



The Study on Quranic surahs' Structured-ness and their Order Organization Using NLP Techniques

Ehsan Khadangi¹  and Hosein Haghi tarooni² 

1. Corresponding author, Assistant Professor of Computer Science Department, Faculty of Engineering, Shahed University, Tehran, Iran: khadangi@shahed.ac.ir
2. Bachelor's student in Computer Engineering at Shahed University, Tehran, Iran: hosein.haghitarooni@gmail.com

Article Info

Article type:
Research Article

Article history:
Received: 20 February 2025
Received in revised form:
10 May 2025
Accepted: 22 June 2025
Available online:
23 August 2025

Keywords:
Natural Language
Processing,
Word2vec,
Quran,
Topic Sameness,
Surahs' Structured-ness,
Tf-idf.

ABSTRACT

In recent years, the structure of the Quranic surahs has garnered significant attention from Quranic structural researchers. One of the theories presented in this regard is the theory of thematic unity, which posits that each surah of the Quran is formed around a single, unified topic. The theory of "commandments and elaboration," as one of the most important branches of thematic unity, suggests that God presents the subject of each surah in the first section, elaborates on it in various ways in the following sections, and finally draws a conclusion in the closing section. In this paper, we aim to study these two theories using natural language processing techniques. In this regard, we used three methods: TF-IDF, Word2Vec word embeddings, and the co-occurrence of roots within the verses, to calculate the similarity of Quranic roots. Subsequently, the degree of conceptual similarity between the surahs was computed and compared to a random state. The results show that the surahs of the Quran exhibit semantic coherence, in the sense that the concepts and topics within the surahs revolve around a specific subject. Furthermore, the examination of the similarity between the first and last sections, as well as the first and subsequent sections of each surah, indicates that the theory of "commandments and elaboration" holds true for many surahs. Finally, by analyzing the similarity between the surahs based on their sequential order in the Quran and their chronological revelation throughout history, it was determined that surahs that are closer in arrangement also show greater similarity to each other. Thus, it appears that the Quran as a whole has a structured arrangement, with surahs grouped into related clusters within the Quran.

Cite this article: Khadangi, E., & Haghi tarooni, H. (2025). The Study on Quranic surahs' Structured-ness and their Order Organization Using NLP Techniques. *Digital Islamic Studies and Humanities*, 1 (1), 123-154. <https://doi.org/10.22034/disah.2024.716139>



© The Author(s). **Publisher:** Research Center for Digital Islamic Studies and Humanities (RCDISAH).
DOI: <https://doi.org/10.22034/disah.2024.716139>

بررسی ساختار سوره‌های قرآن و ترتیب چینی آن‌ها با بهره‌گیری از

تکنیک‌های پردازش زبان طبیعی

احسان خدنگی^۱ و حسین حقی طارونی^۲

۱. نویسنده مسئول، استادیار، گروه علوم کامپیوتر، دانشگاه شاهد، تهران، ایران، رایانامه: khadangi@shahed.ac.ir

۲. دانشجوی کارشناسی مهندسی کامپیوتر دانشگاه شاهد، تهران، ایران، رایانامه: hosein.haghitarooni@gmail.com

اطلاعات مقاله

چکیده

نوع مقاله:

مقاله پژوهشی

تاریخ دریافت: ۱۴۰۳/۱۲/۰۲

تاریخ بازنگری: ۱۴۰۴/۰۲/۲۰

تاریخ پذیرش: ۱۴۰۴/۰۴/۰۱

تاریخ انتشار: ۱۴۰۴/۰۶/۰۱

کلیدواژه‌ها:

ساختار سوره‌ها،

انسجام معنایی سوره‌ها،

نظریه احکام و تفصیل،

پردازش زبان طبیعی،

تعبیه کلمات.

بررسی ساختار سوره‌های قرآن در سال‌های اخیر توجه پژوهشگران ساختاری قرآن را به خود جلب کرده است. یکی از نظریه‌های مطرح در این زمینه، نظریه وحدت موضوعی است که بر این باور است که هر سوره قرآن حول یک موضوع واحد شکل گرفته است. نظریه «احکام و تفصیل» به‌عنوان یکی از مهم‌ترین شاخه‌های وحدت موضوعی بیان می‌کند که خداوند متعال در فصل اول هر سوره موضوع اصلی را بیان کرده و در فصول بعدی آن را به شیوه‌های مختلف تفصیل می‌دهد و در نهایت، در فصل پایانی نتیجه‌گیری می‌کند. در این مقاله، قصد داریم این دو نظریه را با استفاده از تکنیک‌های پردازش زبان طبیعی مورد بررسی قرار دهیم. بدین منظور، با به‌کارگیری سه روش TF-IDF، تعبیه کلمات (Word2Vec)، و با هم‌آیی ریشه‌ها در آیات، به محاسبه تشابه ریشه‌های قرآنی پرداخته شد. سپس میزان شباهت مفاهیم موجود در سوره‌ها با یکدیگر سنجیده و با حالت تصادفی مقایسه گردید. نتایج به‌دست‌آمده نشان می‌دهد که سوره‌های قرآن انسجام معنایی دارند و مفاهیم و مطالب درون هر سوره حول یک موضوع خاص شکل گرفته‌اند. همچنین، بررسی شباهت فصل اول و آخر و نیز فصل اول و فصول بعدی هر سوره نشان می‌دهد که نظریه احکام و تفصیل برای بسیاری از سوره‌ها صادق است. در انتها، با بررسی شباهت سوره‌ها بر اساس فاصله ترتیبی آن‌ها در قرآن و نیز فاصله زمانی نزولشان در طول تاریخ، مشخص شد که سوره‌هایی که از نظر چینی به یکدیگر نزدیک‌ترند، شباهت بیشتری نیز با یکدیگر دارند. بنابراین، به نظر می‌رسد که کل قرآن دارای نظم چینی است و سوره‌های آن در خوشه‌های مرتبط با یکدیگر مرتب شده‌اند.

استناد: خدنگی، احسان؛ و حقی طارونی، حسین (۱۴۰۴). روش‌شناسی حل مسأله فقهی. علوم انسانی و اسلامی دیجیتال،

۱(۱)، ۱۲۳-۱۵۴. doi.org/10.22034/disah.2024.716139



ناشر: پژوهشکده علوم اسلامی و انسانی دیجیتال (مرکز تحقیقات کامپیوتری علوم اسلامی نور). © نویسندگان.

