



New Approaches to
Quran and Hadith Studies

Vol. 50, No. 2, Issue 101

Autumn & Winter 2018 - 2019

DOI: 10.22067/naqhs.v50i2.44674

رهیافت‌هایی در
علوم قرآن و حدیث

سال پنجاهم، شماره ۲، شماره پیاپی ۱۰۱

پاییز و زمستان ۱۳۹۷، ص ۱۲۰-۱۰۳

خوشه‌بندی سوره‌های قرآن با تکنیک‌های داده‌کاوی*

محسن صوفی

دانش‌آموخته کارشناسی ارشد، دانشگاه علم و صنعت ایران

Email: mohsen.soofi@gmail.com

دکتر علی‌رضا علی احمدی^۱

استاد دانشگاه علم و صنعت ایران

Email: pe@iust.ac.ir

دکتر حسین علی احمدی

استادیار دانشگاه علم و صنعت ایران

Email: aliahmadi@iust.ac.ir

دکتر بهروز مینایی

دانشیار دانشگاه علم و صنعت ایران

Email: b_minaei@iust.ac.ir

چکیده

قرآن معجزه جاوید پیامبر اسلام و کتاب آسمانی مسلمانان است. این کتاب مقدس از ۱۱۴ سوره تشکیل شده که هر سوره در بر دارنده موضوعات مختلفی است. بیشتر موضوعاتی که قرآن به آن‌ها پرداخته است در بیش از یک سوره مطرح شده‌اند و این موضوعات مشترک بین سوره‌ها، نشانه ارتباط بین آن‌ها است. شناخت این ارتباطات موضوع مهمی است که می‌تواند به فهم بهتر قرآن کمک نماید. در این مقاله با کمک یکی از تفاسیر قرآن (تفسیر راهنما)، موضوعات مطرح شده در هر یک از سوره‌ها شناسایی گردیده و میزان شباهت بین سوره‌ها بر اساس موضوعات مشترک آن‌ها، با استفاده از معیاری کمی و بهره‌گیری از روش‌های داده‌کاوی محاسبه گردیده است. در نهایت درختواره شباهت موضوعی سوره‌ها بر اساس معیار کمی مورد نظر ترسیم و با استفاده از آن سوره‌های قرآن به خوشه‌هایی افزاز گردیده‌اند.

کلیدواژه‌ها: فهم قرآن، فناوری اطلاعات، داده‌کاوی، خوشه‌بندی، شباهت موضوعی سوره‌ها.

*. تاریخ وصول: ۱۳۹۳/۱۲/۰۵؛ تاریخ تصویب نهایی: ۱۳۹۶/۱۱/۳۰.

۱. نویسنده مسئول

مقدمه

قرآن نخستین منبع شناخت فرهنگ و معارف اسلامی است که در دسترس مسلمانان قرار دارد. قرآن مجید از ۱۱۴ سوره تشکیل شده است. سوره در اصطلاح علوم قرآنی، به گروهی از آیات قرآن اطلاق می‌شود که دارای آغاز و انجامی است (سیوطی، ۱۳۶۰). نباید تصور کرد که هر سوره‌ای از قرآن در حکم فصلی از یک کتاب است چرا که معمولاً در یک کتاب، فصول از نظر محتوا مستقل از یکدیگر بوده و یا حداکثر مرتب بر هم هستند، در صورتی که سوره‌های قرآن، از نظر مباحثی که در آن مطرح شده هیچگونه ترتب و یا استقلالی نسبت به یکدیگر ندارند و موضوع واحدی مانند خداشناسی، معاد و یا حتی سرگذشت یکی از پیامبران در سوره‌های مختلف، و در هر سوره از زاویه‌ای و به شکلی (مختصر یا مفصل) آمده است. بسیاری از سوره‌های قرآن به طور همزمان چند موضوع را مورد بحث قرار داده و همچنین بسیاری از سوره‌ها به موضوعات مشابه و مرتبطی می‌پردازند. با توجه به گستردگی مباحث و نیز حجم بالای سوره‌های قرآن، بررسی این ارتباطات برای انسان کار بسیار دشواری است که امکان خطا در آن زیاد است.

در عصر حاضر، داده‌کاوی^۱ به عنوان ابزاری برای تحلیل داده‌های حجیم، به وفور در علوم مختلف مورد استفاده قرار می‌گیرد. استفاده از الگوریتم‌های داده‌کاوی، برای تحلیل و بررسی محتوای قرآن امری بدیع بوده که می‌تواند نتایج مفیدی را در بر بگیرد. برخی از الگوریتم‌های داده‌کاوی به منظور «خوشه‌بندی»^۲ داده‌های حجیم به کار می‌روند. سنجش و ارزیابی میزان شباهت موضوعی سوره‌ها به یکدیگر و نیز شناسایی سوره‌هایی از قرآن، که از نظر موضوعی به هم نزدیک و مرتبط هستند و قرار دادن آن‌ها در یک خوشه موضوعی است که در این مقاله با استفاده از تکنیک‌های داده‌کاوی به دنبال انجام آن هستیم.

پیشینه پژوهش

قرآن کریم برای مسلمانان منبعی معتبر و غنی از دانش و حکمت است. این دانش گاه آشکار و گاه پنهان است. دسترسی به علوم قرآن در منابع اسلامی، نیاز به تفسیر و استنباط دارد. بیش از هزار سال است که محققان اسلامی به دنبال استخراج دانش و قوانین، از متن قرآن و نگارش تفاسیر و زنجیره‌های استنتاج برای آن هستند و تحقیقات گسترده‌ای را در این رابطه انجام داده‌اند. اکثر این تحقیقات به شکل تدبر و تأمل و بحث پیرامون آیات بوده است. قدیمی‌ترین نوع از پژوهش‌های صورت گرفته بر روی قرآن، تحقیقاتی است که با نام «تفسیر» شناخته می‌شوند. طبق تعریف ارائه شده توسط علامه سید محمدحسین طباطبایی

^۱ Data Mining

^۲ Clustering

(۱۲۸۱-۱۳۶۰ ش)، «تفسیر عبارت است از بیان کردن معنای آیه‌های قرآن، روشن کردن و پرده‌برداری از اهداف و مفاهیم آیه‌ها».

دسته‌ای دیگر از پژوهش‌های صورت گرفته در مورد قرآن، جنبه آماری دارد و بر داده‌های عددی متکی است. مشهورترین این تحقیقات توسط رشاد خلیفه (۱۹۳۵-۱۹۹۰ م) و عبدالرزاق نوفل (۱۹۱۷-۱۹۸۴ م) صورت گرفته است. البته از این میان، نظرات رشاد خلیفه به شدت توسط مجامع رسمی مسلمانان مردود اعلام گردیده است.

