گذاری یکتا با علم به تکراری بودن بسیاری از کلمات قرآن، در ادامه الگوریتم مشخص می‌شود.

برخی دیگر از بانک‌هایی که تا کنون در ارتباط با متن‌کاوی متون قرآنی در سیستم پیشنهادی ایجاد شده است، عبارتند از: بانک رسم‌الخط‌های متفاوت، بانک عبارات تک کلمه‌ای و دو کلمه-ای و بانک عبارات سه کلمه‌ای. توضیح در مورد این بانک‌ها در این مجال نمی‌گنجد.

دلیل این دو نوع دسته‌بندی این است که هدف جستجوی آیات در این دو نوع منبع متفاوت است. مثلاً در متونی که فراوانی آیات در آنها کم است، مسئله اساسی یافتن آیات نیست، بلکه مشخص کردن بخش‌های زیادی از متن است که فاقد آیات قرآنی هستند، چرا که در این صورت از جستجوی اضافی صفحاتی که در آنها آیه قرآنی وجود ندارد، پرهیز می‌شود.

۵-۲- معیارهای ارزیابی

الگوریتم سیستم تشخیص آیات قرآنی جزو سیستم‌های بازیابی اطلاعات است، از این رو بر خلاف سیستم‌های بازیابی داده که در آن کارایی سیستم از نظر سرعت و فضا معیار ارزیابی در نظر گرفته می‌شود، در این سیستم نیز همانند سیستم‌های بازیابی اطلاعات، معیار صحت^۱، دقت^۲ و معیار^۳ F معیارهای اصلی ارزیابی هستند. در الگوریتم ما، «صحت» تشخیص یک آیه، نسبت تعداد کلماتی است که به درستی به عنوان قرآن تشخیص داده شده‌اند به تعداد کل کلمات قرآنی موجود در متن. «دقت» تشخیص یک آیه، تعداد کلماتی است که به درستی به عنوان قرآن تشخیص داده شده‌اند نسبت به تعداد کل کلماتی که قرآن تشخیص داده شده‌اند. معیار F میانگین هارمونیک صحت و دقت است و تأثیر این دو معیار را به طور هم‌زمان نشان می‌دهد.

۵-۳- نتایج آزمایش‌ها

برای ارزیابی سیستم از دو نوع منبع یکی با فراوانی آیات قرآنی بالا و دیگری با فراوانی کم آیات قرآنی استفاده شده و نتایج آزمایش‌ها برای هر کدام به تفکیک آرایه شده است. هنگامی که موتور هوشمند تشخیص آیات قرآنی، متنی را بررسی می‌کند، برای هر کلمه دو حالت به وجود می‌آید: کلمه، یا قرآن تشخیص داده می‌شود و یا غیر قرآن. بر همین اساس می‌توان دو معیار برای ارزیابی سیستم تعریف کرد:

یافت شود، مشخص می‌شود این کلمات یک آیه و یا بخشی از یک آیه (یا چند آیه در چند جای مختلف از قرآن) هستند. در این مثال، با پردازش آرایه‌ها، ۲ رشته توالی اندیسی مشخص می‌شود (به شکل ۲ توجه کنید):

- «هو الحی القيوم»: بخشی از آیه ۳ سوره آل عمران.
- «هو الحی القيوم لا تأخذه سنة و لا نوم»: بخشی از آیه ۲۵۵ سوره بقره.

در این مرحله، بلندترین رشته تطبیق یافت شده به عنوان آیه تشخیص داده شده و برگردانده می‌شود.

در آخرین بخش این مرحله نیز با استفاده از اندیس الگوها (که در مرحله پیش پردازش انجام شد)، آیات قرآنی بازیابی شده و با استفاده از اندیس کلمات در رشته ورودی، مکان حضور آیات قرآنی در متن نیز مشخص شده و به گونه‌ای برجسته می‌شود (به طور مثال با تغییر رنگ کلمات قرآنی موجود در متن). مزیت این روش این است که چون آرایه‌های اندیسی مربوط به کلمات را مرتب می‌کنیم، به راحتی می‌توانیم از الگوریتم‌های جستجوی بهینه مانند جستجوی دودویی، استفاده کنیم.