مقدمه

نگرش ساختاری به سوره‌های قرآن، تلاش در جهت یافتن عنوان و موضوع محوری سوره‌ها، ساماندهی آیات حول یک هدف واحد و تفسیر مبتنی بر ساختار سوره‌ها، مورد توجه جدی برخی قرآن‌پژوهان متأخر قرار گرفته است (خامه‌گر، ۱۳۸۵: ۲۷۸/۴۹-۲۹۱). پایه و فرض اصلی تحلیل ساختار سوره‌های قرآن توسط این محققان، توقیفی و وحیانی بودن چینش آیات در سوره‌های قرآن است (خامه‌گر، ۱۳۸۶: ۱۸۲/۸۶-۲۱۳). از دید این پژوهشگران، قرآن یک سیستم به هم پیوسته و منظم است که برای فهم آن باید ارتباطات بین اجزاء سیستم اعم از مفاهیم، آیات و سیاق‌های درون سوره‌ها را شناسایی کنیم. بر این اساس و بر پایه بررسی‌های مفصلی که توسط پژوهشگران متأخر علوم قرآنی مانند خامه‌گر و لسانی فشارکی انجام شده، قرآن یک سیستم یکپارچه و دقیق است که اجزای آن (سوره‌ها و سیاق‌ها) نیز دارای نظم دقیق و وحیانی است (خامه‌گر، ۲۰۰۲: ۱۳۸-۹/۱۳). تحلیل این سیستم واحد می‌تواند به کشف افق‌های جدیدی از اعجاز قرآن، استخراج دانش غنی قرآنی، یافتن غرض و محور اصلی سوره‌ها، نظام چینش سوره‌ها و... منتهی شود. پژوهشگران بسیاری مبتنی بر این فرضیات به بررسی ساختار سوره‌های مختلف پرداخته‌اند، از جمله بررسی ساختار سوره مائده توسط آرام و لایقی (۱۳۹۵: ۵۵/۱۳-۷۷)، ساختار سوره کهف توسط فتاحی‌زاده و ذاکری (۱۳۹۵: ۱۰۱/۲۵-۱۲۰)، ساختار سوره انشقاق توسط دهقانی‌فارسانی (۱۳۸۷: ۳/۱۴-۱۴) و ساختار سوره انفطار توسط جیگاره و صادقی (۱۳۹۵: ۵۱/۱۳-۷۴) اشاره کرد. همچنین افرادی چون خامه‌گر به ارائه ترجمه قرآن مبتنی بر ساختارمندی سوره‌ها پرداخته‌اند (خامه‌گر، ۱۳۸۵: ۲۷۸-۲۹۱/۴۹).

با وجود نشانه‌های مختلف کشف شده توسط قائلین به ساختارمندی سوره‌ها، برخی مستشرقان با مشاهده نشانه‌هایی چون بیگانگی سبک بیان قرآن با معیارهای خاص و با فرض دخالت صحابه در تعیین ترتیب آیات و حتی غیر وحیانی بودن قرآن به این نتیجه رسیده‌اند که محتوای قرآن از هم گسیخته و بدون ارتباط منطقی است. به‌طوریکه ریچارد بل، یکی از قرآن‌پژوهان اروپایی، در مقدمه ترجمه معتبری که از قرآن به انگلیسی ارائه کرده می‌نویسد: «یکی از ویژگی‌های اصیل سبک قرآن این است که جسته‌جسته است و به ندرت می‌توان در طول بخش عمده‌ای از یک سوره انسجام

معنایی دید» (R. Bell, 1953). «آرتور جان آربری» نیز در بخشی از مقدمه خود بر ترجمه قرآن می‌نویسد: «قرآن از هر انسجامی که مربوط به ترتیب نزول باشد و نیز از انسجام منطقی بسی دور است. خواننده قرآن، به‌ویژه اگر ناچار باشد به یک ترجمه اکتفا کند، هرچند که آن ترجمه از نظر زبان‌شناختی دقیق باشد، بی‌شک از حالت ظاهراً بی‌نظم بسیاری از سوره‌ها حیرت‌زده می‌شود» (A. J. Arberry, 1996). تأکیدهای این پژوهشگران بر گسیختگی و بی‌نظمی آیات قرآن، به خواننده این نکته را گوشزد می‌کند که قرآن از دستبردهای بشری در امان نمانده است و حداقل چینش آیات و حیانی نیست. این در حالی است که دلایل متعدد عقلی مبتنی بر مطالعه ساختار سوره‌ها و مستندات تاریخی تصریح می‌کنند که ترتیب آیات و حیانی بوده و در طول زمان ثابت و بدون تغییر باقی مانده است (خامه‌گر، ۱۳۸۶: ۱۸۲/۸۶-۲۱۳).

اگرچه در مورد سازماندهی سوره‌ها کارهای مختلفی انجام شده و ساختار برخی از سوره‌ها توسط پژوهشگران علوم قرآنی مورد بررسی قرار گرفته است، اما به نظر می‌رسد در هیچ‌یک از این آثار از الگوریتم‌های متن‌کاوی و پردازش زبان‌های طبیعی استفاده نشده است.

مقاله حاضر به بررسی یکپارچه سیستم قرآن و مطالعه ساختارمندی سوره‌ها از منظر درون‌سوره‌ای و میان‌سوره‌ای با استفاده از متدولوژی‌ها و تکنیک‌های متن‌کاوی می‌پردازد. ساختارمندی درون‌سوره‌ای از منظر وحدت موضوعی و ساختارمندی برون‌سوره‌ای از منظر چینش سوره‌های قرآن بررسی خواهد شد. بدین منظور، پژوهش حاضر دارای دو سؤال اصلی است: ۱. آیا سوره‌های قرآن پیرامون یک موضوع واحد هستند؟ ۲. آیا ترتیب سوره‌های قرآنی سامان‌یافته است؟ باقی‌مانده مقاله به شرح زیر تنظیم شده است: بخش اول برخی از کارهای مرتبط را ارائه می‌کند. تعاریف اساسی به‌طور خلاصه در بخش دوم توضیح داده شده است. بخش سوم شامل روش‌شناسی پژوهش است. بخش چهارم به پیش‌پردازش می‌پردازد. در بخش پنجم روش و معیار ارزیابی خود را با جزئیات بیشتری توضیح می‌دهیم. سپس نتایج آزمایش‌های خود را در بخش ششم ارائه می‌کنیم.

الف. کارهای مرتبط

علاوه بر آثار پژوهشگران علوم قرآنی و خاورشناسان که پیش‌تر ذکر شد، پژوهشگران علوم رایانه نیز کارهای زیادی در زمینه تحلیل قرآن انجام داده‌اند. با توجه به اهمیت جستجوی معنایی در قرآن،

بسیاری از آثار به دنبال روش‌های جدید جستجوی معنایی بوده‌اند. از جمله این آثار، (A. R. Yauri و همکاران، ۲۰۱۳: ۴۴۹۸-۴۴۹۲/۶؛ H. U. Khan و همکاران، ۲۰۱۳: ۵۷۰/۲). و همچنین (M. Shoaib و همکاران، ۲۰۰۹: ۳۴-۲۹؛ M. Alhwarat و همکاران: ۲۰۱۵: ۲۸۸/۶-۲۹۴) هر کدام به ترتیب روش‌هایی مبتنی بر هستی‌شناسی، شبکه کلمات و مدل‌سازی موضوعی برای جستجوی معنایی قرآنی ارائه کرده‌اند.

آثار مختلفی نیز به ساخت هستی‌شناسی‌های قرآنی پرداخته‌اند که بیشتر آنها بر حوزه‌ای خاص متمرکز شده‌اند (R. Ismail، ۲۰۱۵). اقبال و همکاران نقاط ضعف هستی‌شناسی‌های موجود قرآن را برجسته کرده و هستی‌شناسی جدیدی را ارائه کرده‌اند (R. Iqbal، ۲۰۱۳: ۳۴۳/۷-۳۳۳). محمدصفی و همکاران روش‌های مختلف بازیابی آیات را نیز مورد بررسی قرار داده و نقاط ضعف و قوت آنها را مطرح کرده‌اند (S. A. Pitchay، ۲۰۱۶: ۶۳۴/۱۰۰-۶۲۹). یافته‌های آنان نشان می‌دهد که برای رفع تناقض میان هستی‌شناسی‌های موجود، نیاز به یادگیری و ساخت هستی‌شناسی‌های قرآنی جدید وجود دارد.