رشاد خلیفه (خلیفه، ۱۹۸۲ و ۱۹۸۳) مدعی کشف یک نظم بر پایه ریاضی در قرآن شد. وی عدد ۱۹ را مخرج مشترک سراسر سیستم ریاضی قرآن می‌دانست. اینکه قرآن ۱۱۴ سوره دارد، و این عدد مضرب ۱۹ است، یا اینکه «بسم الله الرحمن الرحیم» ۱۹ حرف دارد، و تکرار هر یک از کلمات آن در قرآن، باز هم مضرب ۱۹ است، نمونه‌ای از مسائل مطرح شده در نظریه اوست. طبق ادعای او، این نظم در مورد تمام آیات قرآن، به غیر از دو آیه پایانی سوره توبه صادق است. از همین رو، وی ادعا کرد که این دو آیه جزء قرآن نازل شده بر پیامبر (ص) نیست و بعدها به آن افزوده شده است (پهلوان و شفیع، ۱۳۸۸؛ عطایی، ۱۳۸۹).

عبدالرزاق نوفل (نوفل، ۱۹۸۷) تحقیقات خود را بر پایه اعجاز عددی قرآن انجام داده است. وی فراوانی واژه‌ها و اصطلاحات استفاده شده در قرآن را شمارش کرده و ضمن مقایسه اعداد حاصله با حقایق بیرونی، عمدتاً این واژه‌ها را به صورت دوتایی و در مواردی نادر، سه و حتی چهارتایی از لحاظ فراوانی مورد مقایسه قرار داده است؛ او مدعی کشف روابط عددی جالبی در قرآن شده؛ مانند اینکه تعداد دفعات ظهور لفظ «دنیا» و «آخرت» در متن قرآن برابر، و به تعداد ۱۱۵ بار است، یا تعداد دفعات ظهور کلمه «شهر» به معنای ماه در قرآن ۱۲ مرتبه است که برابر با تعداد ماه‌های یک سال است.

شروان عطایی (عطایی، ۱۳۸۹) یکی دیگر از محققانی است که با استفاده از اطلاعات آماری، به نوعی تحلیل محتوا را بر روی قرآن انجام داده است. وی در تحقیقی ضمن محاسبه نمودن فراوانی تکرار کلمات در سوره‌های قرآن، تعداد کلمات مشابه بین هر دو آیه را به عنوان معیاری برای ارزیابی شباهت آیات در نظر گرفته و با توجه به تعداد آیات مشابه در هر دو سوره قرآن، ماتریس شباهت سوره‌ها را ایجاد نموده است. همچنین با توجه به فراوانی کلمات در هر سوره، کلماتی با بیشترین فراوانی به عنوان شاخصی برای تشخیص محورهای اصلی یک سوره در نظر گرفته شده است.

در سال‌های اخیر گروهی از محققان تلاش نموده‌اند تا دانش نهفته و پنهان موجود در قرآن را با کمک سیستم‌های کامپیوتری استخراج و ارائه نمایند. در چند پژوهش از تکنیک‌های داده‌کاوی، برای کشف

دانش پنهان استفاده شده است. الدهلان^۱ و زکی^۲ (الدهلان و زکی، ۲۰۱۰) در پژوهشی تلاش نموده‌اند این مساله را روشن سازند که چگونه تکنیک‌های داده کاوی می‌توانند دانش اسلامی را از منابع آن استخراج، و آن را به شکلی معنادار و بامفهوم برای کاربر ارائه نمایند.

در پژوهشی دیگر که توسط عمران علی^۳ (علی، ۲۰۱۲) صورت گرفته، روشی برای ارائه پیکره متنی قرآن به شکل گراف پیشنهاد شده است. وی برای استخراج و شناسایی الگوهای پرتکرار در قرآن، الگوریتمی برای کشف دنباله مفاهیم پرتکرار در آیات را با استفاده از گراف مذکور اعمال نموده است. در این تحقیق، روش پیشنهادی برای چهار سوره ابتدایی قرآن و به منظور یافتن دنباله مفاهیم پرتکرار پیاده‌سازی گردیده و در نهایت توضیح داده شده که چگونه الگوهای پرتکرار، می‌توانند برای نمایه سازی مفهومی^۴ و خوشه‌بندی آیات مشابه استفاده شوند.

طریق محمود^۵ و همکارانش (محمود و همکاران، ۲۰۱۳) نیز در پژوهشی، تکنیک‌های داده‌کاوی را برای استخراج الگوهای پنهان و غیربديهی از متن قرآن به کار گرفته‌اند. آن‌ها در پژوهش خود، آیات قرآن را با توجه به جزئی که در آن قرار دارند، تقسیم‌بندی و مجزا نموده‌اند. آیات هر جزء، به عنوان یک نمونه یا رکورد^۶، که سطری را در جدولی تشکیل می‌دهند، دیده شده‌اند. هر کلمه سازنده آیات، در رکورد مربوطه، خصیصه‌ای را ایجاد کرده است. البته در این بین، کلماتی که مستقلاً بار معنایی آن‌ها مشخص نبوده است و در خوشه‌بندی نقش آفرینی نداشته‌اند، حذف گردیده‌اند (نظیر برخی از حروف مقطعه). در مرحله بعدی یکی از الگوریتم‌های خوشه‌بندی بر جدول حاصل از آیات اعمال گردیده و نتیجه حاصله، خوشه‌بندی آیات هر جزء با توجه به کلمات تشکیل دهنده آن آیات است. در نهایت با توجه به خصایص مشترک اعضای یک خوشه، مهم‌ترین عناوین و موضوعات برای هر یک از خوشه‌ها، همراه با تعداد خوشه‌های حاوی این موضوعات استخراج گردیده است. تعداد زیاد تکرار یک عنوان در خوشه‌ها نشانه اهمیت موضوع در آن جزء است. در تحقیق مذکور با استفاده از این شیوه، در مجموع ۲۲۶ موضوع استخراج شده است. در پایان، به منظور تجزیه و تحلیل چگونگی احتمال ظهور موضوعات استخراج شده با یکدیگر، از الگوریتم Apriori، که در کاوش قوانین انجمنی کاربرد دارد، استفاده شده است.

۱ Kawther A. Aldhlan

۲ Ahmed M. Zeki

۳ Imran Ali

۴ Subjective indexing

۵ Tariq Mahmood

۶ Record

ابن کثیر (۷۰۱-۷۷۴ ق) یکی از مفسران قرآن، در مقدمه کتاب خود با عنوان «تفسیر القرآن العظیم» بهترین و صحیح‌ترین شیوه تفسیر را آن می‌داند که قرآن با خود قرآن تفسیر گردد. بر همین اساس و برای شناسایی آیات که با هم ارتباط دارند، در تحقیقی که توسط عبدالباقی شرف^۱ و آتول^۲ (شرف و اتول، ۲۰۱۲) انجام شده، ارتباط آیات قرآن با یکدیگر، بر اساس الفاظ استفاده شده در آیات مورد مطالعه قرار گرفته است. روش استفاده شده در این تحقیق، شناسایی ریشه الفاظ بکار رفته در هر آیه و سپس محاسبه شباهت آیات بر اساس ریشه‌های اشتراکی است.