۵-۴- پیاده‌سازی و نتایج

برای بررسی کارایی الگوریتم پیشنهادی آزمایش‌های متفاوتی انجام داده‌ایم. در ادامه ابتدا داده‌های مورد آزمایش توصیف شده، سپس معیارهای ارزیابی معرفی شده‌اند و در انتها نتایج به دست آمده آرایه و تحلیل شده است. نتایج آرایه شده از میانگین نتایج آزمایش‌های مختلف به دست آمده است.

۵-۱- مجموعه داده‌ها

برای بررسی نتایج الگوریتم از دو نوع منبع استفاده شده است:

۱. منابعی که فراوانی آیات قرآنی در آن بالاست.
۲. منابعی که آیات قرآنی در آن کمتر وجود دارد.

نوم	لا	و	سنة	تأخذه	لا	القيوم	الحي	هو	الذي	الله
122 376 7790 13482	117 121 8512 15890	17 53 120 6246 9452	12 34 79 119 6200	118 9430 14560	117 121 8512 15890	116 6242 9678	72 115 5678 6241	114 743 4567 6240 10451	42 193 3556 7467	2 332 7546 9567

شکل (۱): آرایه‌های عددی مربوط به اندیس کلمات در قرآن

نوم	لا	و	سنة	تأخذة	لا	القيوم	الحي	هو	الذي	الله
122 376 7790 13482	117 121 8512 15890	17 53 120 6246 9452	12 34 79 119 6200	118 9430 14560	117 121 8512 15890	116 6242 9678	72 115 5678 6241	114 743 4567 6240 10451	42 193 3556 7467	2 332 7546 9567

شکل (۲): توالی اندیسی بین برخی از عناصر آرایه‌ها و مشخص شدن آیه‌های قرآنی

سپاسگزاری

این مقاله با حمایت و پشتیبانی مرکز تحقیقات کامپیوتری علوم اسلامی نور نگارش یافته است. همچنین از جناب آقای بیژنی که در پیاده‌سازی این سیستم ما را یاری فرمودند کمال تشکر و سپاسگزاری را داریم.

مراجع

- [1] Bhukya Raju, Somayajulu DVLN, "Index Based Multiple Pattern Matching algorithm using DNA Sequence and pattern count", *International Journal of Information Technology and Knowledge Management*, Volume 4, No. 2, pp. 431-441, July-December 2011.
- [2] Kouzinopoulos Ch S., Michailidi P D., Margaritis K G., "Parallel Processing of Multiple Pattern Matching Algorithms for Biological sequences: Methods and Performance Results", *Systems and Computational Biology - Bioinformatics and Computational Modeling*, ISBN 978-953-307-875-5, pp. 161-182, InTech, September 2011.
- [3] Moraru Iulian, Andersen David G., "Exact Pattern Matching with Feed-Forward Bloom Filters". 2011 by SIAM.
- [4] Salmela Leena, "Multi-Pattern String Matching with Very Large Pattern Sets". *ACM Journal of Experimental Algorithmic*, Volume 11, 2006, November 1st 2007.
- [5] Scarpazza, D.P., Villa, O.; Petrini, F., "High-speed string searching against large dictionaries on the Cell/B.E. Processor". *Parallel and Distributed Processing*, 2008. IEEE International Symposium on. ISSN: 1530-2075, pp. 1 - 12. 14-18 April 2008, Miami, FL.
- [6] Aho A., Corasick M.; Laboratories Bell. "Efficient String Matching: An Aid to Bibliographic Search", in *Communications of the ACM*, vol. 8, 1975, pp. 333-340.
- [7] C. Allauzen, M. Crochemore ; M. Raffinot, "Efficient Experimental String Matching by Weak Factor Recognition", in *Proc. 12th Annu. Symp. on Combinatorial Pattern Matching*, Jerusalem, July 1-4, 2001, pp. 51-72.
- [8] T Jianlong, L Xia; L Yanbing; L Ping "Speeding up Pattern Matching by Optimal Partial String Extraction", in *The First International Workshop on Security in Computers, Networking and Communications (INFOCOM WKSHPs)*, 2011 IEEE, E-ISBN: 978-1-4577-0248-8, Shanghai, 23 June 2011, pp. 1030 - 1035.