مجموعه‌ای از آثار دیگر به ارائه پیکره‌های مناسب برای تحلیل قرآن پرداخته‌اند. از این میان، (K. Dukes، ۲۰۱۰: ۷-۱؛ E. Atwell، ۲۰۰۹: ۱۶۹) از درخت-بانک آیات قرآن با توجه به دستور زبان عربی بهره‌برداری کرده و آن را با نمودارهای وابستگی نشان داده‌اند. «شرف و همکاران» در (A. B. M. Sharaf and E. Atwell، ۲۰۱۲: ۲۳۰۲-۲۲۹۵) پیکره‌ای را ارائه کرده‌اند که آیات مشابه از نظر مفهومی را به هم متصل می‌کند. این پیکره «قرسیم» نامگذاری شده و می‌تواند برای کاربردهای مختلفی مانند ترجمه قرآن استفاده شود. پیکره «قرآنا» نیز ضمائر شخصی قرآن را بر اساس ارجاعات آنها برچسب‌گذاری کرده است (A. B. M. Sharaf and E. Atwell، ۲۰۱۲: ۱۳۷-۱۳۰). «شریف و همکاران» مجموعه داده‌ای بر اساس RDF از ترجمه قرآن به ۴۳ زبان مختلف استخراج کرده‌اند که می‌تواند برای کاربردهای مختلف در پردازش زبان طبیعی استفاده شود (M. A. Sherif، ۲۰۱۵: ۳۴۵/۶-۳۳۹). علاوه بر مجموعه داده‌های ارائه شده، ابزارهایی برای جستجو و تحلیل پیکره‌های قرآن معرفی شده است. «الفایفی و همکاران» این ابزارها را در (Alfaifi and E، ۲۰۱۶: ۳۴۷-۳۵۷/۱۹) بررسی و مقایسه کرده‌اند.

رویکرد برخی از پژوهش‌ها نیز تحلیل قرآن برای کاربردهای مختلف از جمله توسعه سیستم پاسخگویی به سؤالات قرآنی (S. K. Hamed and M. J. Ab Aziz, ۲۰۱۶: ۱۲/۱۷۷-۱۶۹) و طبقه‌بندی آیات است. به‌عنوان مثال، «ابوالبقایی و همکاران» به دنبال طبقه‌بندی سوره‌های قرآنی به دو دسته مکی و مدنی با استفاده از طبقه‌بند درختی تصمیم بودند. آنها از ویژگی‌هایی همچون طول سوره، کلمات و عبارات به کار رفته در سوره‌ها و آیات سجده‌دار استفاده می‌کردند.

ب. تعاریف اولیه

۱. وحدت موضوعی سوره‌ها

برخی محققان علوم قرآنی بر ساختارمندی سوره‌های قرآن تأکید دارند و با بررسی نظم در سوره‌های مختلف نظریه‌هایی ارائه کرده‌اند که مهم‌ترین آن‌ها نظریه وحدت موضوعی است. بر اساس این نظریه، هر سوره دارای یک موضوع محوری است و تمامی آیات و مباحث مطرح‌شده در سوره به آن موضوع مرتبط می‌شوند. پژوهشگرانی که به وحدت موضوعی سوره‌ها قائلند، مطرح می‌کنند که خداوند در آیات ابتدایی، موضوع محوری سوره را بیان می‌کند. سپس در بخش‌های مختلف سوره، این موضوع را از طریق مثال‌ها، داستان‌ها و مصادیق مختلف باز می‌نماید و در نهایت، مبتنی بر همان موضوع محوری جمع‌بندی می‌کند (P. Soucy, ۲۰۰۵: ۱۱۳۵-۱۱۳۰؛ Y. Zhang, ۲۰۱۰: ۱/۵۲-۴۳).

۲. نظریه احکام و تفصیل

نظریه «احکام و تفصیل» بیان می‌کند که خداوند متعال موضوع اصلی را در آیات ابتدایی مطرح می‌کند، سپس در بخش‌های مختلف سوره، آن را با قیاس‌ها، حکایات و مثال‌ها تفصیل می‌دهد و در نهایت بر اساس موضوع اصلی نتیجه‌گیری می‌کند (محمد خامه‌گر، ۱۳۸۳: ۱۷۹/۱۷-۱۳؛ ۱۳۸۱: ۱۷/۲۹-۱۳). بر اساس این نظریه، موضوع اصلی هر سوره در قسمت ابتدایی سوره که مقدمه نامیده می‌شود، معرفی شده است. به همین ترتیب، بخش پایانی سوره که به نوعی نتیجه‌گیری از مباحث مطرح‌شده در سوره است، به‌عنوان فصل نتیجه‌گیری شناخته می‌شود.

۳. آیه محوری

موضوع محوری سوره که کل سوره حول آن می‌چرخد، عموماً در یک آیه از همان سوره مطرح می‌شود که در این مقاله به آن «آیه محوری سوره» اطلاق می‌شود.

۴. بخش

در این مقاله، آیاتی از قرآن که در کنار هم درباره موضوع خاصی صحبت می‌کنند، «بخش» نامیده می‌شوند.

۵. مقدمه و نتیجه‌گیری سوره

بر اساس نظریه وحدت موضوعی، موضوع محوری یک سوره در سیاق ابتدایی هر سوره عنوان می‌شود که به آن مقدمه می‌گوییم. همچنین به فصل انتهایی سوره که به نوعی جمع‌بندی مباحث مطرح‌شده در سوره است، «نتیجه‌گیری سوره» اطلاق می‌شود.

۶. مدل‌های فضای برداری

روش‌های مختلفی برای نمایش متون و مفاهیم به‌صورت برداری وجود دارد. در مدل‌های فضای برداری، متن یا کلمه به‌صورت بردار نمایش داده می‌شود که هر جزء آن با اهمیت تخمینی کلمه در متن مرتبط است (P. Soucy, ۲۰۰۵: ۱۱۳۵-۱۱۳۰). روش‌های «کیسه کلمات» و «N-gram» یکی از کاربردی‌ترین روش‌ها برای نمایش متون هستند که با وجود سادگی، برای بسیاری از کاربردهای متن‌کاوی مناسب عمل می‌کنند (Y. Zhang, ۲۰۱۰: ۵۲/۱-۴۳).

ج. مواد و روش‌ها

در این مقاله، ساختارمندی سوره‌های قرآنی بر اساس نظریه وحدت موضوعی بررسی شده است. روش تحقیق حاضر شامل هفت بخش است:

۱. پیش‌پردازش و تهیه داده‌ها؛

۲. سیاق‌بندی سوره‌ها؛

۳. محاسبه تشابه ریشه‌های قرآنی؛

۴. محاسبه تشابه سیاق‌ها و سوره‌های قرآنی؛

۵. بررسی رابطه عنوان سوره‌ها با محتوای آن‌ها؛

۶. مطالعه وحدت موضوعی سوره‌ها؛

۷. ساختارمندی قرآن از منظر چینش سوره‌ها؛

بر اساس این روش، ابتدا دادگان ورودی برای پردازش‌های بعدی آماده‌سازی شد. سپس با استفاده از روش‌های مختلف بردارسازی متون، ریشه‌های قرآنی به فضای برداری منتقل شدند و شباهت ریشه‌ها از طریق محاسبه شباهت بردارهای متناظر آن‌ها به دست آمد. در ادامه، وحدت موضوعی سوره‌ها به دو شیوه مختلف مورد مطالعه قرار گرفت. ابتدا شباهت مفاهیم درون سوره‌ای محاسبه شد و با حالت تصادفی مقایسه گردید. سپس شباهت فصل ابتدایی سوره‌ها با سایر فصول و همچنین فصل ابتدایی با نتیجه‌گیری سوره‌ها برای سوره‌های مختلف محاسبه شد و نتایج با حالت تصادفی مقایسه شد. در نهایت، نظام‌مند بودن قرآن از منظر چینش سوره‌ها مطالعه گردید. به این صورت که شباهت سوره‌های مختلف محاسبه و ارتباط فاصله چینشی سوره‌ها در قرآن و فاصله زمانی نزول سوره‌ها با میزان شباهت سوره‌ها بررسی شد.