نکته قابل توجه در مورد پژوهش‌های اشاره شده، تمرکز آن‌ها بر لغات و کلمات به صورت منفرد و مجزا، و عدم در نظر گرفتن ارتباطات بین کلمات به کار رفته در یک آیه است. در واقع شباهت لغوی آیات، معیار خوشه‌بندی، نمایه‌سازی مفهومی و ارتباطات آن‌ها بوده است. این درحالی است که با توجه به شیوه کنایی و تمثیلی قرآن، به نظر می‌رسد لغات مشترک بین آیات، معیار چندان مناسبی برای ارزیابی شباهت بین آن‌ها نیست. چه بسا الفاظی مشترک در دو آیه، با مقاصدی متفاوت به کار رفته باشند. لذا شیوه به کار رفته در این پژوهش‌ها، برای متن قرآن چندان مناسب نیست چون اکثر پژوهش‌ها بر پایه ارتباطات لفظی مندرج در آیات تمرکز داشته‌اند. ایان باربور^۳ (باربور، ۱۹۶۸) نیز در کتاب علم و دین، توجه به زبان کنایی نموده و تفاوت انسان و حیوان را در فهم زبان کنایی می‌داند. فارغ از صحیح یا غلط بودن این نظر، فهم کامل زبان کنایی توسط انسان هم با دشواری‌هایی روبروست، و لزوماً همه افراد قادر به فهم آن نیستند. همچنین به دلیل پیچیدگی‌های زیاد زبان کنایی، این نحوه بیان مطالب، تا به امروز توسط هیچ برنامه پردازش زبان طبیعی^۴، به طور کامل قابل فهم و پردازش نبوده است.

با توجه به مطالب بیان شده، به نظر می‌رسد برای استفاده از برخی تکنیک‌های داده‌کاوی در قرآن و حصول نتیجه مناسب از آن‌ها، بهتر است به جای استفاده از واژگان آیات، از موضوعات آیات استفاده شود. همچنین برای شناخت موضوعات آیات و استخراج آن‌ها نیز، بهتر است که از تفاسیر قرآن بهره ببریم؛ تفاسیر ارزشمندی که توسط علمای اسلام به رشته تحریر درآمده‌اند، و حاصل خیرگی و حکمتی هستند که هزاران سال تلاش و دانش اندیشمندان بزرگ پشتوانه آن است. هرچند این دو از یک سنخ نیستند و این‌ها در طول یکدیگر قرار دارند و این سنت تفسیری است که نزد مفسر و اندیشمند قرآنی بعد از خود نصوص قرار دارد.

^۱ Abdul-Baqee M. Sharaf

^۲ Eric S. Atwell

^۳ Ian Graeme Barbour

^۴ Natural language processing

روش تحقیق

در سال‌های اخیر، با هدف بهره‌برداری از مجموعه‌های بزرگ داده، تحقیقات جدیدی با موضوع کشف دانش از پایگاه داده‌ها (KDD^۱) و داده‌کاوی پدید آمده است. با وجود اینکه اغلب واژه داده‌کاوی و کشف دانش به جای یکدیگر استفاده می‌شود، در معنای واقعی و با بیانی دقیق‌تر، داده‌کاوی یکی از گام‌های فرآیند کشف دانش از مجموعه داده‌های ساختارمند انبوه است که اصطلاحاً در حوزه علوم کامپیوتر با نام پایگاه داده^۲ نامیده می‌شود. اسامه فیاض^۳ و همکارانش (فیاض و همکاران، ۱۹۹۶) کشف دانش از پایگاه داده‌ها را فرآیند شناسایی الگوها و مدل‌های جدید، بالقوه مفید، قابل فهم و پنهان در داده‌ها تعریف نموده‌اند. در شکل ۱ پنج مرحله اصلی فرآیند کشف دانش از پایگاه داده‌ها نمایش داده شده است.



شکل ۱- گام‌های فرآیند کشف دانش از پایگاه داده‌ها (فیاض و همکاران، ۱۹۹۶)

داده‌کاوی تکنیک‌های مختلفی را در بر می‌گیرد؛ همه این تکنیک‌ها به دنبال یافتن ارتباطات، وابستگی‌ها و الگوهای جدید و پنهان در داده‌ها هستند، اما اهداف آن‌ها متفاوت است. یکی از تکنیک‌های مهم و پرکاربرد داده‌کاوی خوشه‌بندی است. خوشه‌بندی به منظور گروه‌بندی n داده که هر یک شامل p خصیصه هستند طراحی گردیده، به طوری که بین اعضای موجود در یک خوشه شباهت حداکثری وجود داشته باشد و اعضای موجود در خوشه‌های مختلف شباهتشان حداقل باشد. در این تکنیک معمولاً تعداد خوشه‌ها از قبل مشخص نیست.

خوشه‌بندی به دو شیوه عمده می‌تواند صورت گیرد. یکی به صورت افزایشی^۴ و دیگری به شیوه سلسله‌مراتبی^۵. در شکل افزایشی، داده‌ها به نحوی خوشه‌بندی می‌شوند که اشتراک دو به دو خوشه‌ها با هم تهی است و اجتماع آن‌ها، همه داده‌ها را در بر می‌گیرد. به عبارتی دیگر هر یک از داده‌ها عضو یک و فقط یک

^۱ Knowledge Discovery in Databases

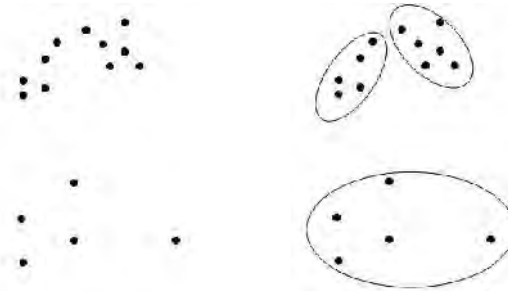
^۲ Data base

^۳ Fayyad Usama

^۴ Partitional clustering

^۵ Hierarchical clustering

خوشه هستند. در شکل ۲ خوشه‌بندی افزایی به صورت شماتیک برای داده‌های فرضی نمایش داده شده است.