۱. بر اساس کلماتی که قرآن اعلام شده‌اند.
 ۲. بر اساس کلماتی که قرآن اعلام نشده‌اند.
- در این مقاله به هر دو مورد اشاره شده و سپس برای نتیجه‌گیری کلی، میانگین هندسی این دو نیز محاسبه شده است. جدول‌های ۱ و ۲ نتایج این آزمایش‌ها را نشان می‌دهند.

جدول (۱): متنی ۲۴۱۸ کلمه و فراوانی آیات بالا

F-Measure	Precision	Recall	نوع کلمات
۰.۹۲۸۶۴۱	۰.۸۶۶۷۸۸	۱	کلماتی که قرآن تشخیص داده شده‌اند
۰.۹۸۰۸۵	۰.۹۶۲۴۲	۱	کلماتی که قرآن تشخیص داده نشده‌اند
۰.۹۵۴۰۳	۰.۹۲۸۶۴	۱	میانگین هندسی کل

جدول (۲): متنی با ۳۸۰۲ کلمه و فراوانی آیات کم

F-Measure	Precision	Recall	نوع کلمات
۰.۸۸۰۱۰	۰.۸۱۳۶۷	۰.۹۵۸۳۳	کلماتی که قرآن تشخیص داده شده‌اند
۰.۹۸۶۲۱	۰.۹۹۵۵۵	۰.۹۷۷۰۴	کلماتی که قرآن تشخیص داده نشده‌اند
۰.۹۳۰۱۴	۰.۸۹۵۴۷	۰.۹۶۷۶۰	میانگین هندسی کل

۶- نتیجه‌گیری

در این مقاله الگوریتم سیستم تشخیص هوشمند آیات قرآن در متون دیجیتالی ارزیابی شد. در حال حاضر، گام اول تدوین نرم‌افزار، یعنی «پیاده‌سازی موتور جستجو» برداشته شده است. این موتور جستجو ضریب اطمینان بسیار بالایی از حیث جامعیت و درصد مناسبی از مانعیت را دارد و جهت‌گیری کنونی آن است که مانعیت نرم‌افزار افزایش یابد (هشدارهای غیر مفید کاهش یابند).

هم‌چنین برخی چشم اندازهای دیگر این سیستم عبارتند از:

- ایجاد سیستم‌های هوشمند مکمل جستجو، با استفاده از متون جستجوی تکمیلی کمکی، نظیر: «قال تعالی»، «قرآله عز و جل» و... که در اطراف متون قرآنی مشاهده می‌شوند
- تشخیص هوشمند اختلاف قرائات و رسم الخطها
- انتخاب هوشمند آدرس صحیح از میان آدرس‌های مشترک.
- اعراب‌گذاری هوشمند آیات قرآن

- [9] N Singla, D Garg, "String Matching Algorithms and their Applicability in various Applications", *International Journal of Soft Computing and Engineering (IJSC)*, ISSN: 2231-2307, Volume-I, Issue-6, pp. 218-222, January 2012.
- [10] R Kandhan, N Teletia, J M. Patel, "SigMatch: Fast and Scalable Multi-Pattern Matching", *Proceedings of the 36th International Conference on Very Large Data Bases*, Vol. 3, No. 1, September, 2010, Singapore.
- [۱۱] محمدی نصیر، مجتبی، برخی مسایل در طراحی موتور جستجوی قرانی وب، رساله آرایه شده برای دریافت کارشناسی ارشد نرم افزار کامپیوتر، دانشگاه صنعتی شریف، دانشکده مهندسی کامپیوتر، استاد راهنما، دکتر محمد قدسی، دی ماه ۱۳۸۶.
- [12] Hammo, B., Sleit, A; El-Haj, M.. "Effectiveness of Query Expansion in searching the Holy Quran". In *Proceeding of the Second International Conference on Arabic Language Processing CITALA'07*, pages 1-10, Rabat, Morocco 2007. CITALA.

زیر نویس ها

-
- ¹- Multiple Pattern Matching
 - ²- Information Retrieval
 - ³- Tree-Based Algorithms
 - ⁴- Exact String Matching
 - ⁵- Inexact/approximate String Matching
 - ⁶- Quick-Search
 - ⁷- Exact Multi-Pattern Matching
 - ⁸- Shift table-based
 - ⁹- Automaton based
 - ¹⁰- Data sets
 - ¹¹- Recall
 - ¹²- Precision
 - ¹³- F-Score