۱. مجموعه داده

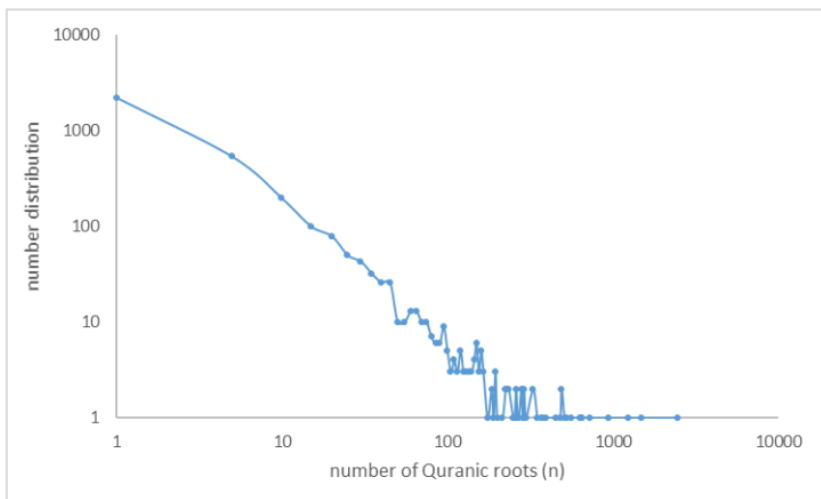
با توجه به تعداد زیاد واژگان متمایز قرآن، به جای استفاده از مشتقات کلمات قرآنی، تمرکز اصلی بر روی ریشه‌های قرآنی قرار گرفت. داده اولیه به صورت جدولی بود که در هر سطر آن، کلمات قرآنی به طور کامل اعراب‌گذاری شده، سوره‌ای که کلمه در آن قرار گرفته، آیه مربوط به کلمه و در نهایت ریشه متناظر با کلمه مدنظر ذخیره شده بود. در فاز آماده‌سازی داده‌ها، ابتدا حروف ربط و حروف اضافه از مجموعه داده حذف و اعراب کلمات نیز کنار گذاشته شد. سپس مجموعه داده از نظر کدینگ یکسان‌سازی گردید، کلمات با ریشه‌های قرآنی جایگزین شده و ریشه‌ها شماره‌گذاری شدند. علاوه بر این، یک دیتاست دیگر که در آن ریشه‌های مربوط به هر آیه به تفکیک آیات ذخیره شده بودند ایجاد شد. همچنین دیتاست‌های تعداد تکرار ریشه‌ها و شماره سوره‌ها نیز تولید گردید.

۲. پیش‌فرض‌های پژوهش

پیش‌فرض اصلی پژوهش حاضر این است که اگر فاصله دو کلمه در یک متن از یکدیگر کم باشد، احتمال شباهت یا ارتباط این دو کلمه به یکدیگر زیاد است.

د. نمونه داده

شکل (۱) توزیع فراوانی ریشه‌ها را در کل قرآن نشان می‌دهد.



شکل (۱): توزیع فراوانی تعداد ریشه‌های قرآنی در کل قرآن

[محور عمودی: عدد توزیع؛ محور افقی: تعداد ریشه‌های قرآنی]

از آنجا که توزیع تعداد ریشه‌ها در مقیاس لگاریتمی تقریباً به صورت خطی است، می‌توان گفت که تعداد ریشه‌های قرآنی در کل قرآن از توزیع قانون توانی^۱ پیروی می‌کند. این امکان وجود دارد که توزیع فراوانی با توزیع‌های لگاریتم نرمال (M. Mitzenmacher, ۲۰۰۴: ۲۵۱/۱-۲۲۶) یا توزیع توانی با برش نمایی (A. Clauset, ۲۰۰۹: ۷۰۳/۵۱-۶۶۱) تطابق داشته باشد، اما نمودار نشان داده‌شده در شکل (۱) از هر دو توزیع ذکر شده متمایز است. علت تمایز این توزیع با توزیع لگاریتم نرمال، تعداد زیاد ریشه‌هایی است که فراوانی بسیار کمی دارند. علاوه بر این، دُم توزیع به اندازه‌ای بلند است که نیازی به تطبیق آن با قانون توانی برش دار نیست.

1. power-law.

ه. روش پیشنهادی

پس از تهیه داده‌ها، سوره‌های مختلف تقسیم‌بندی شدند. در این مقاله، از سیاق‌بندی پیشنهادی علامه طباطبایی در کتاب المیزان با اصلاحاتی استفاده شده است (M. H. Tabatabai, ۱۹۹۶: ۲). بر اساس روش‌های پیشنهادی زیر، شباهت‌های ریشه‌ها، بخش‌ها و سوره‌های قرآنی محاسبه می‌شوند.

۱. محاسبه تشابه ریشه‌های قرآنی

روش tf-idf از ترکیب فراوانی عبارت در سند و فراوانی معکوس عبارت در اسناد دیگر به دست می‌آید. فراوانی عبارت t در سند d که با tft, d نشان داده می‌شود، وزنی است که به هر عبارت بر اساس تعداد وقوع آن در سند اختصاص داده می‌شود. فراوانی معکوس سند نیز از فرمول زیر محاسبه می‌شود، که در آن N تعداد کل اسناد مجموعه و dft برابر با تعداد اسنادی است که عبارت t در آن‌ها وجود دارد:

$$idf_t = \log \frac{N}{dft}$$

بر این اساس tf-idf، به عبارت t در سند d وزنی متناسب با فرمول زیر انتساب می‌دهد (R. Larson, ۲۰۱۰: ۸۵۲/۶۱).

$$tf_idf_{t,a} = tft,a \times idf_t$$

برای محاسبه میزان شباهت دو ریشه بر اساس tf-idf، کافی است بردارهای tf-idf ریشه‌ها را بر اساس وزن‌های به دست آمده در سوره‌های مختلف محاسبه کنیم. شباهت کسینوسی بین بردارهای tf-idf ریشه‌ها می‌تواند معیار مناسبی برای اندازه‌گیری شباهت ریشه‌ها باشد. روش دیگری که در این مقاله مورد استفاده قرار گرفته است، word2vec می‌باشد. این روش که توسط میکولوو و همکارانش در گوگل ارائه شد، هر کلمه را به یک بردار پیوسته نگاشت می‌کند [۳۲]. زمانی که هدف بازنمایی عناصر بزرگ‌تر است، نسخه تعمیم‌یافته word2vec با نام Sent2vec قابل استفاده می‌باشد. [33]

روش دیگری که در این مقاله استفاده شده، باهم‌آیی ریشه‌ها است. این روش بر اساس کنار هم قرار گرفتن ریشه‌ها در آیات تمرکز دارد و فرض بر این است که اگر دو ریشه در یک آیه کنار یکدیگر

قرار گیرند، ارتباطی بین آن‌ها وجود دارد و هر چه تعداد کنار هم قرار گرفتن دو ریشه بیشتر باشد، میزان ارتباط آن‌ها بیشتر خواهد شد. مبتنی بر این فرض، شباهت دو ریشه را می‌توان با استفاده از فرمول زیر محاسبه کرد:

$$S_{ij} = \frac{N_{ij}}{\sqrt{N_i N_j}}$$

که در آن N_{ij} تعداد باهم‌آیی ریشه‌های i و j است. مخرج کسر نیز میانگین هندسی تعداد تکرار ریشه‌های i و j در قرآن است. سپس ماتریس حاصل مجدداً به نحوی نرمال‌سازی می‌شود که مجموع درایه‌های ماتریس شباهت برای هر سطر برابر با ۱ باشد:

$$W_{ij} = \frac{S_{ij}}{C_i}$$

$$\sum_i W_{ij} = 1$$

که در آن C_i مجموع عناصر هر سطر از ماتریس است. این نرمال‌سازی به این صورت است که مجموع مقدار شباهت هر ریشه با ریشه‌های قرآنی دیگر برابر با ۱ باشد.