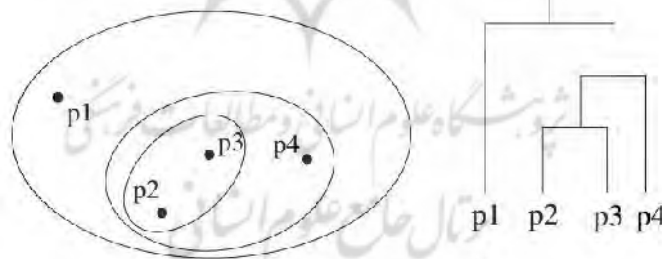


الف- داده‌های اولیه

ب- نمونه‌ای از خوشه‌بندی افزایی بر روی داده‌ها

شکل ۲- خوشه‌بندی افزایی

در شیوه سلسله مراتبی، بسته به دقت دید تعداد خوشه‌ها متفاوت است. در یک نگاه کلان، کل داده‌ها در یک خوشه قرار دارند. اما اگر دقت دید را بالاتر ببریم داده‌ها به دو خوشه تقسیم می‌شوند. با افزایش دقت دید، تعداد خوشه‌ها نیز افزایش خواهد یافت تا آنجا که در دقیق‌ترین دید هر یک از داده‌ها دارای خوشه‌ای مجزا هستند. بنابراین در شکل سلسله مراتبی، مجموعه‌ای از خوشه‌های تودرتو را داریم که در قالب یک درختواره^۱ ساماندهی شده‌اند. در شکل ۳ خوشه‌بندی سلسله مراتبی به صورت شماتیک برای داده‌های فرضی نمایش داده شده است.



الف- خوشه‌های تودرتو

ب- درختواره

شکل ۳- خوشه‌بندی سلسله مراتبی

با استفاده از یک خط افقی می‌توان هر سطحی از درختواره را برش زد و به یک خوشه‌بندی افرازی دست یافت. بدین شکل که در پایین خط افقی ترسیم شده بر روی درختواره، تعدادی درختواره جدید خواهیم داشت، که اشتراک دو به دوی عناصر تشکیل دهنده آن‌ها تهی است و اجتماع همه آن‌ها کل داده‌ها را در بر می‌گیرد.

پیش از این گفته شد که هدف از خوشه‌بندی ایجاد دسته‌بندی در داده‌ها است، به نحوی که داده‌های درون یک خوشه بیشترین شباهت را با یکدیگر داشته و با داده‌های سایر خوشه‌ها در کمینه تشابه باشند. حال سؤال اصلی این است که معیار ما برای ارزیابی و سنجش میزان تشابه داده‌ها چه می‌تواند باشد. مشابهت بین دو چیز معمولاً توسط یک عدد بیان می‌گردد که هر چه مقدار آن بیشتر باشد بیانگر شباهت بیشتر است. البته اغلب از مقداری بین صفر (عدم تشابه) و یک (کامل تشابه) برای بیان میزان شباهت استفاده می‌کنند.

معیارهای زیادی برای اندازه‌گیری شباهت، بین دو شیء داده‌ای که تنها شامل ویژگی‌های دودویی^۱ هستند وجود دارد. در ادامه برخی از این معیارها را معرفی خواهیم نمود.

فرض کنید p و q دو شیء داده هستند، که هر یک دارای n ویژگی دودویی می‌باشند. از مقایسه نظیر به نظیر ویژگی‌های این اشیاء داده، چهار کمیت حاصل می‌گردد، که عبارت‌اند از:

$$- M_{01} = \text{تعداد ویژگی‌هایی که ارزش آن‌ها در } p \text{ صفر و در } q \text{ یک است.}$$

$$- M_{10} = \text{تعداد ویژگی‌هایی که ارزش آن‌ها در } p \text{ یک و در } q \text{ صفر است.}$$

$$- M_{00} = \text{تعداد ویژگی‌هایی که ارزش آن‌ها در } p \text{ صفر و در } q \text{ نیز صفر است.}$$

$$- M_{11} = \text{تعداد ویژگی‌هایی که ارزش آن‌ها در } p \text{ یک و در } q \text{ نیز یک است.}$$

یکی از معیارهایی شباهتی که به طور معمول مورد استفاده قرار می‌گیرد، ضریب تطابق ساده^۲ است.

این معیار اغلب با نماد SMC نمایش داده می‌شود و مقدار آن با استفاده از رابطه ۱ محاسبه می‌گردد.

$$\text{رابطه ۱} \quad SMC = \frac{M_{01} + M_{10} + M_{11} + M_{00}}{\text{تعداد کل ویژگی‌ها}}$$

در معیار SMC، عدم وجود یک ویژگی در هر دو شیء داده (M_{00})، به عنوان شباهت در نظر گرفته می‌شود. مثلاً در مقایسه دو آیه از منظر این معیار، اینکه هیچ کدام از آن‌ها راجع به نماز صحبت نکرده‌اند به عنوان دلیلی برای شباهت بیشتر دیده خواهد شد. این موضوع چندان منطقی نیست، چرا که در مورد آیات،

^۱ Binary

^۲ Simple Matching Coefficient

موضوعات مطرح شده مهم هستند و نه موضوعاتی که مطرح نشده‌اند؛ البته این معیار در برخی موارد می‌تواند مناسب باشد، اما برای منظوری که در این مقاله به دنبالش هستیم مناسب نیست.

با حذف M_{00} از معیار SMC معیار جدیدی به نام جاکارد^۱ حاصل می‌گردد. نحوه محاسبه معیار جاکارد در رابطه ۲ بیان گردیده است.

$$J = \frac{M_{01} + M_{10} + M_{11}}{\text{تعداد ویژگی‌هایی که حداقل در یکی از داده‌ها وجود دارند}}$$

رابطه ۲

بعنوان مثال اگر در آیه A راجع به a موضوع و در آیه B راجع به b موضوع صحبت شده باشد، و c موضوع در این دو آیه مشترک باشد، مقدار معیار جاکارد برابر $\frac{c}{a+b-c}$ خواهد بود. صورت این کسر موضوعات مشترک بین دو آیه و منخرج آن اجتماع موضوعات مطرح شده در دو آیه می‌باشد.

با بکارگیری هر یک از معیارهای شباهت، می‌توان ماتریسی مربعی را تشکیل داد، که هر سطر و ستون آن نماینده یکی از اشیاء داده است. در این ماتریس، در تقاطع سطرها و ستون‌ها میزان مجاورت دو عنصر داده مشخص گردیده است. این ماتریس به نام ماتریس مجاورت^۲ معروف است و در همه الگوریتم‌های خوشه‌بندی مورد استفاده قرار می‌گیرد. در کل به نظر می‌رسد از بین معیارهای موجود با استفاده از معیار جاکارد، با منطقی درست می‌توان شباهت موضوعی بین سوره‌های قرآن را محاسبه نمود. بنابراین در این تحقیق برای محاسبه شباهت سوره‌ها از این معیار استفاده خواهیم کرد.

تا اینجا معیارهایی برای محاسبه میزان مجاورت دو شیء داده بیان گردید. در خوشه‌بندی سلسله مراتبی در مرحله اول نزدیک‌ترین اشیاء داده، با هم خوشه‌ای را تشکیل می‌دهند. بدین ترتیب در مراحل بعدی به جای اشیاء داده با خوشه‌های داده‌ها مواجه هستیم و لازم است که معیارهایی را هم برای سنجش میزان مجاورت خوشه‌ها با یکدیگر در نظر بگیریم. برای سنجش میزان مجاورت خوشه‌ها هم معیارهای زیادی تعریف شده است. معیارهای اتصال کامل (CL^۳) یا دورترین همسایه، تک اتصال (SL^۴) یا نزدیک‌ترین همسایه، میانگین همه فاصله‌ها (AL^۵ یا UPGMA^۶) و روش وارد^۷ از مشهورترین معیارها در این زمینه هستند. در این پژوهش از روش وارد برای سنجش میزان فاصله بین خوشه‌ها استفاده شده است. در این

۱ Jaccard

۲ Proximity Matrix

۳ Complete Link

۴ Single Link

۵ Average Link

۶ Unweighted Pair Group Method using arithmetic Average

۷ Ward

روش، مجموع مربع فواصل^۱ هر یک از اعضای خوشه‌هایی که قرار است با هم ادغام گردند از مرکز خوشه فرضی حاصل از ادغام محاسبه می‌گردد و ابتدا خوشه‌هایی با هم ادغام می‌شوند که در آن‌ها، این مقدار نسبت به سایرین کمترین باشد. به بیان دیگر، روش وارد ابتدا خوشه‌هایی را با هم ادغام می‌کند که اعضای آن‌ها با شدت بیشتری نسبت به سایر خوشه‌ها حول مرکز خوشه فرضی حاصل از ادغام متمرکز شده باشند (تن و همکاران، ۲۰۰۵).

داده‌های تحقیق

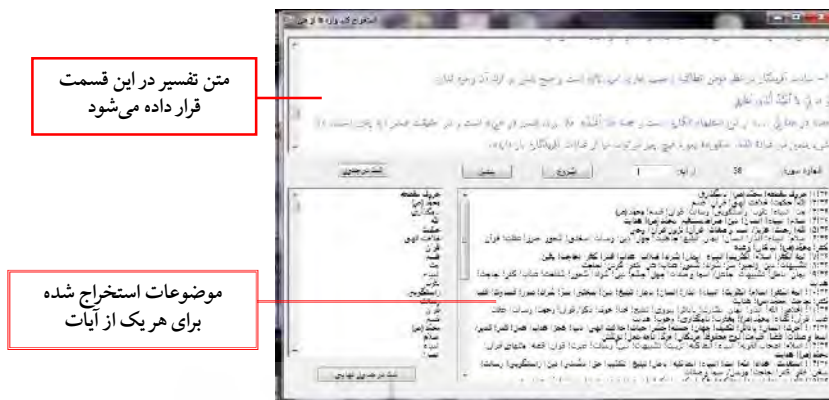
به دلیل تقدس قرآن و نیز زبان کنایی و تمثیلی آن، متن کاوی مستقیم آیات به شیوه معمول چندان جالب نخواهد بود. مشخص نمودن موضوعات و مفاهیم آیات با توجه به شیوه تمثیلی و کنایی قرآن، امری حساس، پیچیده و تخصصی است، و حتماً باید با نظر خبره^۲ صورت گیرد. در این تحقیق از تفسیر قرآن به جای خود قرآن استفاده شده است و موضوعات آیات با تشخیص خبره (مفسران) از آیات استخراج گشته است. باید توجه داشت که در بسیاری از تفاسیر قرآن، تفسیر لزوماً به صورت آیه به آیه انجام نشده است. به عبارتی دیگر، در تفسیر بسیاری از آیات، دسته آیات، کل سوره، و حتی کل قرآن مدنظر مفسر بوده و در تفسیر هر آیه، از مطالب سایر آیات نیز استفاده شده است. آنچه ما در انجام این تحقیق به دنبالش هستیم یافتن همین ارتباطات است؛ لیکن با استفاده از تحلیل‌های محاسباتی و الگوریتم‌های ماشینی، و نه با اعمال نمودن نظرات خبرگان. به بیان دیگر ما نیازمند تفسیری هستیم که در آن خبره هر آیه را مستقل از سایر آیات مورد بررسی قرار داده باشد. از بین تفاسیر موجود، تفسیر راهنما (هاشمی رفسنجانی، ۱۳۷۹) آیه به آیه پیش رفته و ارتباطات آیات را با هم در نظر نگرفته است. به همین دلیل داده‌های خود را از این تفسیر جمع‌آوری نموده‌ایم. برای انجام این پژوهش و به منظور جمع‌آوری موضوعات آیات، نگارندگان، نرم‌افزاری را توسعه دادند که متن الکترونیکی تفسیر را به عنوان ورودی دریافت و با پویس^۳ این متن، و نمایه‌های موضوعی که توسط تفسیر راهنما در ذیل هر آیه با انواع دلالت‌های مطابقی و نیز غیرمطابقی نظیر دلالت‌های التزامی و سیاقی استنباط شده است را در قالب پایگاه داده‌ای ذخیره‌سازی شده و بستری برای گونه‌های مختلف جستجوی لفظی و موضوعی فراهم نموده است. نرم‌افزار فراهم شده می‌تواند بسامد لفظ یا مفهوم مورد جستجو و محل رخداد آن‌ها را در متن تفسیر ذیل هر آیه قرآن ارائه نماید. مفهوم شکل ۴

^۱ Sum of the Square Distance

^۲ Expert

^۳ Scan

واسط کاربری^۱ این نرم‌افزار را نمایش می‌دهد.



شکل ۴- واسط کاربری نرم‌افزار جمع‌آوری موضوعات آیات از تفسیر راهنما

این نرم‌افزار پس از آنکه از میان صفحات تفسیر موضوعات آیات را استخراج نمود، با پایگاه داده‌ای که از قبل برای نگهداری داده‌ها ایجاد شده ارتباط برقرار می‌کند تا موضوعات استخراج شده برای هر آیه را در جدولی ذخیره نماید. نحوه کار به این صورت است که نرم‌افزار پس از استخراج موضوعات یک آیه، آن‌ها را در ستون‌ها یا همان خصیصه‌های جدول جستجو می‌کند. در صورتی که یک موضوع پیش از این، برای یک آیه آورده شده باشد حتمه در بین فیلدها خواهد بود اما اگر این نخستین بار باشد که چنین موضوعی برای یک آیه استخراج شده، ستونی جدید به جدول اضافه خواهد شد که به آن موضوع اختصاص می‌یابد. در هر صورت پس از یافتن یا افزودن فیلد مربوط به یک موضوع، در تقاطع این فیلد با سطر متناظر با آیه، عدد یک قرار می‌گیرد. بدین ترتیب در پایان کار جدولی خواهیم داشت که ۶۲۳۶ سطر دارد، که هر یک از این سطرها نماینده یکی از آیات قرآن هستند و ستون‌های جدول نیز معرف موضوعات مطرح شده در آیات می‌باشند.