۲. محاسبه تشابه بخش‌ها و سوره‌ها

در این مقاله، تشابه دو فصل یا دو سوره i و j با میانگین تشابه تمام ریشه‌هایی که در بخش i هستند با تمام ریشه‌های بخش j به صورت دو به دو تعریف می‌شود. بنابراین، شباهت دو بخش را می‌توان با استفاده از معادلات زیر محاسبه کرد:

$$sim_{i,j} = \frac{\sum_{m \in i, n \in j} M_{sim}[m, n]}{l_i l_j}$$

در این فرمول، i و j به ترتیب سوره یا قسمت i و j ، l_i و l_j تعداد ریشه‌های دو بخش و M_{sim} ماتریس تشابه ریشه‌ها را نشان می‌دهد که طبق روش‌های ذکر شده در بالا به دست می‌آید. راه‌حل دیگر برای اندازه‌گیری شباهت دو بخش، استفاده از تعداد ریشه‌های مشترک و شباهت کسینوسی تعمیم‌یافته آن است. در این روش، برای هر سوره یک بردار به اندازه تعداد ریشه‌های قرآنی ایجاد می‌شود. عناصر این بردار با تعداد ریشه‌های موجود در سوره، مقاداردهی اولیه می‌شوند. بر این

اساس، می‌توان از شباهت کسینوسی دو بردار به‌عنوان معیاری برای شباهت دو سوره استفاده کرد. در برخی آزمایشات این مقاله، از نسخه ساده‌شده این روش، یعنی تعداد یا نسبت ریشه‌های مشترک، بهره‌برداری شده است.

۳. ارزیابی

برای ارزیابی روش‌های مختلف، نتایج شباهت‌ها (مانند عنوان و محتوای سوره، شباهت بین سوره‌های مختلف، شباهت بخش اول و آخر یک سوره و شباهت مفاهیم موجود در یک سوره) با حالت تصادفی مقایسه شده است. در این مقاله برای انتخاب قطعات تصادفی، از انتخاب ریشه‌های قرآنی بر اساس احتمال وقوع هر یک از ریشه‌ها در قرآن استفاده می‌شود.

بر این اساس، روش کار به‌صورت زیر خواهد بود:

۱. یک عدد بین ۱ تا ۱۱۴ به‌طور تصادفی و یکنواخت انتخاب کنید.

۲. سوره متناسب با عدد انتخاب شده را برگزینید.

۳. تعداد ریشه‌های این سوره را با تعداد ریشه‌های بخش مدنظر مقایسه کنید. اگر کمتر بود، به

گام اول برگردید.

۴. پس از انتخاب سوره مدنظر، برای انتخاب ریشه‌های شروع و پایان یک بخش تصادفی از

سوره، از فرمول زیر استفاده کنید:

$$\text{Start} = \text{rand_int}(0, \text{SureSize} - \text{PartSize})$$

$$\text{End} = \text{Start} + \text{PartSize}$$

در این فرمول‌ها، تابع `rand_int` یک عدد تصادفی در بازه مشخص شده انتخاب می‌کند.

`SureSize` تعداد کل ریشه‌های سوره انتخاب شده و `PartSize` تعداد ریشه‌های بخش مورد نظر است.

۵. بخش بین `Start` و `End` را بازگردانید.

روش دوم برای انتخاب ریشه‌های قرآنی مبتنی بر احتمال وقوع هر ریشه در قرآن به‌صورت زیر

عمل می‌کند:

۱. تعداد تکرار ریشه‌های مختلف در قرآن را محاسبه کنید.

۲. بر اساس تعداد تکرار ریشه‌ها در قرآن، یک ریشه را انتخاب کنید.

۳. این مراحل را به تعداد اندازه بخش مدنظر (`PartSize`) تکرار کنید.

۴. بررسی ساختار مندی سوره‌های قرآنی

در این مقاله، سه مجموعه آزمایش برای پاسخ به اولین سوال اصلی طراحی شده است. مجموعه اول به محاسبه شباهت عنوان سوره‌ها و محتوای درونی آنها می‌پردازد. مجموعه دوم در پی سنجش وحدت موضوعی در سوره‌های قرآنی و بررسی شباهت مفاهیم موجود در یک سوره است. سومین و مهم‌ترین مجموعه به ارزیابی میزان ارتباط بخش اول (به‌عنوان بخش مهم) با بخش‌های بعدی سوره‌ها اختصاص دارد.

برای این منظور، با استفاده از روش‌های همنشینی ریشه‌های قرآنی در آیات و الگوریتم word2vec، شباهت ریشه‌های قرآنی مختلف محاسبه و در ماتریسی به نام ماتریس تشابه ذخیره شد. سپس برای محاسبه شباهت مفاهیم درون سوره، چند سوره قرآنی انتخاب شدند که سوره‌هایی با اندازه‌های مختلف در میان آنها وجود داشت. همچنین، با توجه به شباهت میان بخش‌های سوره‌ها، سوره‌های کوچک و سوره‌های جزء سی‌ام قرآن از فهرست انتخاب‌ها خارج شدند. لازم به ذکر است که از آنجایی که حروف مقطعه قرآن تنها یک‌بار در قرآن تکرار شده است، سوره‌هایی مانند یاسین و قاف که نام آنها از حروف مقطعه قرآنی اقتباس شده است نیز انتخاب نشده‌اند. به‌طور خلاصه، شباهت‌های زیر برای سوره‌های انتخاب شده محاسبه گردید:

۱. شباهت عنوان سوره با ریشه‌های قرآنی درون سوره؛

۲. شباهت بین ریشه‌های قرآنی در سوره؛

۳. شباهت فصل اول و آخر سوره؛

۴. میانگین شباهت فصل اول با بخش‌های مختلف سوره.

باید توجه داشت که به دلیل فراوانی بسیار بالای برخی ریشه‌ها مانند "الله" و "رب"، این ریشه‌ها برای روش همنشینی (RA) حذف شدند تا شباهت بین برش‌ها به‌طور دقیق‌تری محاسبه شود. تشابه‌ها ابتدا با حذف ریشه‌هایی با فراوانی بالای ۸۰۰ و سپس با حذف ریشه‌هایی با فراوانی‌های بالای ۶۰۰، ۴۰۰ و ۲۰۰ محاسبه شدند. در نهایت، شباهت‌های محاسبه‌شده با حالت تصادفی و سوره‌ها مقایسه گردید و ساختار مندی قرآن از منظر یکسانی موضوع ارزیابی شد.