نرم‌افزار توسعه داده شده بر روی همه بیست جلد تفسیر راهنما که در آن مفاهیم کلیدی آیات توسط مفسر استخراج شده است که در ترجمه عادی قرآن چنین مفاهیمی صراحتاً وجود ندارد. جدول نهایی حاصل از اجرای نرم‌افزار دارای ۱۶۶۲ ستون می‌باشد، که هر یک از آن‌ها نماینده یکی از موضوعات مطرح شده در آیات هستند. در شکل ۵ جدول نهایی حاصل از اجرای نرم‌افزار، به صورت شماتیک نمایش داده شده است. روشن است که با دسترسی به جدول موضوعات مطرح شده در آیات، به راحتی می‌توان با

^۱ User Interface

تجمع سطریهای مربوط به آیات یک سوره به جدول موضوعات مطرح شده در سوره‌ها دست یافت.

موضوعات استخراج شده از تفسیر راهنما برای آیات

		C1	C2	C3	...	C1662
آیات قرآن	آیه ۱ سوره حمد	۱	۱	۱	...	۰
	آیه ۲ سوره حمد	+	+	۱		+
	آیه ۳ سوره حمد	+	۰	۰		۰
	⋮					
	آیه ۶ سوره ناس	+	۰	۰		۱

C3 جزء موضوعات مطرح شده در آیه ۲ سوره حمد است.

شکل ۵- جدول موضوعات آیات

مراحل و فرآیند انجام کار

در این تحقیق به دنبال آن هستیم که با در نظر گرفتن همه موضوعاتی که با تشخیص خیره از سوره‌های قرآن استخراج گردیده‌اند، ارتباطات موضوعی بین کل سوره‌های قرآن را شناسایی، و این ارتباطات را در قالب درختواره شباهت موضوعی سوره‌های قرآن ارائه نماییم. با ترسیم این درختواره، عملاً سوره‌های قرآن به صورت سلسله مراتبی خوشه‌بندی خواهند شد. بدین ترتیب که در پایین‌ترین سطح سلسله مراتب، تک تک سوره‌ها، به لحاظ موضوعی متفاوت از سایرین دیده شده و برای هر سوره، خوشه‌ای مجزا در نظر گرفته خواهد شد. با تعدیل این دید، سوره‌هایی که بیشترین موضوعات مشترک را دارند با استفاده از معیاری خاص شناسایی شده و در سطحی متناسب با اشتراکاتشان، با یکدیگر تشکیل یک خوشه را خواهند داد. در سطوح بالاتر، این خوشه‌ها هستند که بر اساس میزان شباهت موضوعی، با یکدیگر ادغام و خوشه‌های بزرگ‌تری را ایجاد خواهند نمود. روند ادغام در درختواره آنقدر ادامه خواهد یافت که در نگاهی بسیار کلان، کل ۱۱۴ سوره قرآن در یک خوشه قرار خواهند گرفت.

برای دستیابی به درختواره شباهت موضوعی سوره‌های قرآن، در گام نخست میزان شباهت یا فاصله سوره‌ها از هم، با استفاده از معیار جاکارد محاسبه و در ماتریس مجاورت^۱ ذخیره گردید. از نظر معنایی، مقدار معیار جاکارد بین دو سوره بیانگر نسبت موضوعات مشترک بین دو سوره، به کل موضوعات مطرح شده در دو سوره است. با استفاده از ماتریس مجاورت شبیه‌ترین سوره‌ها به لحاظ موضوعی شناسایی و در درختواره شباهت، توسط خطی در سطحی متناسب با میزان شباهتشان به یکدیگر متصل می‌شوند.

^۱ Proximity Matrix

همانطور که پیش‌تر نیز گفته شد برای تشکیل یک درختواره کامل از سوره‌های قرآن، علاوه بر معیار ارزیابی شباهت بین دو سوره، معیاری هم برای ارزیابی شباهت بین خوشه‌ها احتیاج داریم، تا بعد از شناسایی شبیه‌ترین دو سوره به هم و قرار دادن آن‌ها در یک خوشه، بتوانیم شباهت بین خوشه و سوره و نیز شباهت بین خوشه‌ها را مورد ارزیابی قرار داده و بر حسب میزان شباهت، آن‌ها را نیز در سطوح متناسب با یکدیگر ادغام کنیم. بنابراین در گام بعدی ماتریس مجاورت را مجدداً و این بار با استفاده از معیار جاکارد و وارد برای عناصر جدید، که شامل ۱۱۲ سوره و خوشه حاصل از ادغام شبیه‌ترین دو سوره است، تشکیل خواهیم داد. این عمل آنقدر ادامه خواهد یافت تا در نهایت همه سوره‌ها در یک خوشه قرار گیرند.

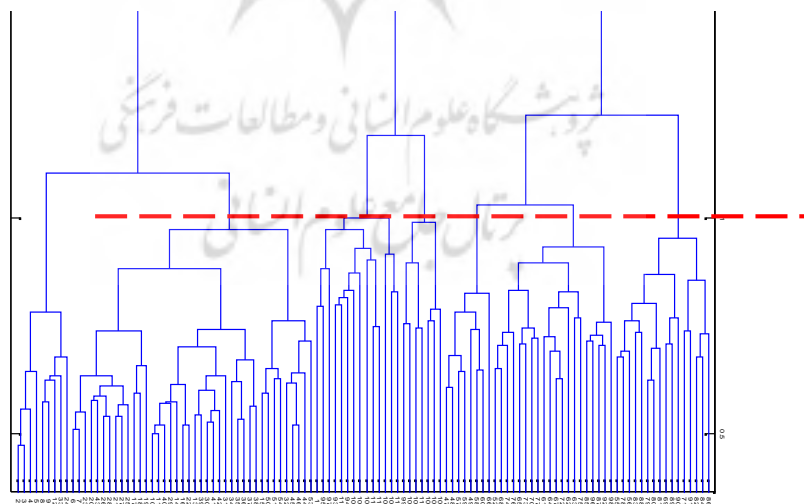
نتایج

با مبنا قرار دادن موضوعات مطرح شده در سوره‌ها و بهره‌گیری از معیارهای جاکارد و وارد می‌توان شباهت سوره‌های قرآن را به صورت سلسله مراتبی و در قالب درختواره‌ای ارائه نمود. این درختواره در شکل ۶ نمایش داده شده است. در پایین‌ترین سطح درختواره، شبیه‌ترین سوره‌های قرآن به لحاظ موضوعی، در سطحی متناسب با شباهتشان با خطی به شکل U وارونه به یکدیگر متصل شده‌اند. همانطور که در شکل ۶ مشخص است دومین و سومین سوره قرآن، یعنی سوره‌های بقره و آل عمران، از لحاظ موضوعی و بر اساس معیار جاکارد، مشابه‌ترین سوره‌های قرآن هستند. میزان شباهت این دو سوره نزدیک به ۵۳ درصد است، به عبارت دیگر موضوعات مطرح شده در این دو سوره، در ۵۳ درصد موارد مشترک است. در نقطه مقابل، بیشترین فاصله موضوعی بین صدونهمین و صدودوازدهمین سوره قرآن، یعنی سوره‌های کافرون و اخلاص است. این دو سوره از نظر موضوعی و بر اساس معیار جاکارد، تقریباً ۱۷ درصد مشابه هستند و این کمترین تشابه موضوعی بین دو سوره قرآن است. این درختواره اطلاعات زیادی را در خود دارد، و می‌تواند در تحقیقات قرآنی، به طرق مختلف مورد استفاده قرار گیرد.