۵. بررسی نظریه احکام و تفصیل

هر یک از پژوهشگران علوم قرآنی بر اساس دیدگاه و پیشینه عقیدتی خود ساختار خاصی برای سوره‌ها فرض کرده‌اند و هر یک به دنبال نشانه‌هایی برای اثبات ادعای خود به شیوه‌ای خاص هستند [۵ و ۷]. خداوند متعال بر اساس رایج‌ترین نظریه ساختارمندی سوره‌ها، موضوع و ایده اصلی سوره را در بخش اول مطرح می‌کند، سپس در بخش‌های مختلف آن را توضیح می‌دهد و در نهایت نتیجه‌گیری را ارائه می‌کند. این ساختار در این مقاله «ساختار مقدمه و توضیح» نامیده شده است. در این بخش، قصد داریم ساختارمندی سوره‌ها را بر اساس نظریه «احکام و تفصیل» بررسی کنیم. برای بررسی این نظریه، ابتدا شباهت بخش اول (که شامل مبحث اصلی است) و بخش پایانی (که از مباحث مطرح شده در سوره به نتیجه می‌رسد) محاسبه می‌شود. سپس این شباهت با حالت تصادفی مقایسه می‌شود. در ادامه، شباهت بخش اول و بخش‌های مختلف سوره محاسبه شده و نشان می‌دهیم که بر اساس نظریه «احکام و تفصیل»، توضیح بخش اول در سوره‌ها به‌طور مؤثری در سایر بخش‌های سوره تبیین می‌شود.

۶. بررسی نظام ترتیب سوره‌های قرآن

برای پاسخ به سوال دوم در نظم ترتیب سوره‌ها، آزمایشی به شرح زیر طراحی شد: در ابتدا، تشابه سوره‌های قرآن به صورت دوه‌دو اندازه‌گیری و در ماتریس تشابه سوره‌ها ذخیره شد. برای سنجش تشابه سوره‌ها، از ماتریس‌های تشابه tf-idf و RA استفاده می‌شود. بر اساس ترتیب شمارش سوره‌ها، مکان هر سوره مشخص شد. بنابراین، سوره‌های فاتحه، بقره و ناس به ترتیب ۱، ۲ و ۱۱۴ در نظر گرفته شده است. بر این اساس، فاصله مکانی دو سوره به صورت زیر محاسبه شد.

$$PD_{S_1, S_2} = |P_{S_1} - P_{S_2}|$$

جایی که P_{S_1} جای سوره S_1 و P_{S_2} جای سوره S_2 است.

بر اساس ترتیب نزول سوره‌های مختلف، فاصله زمانی سوره‌های S_1 و S_2 به شرح زیر محاسبه شد.

$$TD_{S_1, S_2} = |T_{S_1} - T_{S_2}|$$

که T_{S_1} و T_{S_2} زمان نزول سوره‌های S_1 و S_2 هستند.

در این مرحله میانگین شباهت سوره‌ها با فاصله مکانی $0 < pd < 114$ به شرح زیر محاسبه شد.

$$AveDisSim(pd) = \frac{\sum_{PD_{S_1, S_2} = pd} Sim(s_1, s_2)}{114 - pd}$$

که $114 - pd$ تعداد سوره‌هایی است که فاصله آنها برابر با pd است.

به همین ترتیب میانگین شباهت سوره‌ها با فاصله زمانی $0 < td < 114$ به صورت زیر محاسبه شد.

$$AveDisSim(td) = \frac{\sum_{TD_{S_1, S_2} = td} Sim(s_1, s_2)}{114 - td}$$

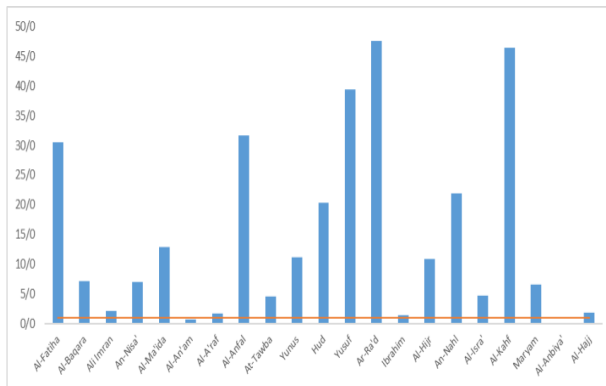
نمودار سوره‌ها در برابر فاصله مکانی رسم شد. همچنین، تشابه سوره‌ها در فاصله زمانی ترسیم و تغییر شباهت سوره‌ها بر اساس فاصله زمانی و مکانی را بررسی گردید. در نهایت، نتایج این بررسی‌ها اعلام شد.

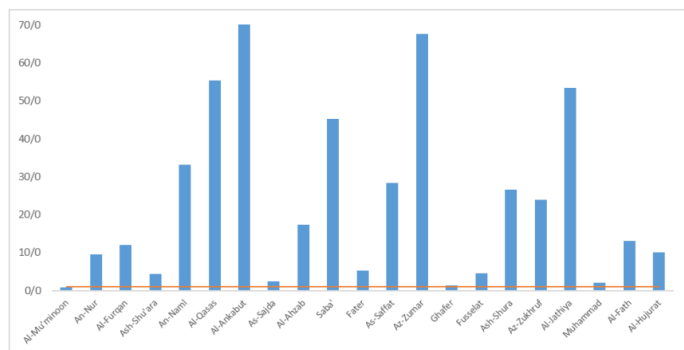
و. نتایج آزمایشات

در این بخش، نتایج آزمایشات انجام شده ارائه می‌شود. مواردی که پیش‌تر اشاره شده است، این مقاله به دنبال پاسخ به دو سوال اصلی است که عبارتند از: ۱. آیا سوره‌های قرآن حول یک موضوع واحد شکل گرفته‌اند؟ ۲. آیا چینش سوره‌ها در قرآن نظم دارد؟ در ادامه، نتایج آزمایشات مربوط به هر یک از این دو سؤال در دو بخش مجزا مورد بررسی قرار خواهد گرفت.

۱. ارتباط عنوان سوره با محتوای سوره

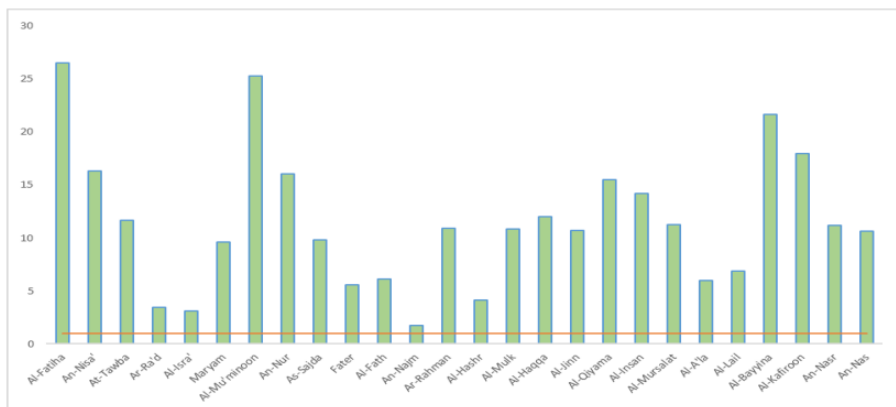
شکل (۳) فراوانی عنوان سوره‌ها را که در خود سوره تکرار شده، در مقابل حالت تصادفی نشان می‌دهد.