شکل ۶- دندروگرام شباهت موضوعی سوره‌های قرآن بر اساس معیار جاکارد و وارد

برای آنکه با استفاده از درختواره شباهت، سوره‌های قرآن را به لحاظ موضوعی خوشه‌بندی نماییم، می‌توان در سطحی مناسب، با خطی افقی درختواره ترسیم شده را به چند زیردرختواره تقسیم نمود. هر یک از این زیردرختواره‌ها، مجموعه‌ای از سوره‌ها را در بر خواهند گرفت که در کنار هم، یک خوشه مجزا را تشکیل خواهند داد. در این میان، مسئله کلیدی و مهم، محل قرار دادن خط افقی است. با نزدیک نمودن این خط به سطوح پایین درختواره، تعداد خوشه‌ها زیاد و اندازه خوشه‌ها کوچک خواهد شد. در این حالت میزان شباهت سوره‌های قرار گرفته در یک خوشه زیاد خواهد بود. در مقابل اگر خط افقی را در سطوح بالای درختواره مستقر نماییم، خوشه‌های بزرگی خواهیم داشت که تعدادشان نسبت به حالت قبلی کمتر است. در این وضعیت، سوره‌های درون یک خوشه به اندازه حالت قبلی به یکدیگر شبیه نیستند. در کل با حرکت خط افقی به سمت سطوح بالای درختواره، تعداد خوشه‌ها کمتر، اندازه آن‌ها بزرگ‌تر و تشابه اعضایشان کمتر خواهد شد. شکل ۷ قسمتی از درختواره حاصل از به کارگیری معیارهای جاکارد و وارد را نمایش می‌دهد؛ در این شکل، خط چین افقی رسم شده بر روی درختواره باعث ایجاد هفت خوشه مجزا گردیده است. البته این خط چین می‌تواند در سطوح دیگری نیز مستقر گردد و خوشه‌بندی‌های متفاوتی را نتیجه دهد. اعضای این خوشه‌ها در جدول ۱ معرفی شده‌اند؛ همچنین موضوعاتی که در بین همه اعضای یک خوشه مشترک هستند نیز در این جدول گنجانده شده است. البته بسیاری موضوعات دیگر نیز وجود دارند که در اغلب سوره‌های یک خوشه مطرح شده‌اند و در شکل‌گیری خوشه مؤثر بوده‌اند. لذا این تصور که تشکیل خوشه‌ها تنها بر اساس موضوعات مشترک در بین همه اعضا بوده است، تلقی صحیحی نیست.



شکل ۷- برشی افقی بر درختواره شباهت موضوعی سوره‌های قرآن بر اساس معیار جاکارد و وارد

جدول ۱- معرفی خوشه‌های حاصل از برش انجام شده بر درختواره شباهت موضوعی

نام خوشه	A	B	C	D	E	F	G
تعداد اعضای خوشه	۹	۴۰	۱۴	۷	۸	۲۰	۱۶
لیست اعضای خوشه	یسره (۲) - آل عمران (۳) - نساء (۴) - مائده (۵) - انفال (۸) - توبه (۹) - یوسف (۱۲) - نور (۲۴) - آل عمران (۳۰) - غافر (۳۹) - فصلت (۴۰) - شعور (۴۲) - زخرف (۴۳) - دخان (۴۴) -	نحل (۱۶) - اسراء (۱۷) - كهف (۱۸) - مریم (۱۹) - طه (۲۰) - انبياء (۲۱) - حج (۲۲) - مؤمنون (۲۳) - فرقان (۲۵) - شعراء (۲۶) - نمل (۲۷) - قصص (۲۸) - عبكورت (۲۹) - روم (۳۰) - لقمان (۳۱) - سجده (۳۲) - سبأ (۳۴) - فاطر (۳۵) - يس (۳۶) - صافات (۳۷) -	فاتحه (۱) - صبحي (۹۳) - انشراح (۹۴) - تين (۹۵) - قدر (۹۷) - عصر (۱۰۳) - فيل (۱۰۵) - قريش (۱۰۶) - كوث (۱۰۸) - كافرون (۱۰۹) - نصر (۱۱۰) - اخلاص (۱۱۲) - فلق (۱۱۳) -	زلزال (۹۹) - عاديات (۱۰۰) - قارعة (۱۰۱) - تكوير (۱۰۲) - ممتززه (۱۰۴) - ماعن (۱۰۷) -	محمد (۴۷) - فتح (۴۸) - حجرات (۴۹) - حديد (۵۷) - مجادل (۵۸) - حشر (۵۹) -	طور (۵۲) - صف (۶۱) - جمعه (۶۲) - منافقون (۶۳) - تغابن (۶۴) - طلاق (۶۵) - ملك (۶۷) - قلم (۶۸) - معارج (۷۰) - جن (۷۲) - مزمل (۷۳) - مدثر (۷۴) - قيامه (۷۵) - انسان (۷۶) -	الر حمن (۵۵) - واقعه (۵۶) - حاقه (۶۹) - نوح (۷۱) - نبا (۷۸) - نازعات (۷۹) - عبس (۸۰) - تكوير (۸۱) - انفطار (۸۲) - مطففين (۸۳) - انشراق (۸۴) - طارق (۸۶) - غاشيه (۸۸) -

یکی از موضوعاتی که در دسترس مفسران می‌باشد موضوع مکی یا مدنی بودن سوره‌های قرآن کریم است. این موضوع ممکن است بتواند تا حدودی اعتبار خوشه‌بندی انجام شده در این تحقیق را حمایت نماید. بر اساس مطالعه انجام شده (ناصحیان، ۱۳۸۶) سوره‌های قرآن کریم از منظر مکی و مدنی بودن در سه دسته گنجانده شده‌اند: ۱- سوره‌هایی که در مکی بودن آن‌ها اتفاق نظر وجود دارد ۲- سوره‌هایی که در مدنی بودن آن‌ها اتفاق نظر وجود دارد ۳- سوره‌هایی که در مکی یا مدنی بودن آن‌ها با توجه به تفاوت در معیار مکی و مدنی بودن، اختلاف نظر وجود دارد. علیرغم برداشت‌های گوناگون و معیارهای متفاوت در تشخیص مکی و مدنی بودن سوره‌ها، نتایج مندرج در این خوشه‌ها، با نتایج این بررسی انطباق دارد، که خلاصه آن در جدول ۲ آمده است.

همانطور که مشخص است به غیر از خوشه F، بقیه خوشه‌ها از نظر مکی و مدنی بودن به میزان زیادی یک‌دست هستند. این یک‌دست بودن، خود می‌تواند شاهی بر صحت خوشه‌بندی انجام شده باشد. بررسی دقیق‌تر این خوشه‌ها نیازمند خبرگی و آشنایی بیشتر با علوم قرآنی است و می‌تواند موضوعی برای تحقیقات آتی باشد.

جدول ۲- بررسی محل نزول (مکی و مدنی بودن) اعضای خوشه‌ها

نام خوشه	A	B	C	D	E	F	G
تعداد کل اعضا	۹	۴۰	۱۴	۷	۸	۲۰	۱۶
تعداد اعضای مکی	۱	۳۹	۱۳	۶	۰	۱۳	۱۵
تعداد اعضای مدنی	۸	۱	۱	۱	۸	۷	۱
موارد خلاف اکثریت	یوسف (۱۲)	یحیی (۲۲)	نصر (۱۱۰)	زلزال (۹۹)	—	صف (۶۱) جمعه (۶۲) مناققون (۶۳) تغابن (۶۴) طلاق (۶۵) انسان (۷۶) بینه (۹۸)	رحمن (۵۵)
میزان یک‌دست بودن اعضای خوشه	%۸۹	%۹۸	%۹۳	%۸۶	%۱۰۰	%۶۵	%۹۴

نتیجه‌گیری و بحث و بررسی

به طور سنتی پژوهشگران و خبرگان علوم قرآنی، به دلیل تقرب و انس زیاد با قرآن، غالباً موضوعات کلان مطرح شده در سوره‌ها را در محفوظات خود دارند و از طریق آن‌ها می‌توانند ارتباطات بین سوره‌ها را در صورت نیاز، به طور نسبی حدس بزنند. ما در این تحقیق به دنبال آن بودیم که با در نظر گرفتن همه موضوعاتی که با تشخیص مفسر از سوره‌های قرآن استخراج گردیده‌اند، ارتباطات موضوعی بین کل سوره‌های قرآن را شناسایی، و این ارتباطات را در قالب درختواره شباهت موضوعی سوره‌های قرآن ارائه

نماییم. با ترسیم این درختواره، عملاً سوره‌های قرآن به صورت سلسله مراتبی خوشه‌بندی شدند. بدین ترتیب که در افراطی‌ترین حالت و در پایین‌ترین سطح سلسله مراتب، تک‌تک سوره‌ها، به لحاظ موضوعی متفاوت از سایرین دیده شده و برای هر سوره، خوشه‌ای مجزا در نظر گرفته شد. با تعدیل این دید، سوره‌هایی که بیشترین موضوعات مشترک را دارند با استفاده از معیاری خاص شناسایی و در سطحی متناسب با اشتراکاتشان، با یکدیگر تشکیل یک خوشه را دادند. در سطوح بالاتر، این خوشه‌ها بودند که بر اساس میزان شباهت موضوعی، با یکدیگر ادغام و خوشه‌های بزرگ‌تری را ایجاد نمودند. روند ادغام در درختواره آنقدر ادامه یافت که در نگاهی بسیار کلان، کل ۱۱۴ سوره قرآن در یک خوشه قرار گرفت. نتایج این تحقیق نشان می‌دهد که استفاده از تکنیک‌های داده‌کاوی برای بررسی و تحلیل قرآن می‌تواند مؤثر و مفید باشد، به طوری که به نظر می‌رسد بخشی از دانشی را که یک فرد خیره در طی سال‌ها انس با قرآن به آن رسیده است با استفاده از این تکنیک‌ها قابل دستیابی است.

منابع

- باربور، ایان، *علم و دین*، ترجمه بهاء‌الدین خرمشاهی، مرکز نشر دانشگاهی، ۱۳۹۳.
- پهلوان، منصور، و سعید شفیعی، «ارزیابی و نقد نظریه اعجاز عددی قرآن کریم»، *دوفصلنامه پژوهش‌های قرآن و حدیث*، دفتر ۱، پاییز و زمستان ۱۳۸۸، صص ۵۷-۷۶.
- خلیفه، رشاد، *معجزة القرآن الکریم*، بیروت: دارالعلم للملایین، ۱۹۸۳ م.
- سیوطی، عبدالرحمن بن ابی‌بکر، *الاتقان فی علوم القرآن*، قاهره، چاپ سوم، ۱۳۶۰.
- طباطبایی، محمد حسین، *المیزان فی تفسیر القرآن*، دفتر انتشارات جامعه مدرسین حوزه علمیه قم، چاپ پنجم، ۱۴۱۷ ق.
- _____، *ترجمه تفسیر المیزان*، مترجم: سید محمدباقر موسوی همدانی، دفتر انتشارات اسلامی وابسته به جامعه مدرسین حوزه علمیه قم، چاپ پنجم، ۱۳۷۴.
- عطایی، شروان، «تدبر در قرآن مجید به کمک روش‌های داده‌کاوی»، *چهارمین کنفرانس داده‌کاوی*، ۹-۱۰ آذر، تهران، ایران، ۱۳۸۹.
- معارف، مجید، «شناخت قرآن»، *مجله گلستان قرآن*، شماره ۱۲۸، (۱۳۸۱)، صص ۴۲-۴۶.
- ناصحیان، علی اصغر، «شناخت مکی و مدنی در پرتو احادیث»، *مجله سفیر*، شماره ۱، ۱۳۸۶، صص ۵-۳۲.
- نوفل، عبدالرزاق، *الاعجاز العددی للقرآن الکریم*، چاپ پنجم، بیروت، دارالکتاب العربی، ۱۹۸۷ م.
- هاشمی رفسنجانی، علی‌اکبر، *تفسیر راهنما*، قم، انتشارات دفتر تبلیغات اسلامی، چاپ سوم، ۱۳۷۹.

Aldhlan, Kawther. A., Zeki, Ahmed. M. & Zeki, Akram. M., "Datamining and Islamic knowledge extraction: alhadith as a knowledge resource", In **Information and Communication Technology for the Muslim World (ICT4M)**, 3rd International Conference on ICT4M 2010, p H-21.

Ali, Imran, "Application of a Mining Algorithm to Finding Frequent Patterns in a Text Corpus: A Case Study of the Arabic", **International Journal of Software Engineering and Its Applications**, (2012), 6(3),pp 127-134.

Fayyad, Usama M, Piatetsky-Shapiro, G., & Smyth, P., "The KDD process for extracting useful knowledge from volumes of data", **Communications of the ACM**, (1996), 39(11),pp 27-34.

Khalifa, Rashad, **Quran: Visual Presentation of the Miracle**, USA, (1982).

Mahmood, T., Atique, N., Razzaq, A., Kazmi, M., & Akhtar, R., "Mining Topical Clusters and Topical Associations from the Holy Qur'an", **International Conference on Advances in Information Technology for the Holy Quran and Its Sciences**, Tibah University (2013).

Pang-Ning, T., Steinbach, M., & Kumar, V., **Introduction to data mining**, Boston, USA, Addison-Wesley, (2005).

Sharaf, A. B. M., & Atwell, E., "QurSim: A corpus for evaluation of relatedness in short texts", In **LREC**, (2012), pp 2295-2302.