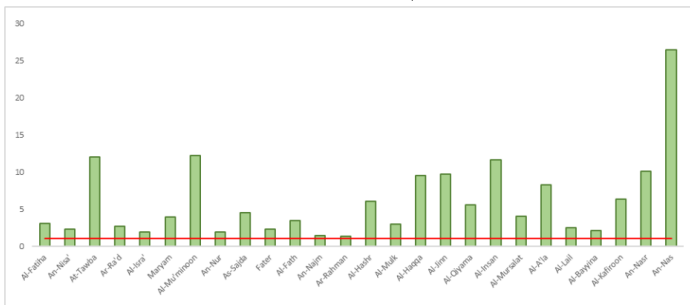
شکل (۳): فراوانی عنوان سوره‌ها در مقایسه با حالت تصادفی

همان‌طور که مشاهده می‌شود، در اکثر سوره‌ها، فراوانی مفاهیم در درون سوره‌ها بسیار بیشتر از حالت تصادفی (خط قرمز رنگ) است. با این حال، این حقیقت برای همه سوره‌ها صادق نیست و در برخی از سوره‌ها فراوانی کمتر است. به عنوان مثال، در حالی که بیشتر سوره انبیاء به پیامبران اختصاص دارد، ریشه مربوط به آن [ن-ب-ء] اصلاً تکرار نشده است. با این حال، در این سوره نام پیامبرانی از جمله ادریس، نوح، ابراهیم، اسماعیل، اسحاق، یعقوب، لوط، یونس، موسی، هارون، داوود، سلیمان، زکریا و یحیی ذکر شده است. بنابراین، می‌توان گفت که ممکن است عنوان سوره کم‌بسامد باشد، اما مفاهیم مشابه یا مرتبط با عنوان در سوره‌های دیگر بارها و بارها تکرار می‌شوند. برای بررسی این موضوع، شباهت عنوان سوره با مفهوم آن محاسبه شد. شکل (۴) این شباهت را با استفاده از روش RA در مقایسه با حالت تصادفی نشان می‌دهد.



شکل (۴): تشابه عنوان سوره‌ها با مفاهیم درون سوره‌ها بر اساس تشابه RA.

در شکل (۴)، بر خلاف شکل (۳)، شباهت و ارتباط بین عنوان سوره‌ها و مفاهیم درونی آن‌ها برای همه سوره‌ها بسیار بیشتر از حالت تصادفی است. به‌عنوان مثال، این شباهت برای سوره انبیاء ۷ برابر حالت تصادفی است. اگر از الگوریتم word2vec (کلمه به بردار) برای سنجش شباهت استفاده کنیم، شباهت عنوان سوره و مفاهیم موجود در آن مشابه شکل (۵) خواهد بود.

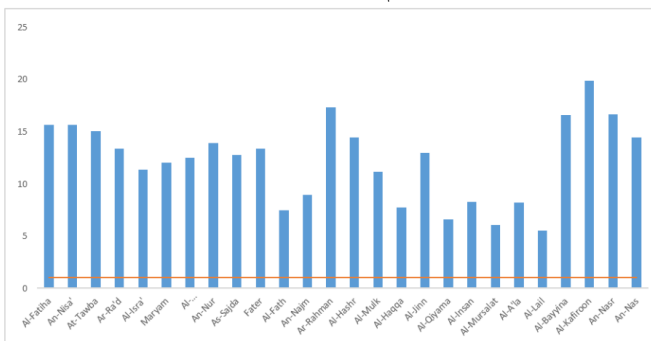


شکل (۵): شباهت عنوان سوره‌ها و مفاهیم موجود در آن بر اساس الگوریتم word2vec.

با توجه به شکل‌های ۴ و ۵، می‌توان گفت که عنوان سوره تقریباً برای همه سوره‌ها مشابه و با مفاهیم درونی سوره مرتبط بوده است. بنابراین، انتخاب عنوان سوره منطقی به نظر می‌رسد و نمی‌توان آن را صرفاً بر اساس انتخاب عمومی مردم دانست. با این حال، شباهت به‌دست آمده توسط الگوریتم word2vec کمتر است که به نظر می‌رسد به دلیل کوچک بودن مجموعه داده‌های آموزشی (قرآن) باشد.

۲. نتایج تجربی در مورد نظریه وحدت موضوعی

پس از بررسی عناوین سوره‌ها، وحدت موضوعی یا به عبارتی ساختارمند بودن مفاهیم درونی سوره‌ها مورد بررسی قرار گرفت. شکل ۶، نسبت شباهت مفاهیم درون سوره‌ای را در مقایسه با حالت تصادفی نشان می‌دهد.

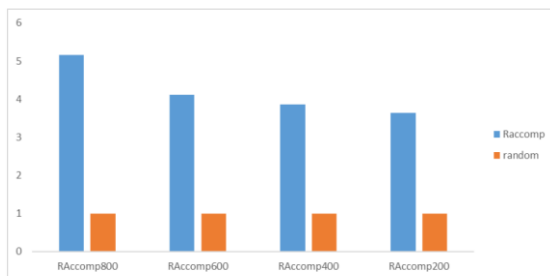


شکل (۶): میزان تشابه مفاهیم درون سوره‌ای در مقابل حالت تصادفی

همان‌طور که مشاهده شد، شباهت مفاهیم درون سوره‌ای برای همه سوره‌های بررسی شده بسیار بیشتر از حالت تصادفی است. به‌طور متوسط، شباهت این مفاهیم بیش از ۱۲ برابر حالت تصادفی است. این نشان می‌دهد که کلمات درون سوره‌ای در همه سوره‌های مورد بررسی با یکدیگر منسجم هستند. این مشاهدات نشان می‌دهد که هر سوره حول یک موضوع صراحتاً واحد یا چندین موضوع مرتبط با یکدیگر شکل گرفته است، اگرچه به‌طور صریح از یک موضوع اصلی پشتیبانی نمی‌کند. لازم به ذکر است که اگرچه نتایج ارائه شده در این مقاله مربوط به ۲۶ سوره، یعنی ربع قرآن است، اما به دو دلیل می‌تواند برای کل قرآن، به جز برخی از سوره‌های خاص، صادق باشد. ابتدا، سوره‌ها به گونه‌ای انتخاب شدند که سوره‌هایی با اندازه‌های مختلف مورد مطالعه قرار گیرند تا اگر اندازه سوره در نتیجه محاسبات تأثیر داشت، تشخیص داده شود. ثانیاً، در این مقاله بیش از ۲۶ سوره مورد مطالعه قرار گرفت که در مورد برخی از سوره‌های دیگر نیز همین نتایج صادق بود، اما به دلیل کمبود جا در اینجا گنجانده نشدند. اما باید توجه داشت که سوره‌هایی که نام آنها از حروف اختصاری قرآنی گرفته شده است یا سوره‌هایی که عنوان آنها در قرآن کم‌بسامد است، از این نتیجه مستثنی هستند. این صرفاً به این دلیل است که اطلاعات کافی در مورد عناوین آنها وجود ندارد، بنابراین نمی‌توان شباهت عناوین این سوره‌ها به ریشه‌های دیگر را با روش‌های موجود NLP به درستی محاسبه کرد.

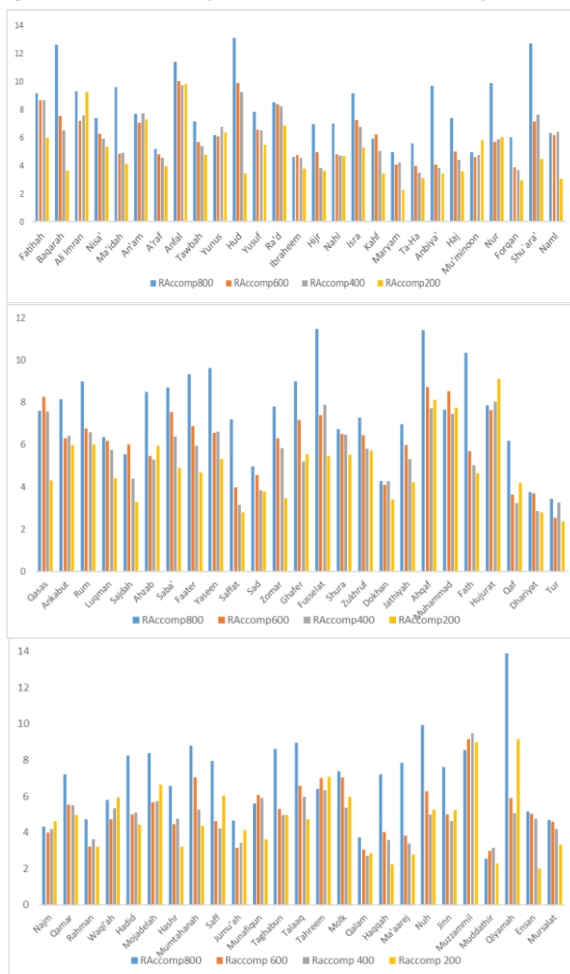
۳. نتایج تجربی در نظریه احکام و تفصیل

شکل (۷) نسبت میانگین تشابه بخش اول و آخر سوره‌ها (مقدمه و نتیجه‌گیری) را در مقایسه با حالت تصادفی نشان می‌دهد. برای این نمودار از روش RA استفاده شده است. به‌عنوان مثال، RAcomp800 محاسبه شباهت را بر اساس باهم‌آیی ریشه‌ها و بدون در نظر گرفتن ریشه‌هایی با بسامد بالاتر از ۸۰۰ نشان می‌دهد و RAcomp200 محاسبه شباهت را بر اساس باهم‌آیی ریشه‌ها و بدون در نظر گرفتن ریشه‌هایی با بسامد بالاتر از ۲۰۰ نشان می‌دهد.



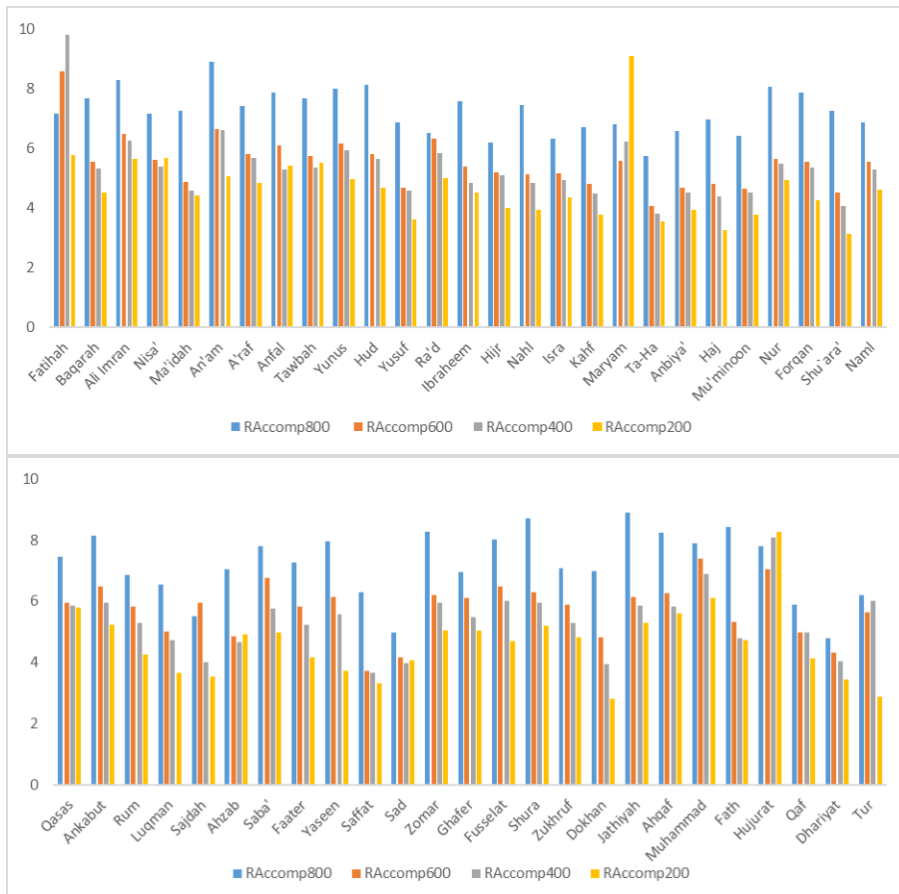
شکل (۷): مقایسه میانگین تشابه فصل اول و آخر سوره‌ها با حالت تصادفی

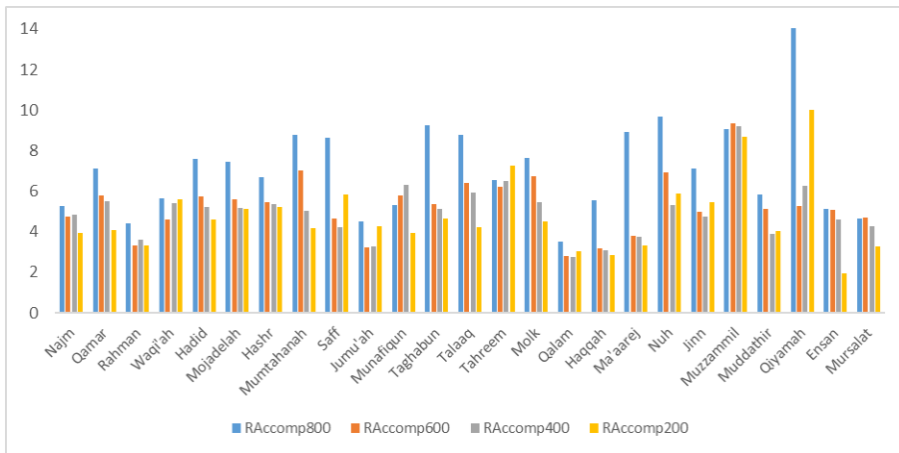
همان‌طور که در شکل (۷) مشاهده می‌شود، میانگین تشابه بخش اول و آخر سوره‌ها بسیار بیشتر از حالت تصادفی است. میانگین شباهت بر اساس روش RA بیش از ۴ برابر است. اگرچه میانگین تشابه بیشتر از حالت تصادفی است، اما مقدار آن به اندازه‌ای قابل توجه نیست که بتوان نتیجه گرفت که ساختار همه سوره‌های قرآن با نظریه احکام و تفصیل مطابقت دارد. به نظر می‌رسد این موضوع به دلیل معدل‌گیری در همه سوره‌ها باشد و از آن جایی که ممکن است برخی از سوره‌ها از مقدمه و توضیح پیروی نکنند، نتیجه نهایی کمتر از پیش‌بینی است. بنابراین تشابه بخش‌ها برای سوره‌های مختلف باید جداگانه بررسی شود. شکل (۸) شباهت بخش اول و آخر هر سوره را بر اساس روش RA نشان می‌دهد.



شکل (۸): نسبت شباهت قسمت اول و آخر سوره‌ها با حالت تصادفی

همان‌طور که در شکل مشاهده می‌شود، شباهت فصل اول و آخر در سوره‌های مختلف بسیار بیشتر از حالت تصادفی است. بدیهی است که شباهت سوره‌های مختلف با حالت تصادفی متفاوت است. در اینجا، به جز حالت RAcomp800، سوره انفال بیشترین شباهت را داشت. در حالت RAcomp800، سوره‌های قیامه و هود بیشترین شباهت را داشتند. علاوه بر این، سوره‌های انفال، مزمل، احقاف، و سایر سوره‌ها، با توجه به میانگین همه حالت‌ها، به ترتیب بیشترین شباهت را در فصل اول و آخر داشتند. پس از مطالعه شباهت میان بخش اول و آخر، به بررسی وجود مفاهیم بخش اول در کل سوره پرداختیم. در شکل ۹، میانگین شباهت بین بخش اول و تمامی بخش‌های سوره‌های فاتحه تا مرسلات نشان داده شده است.





شکل (۹): نسبت شباهت بخش اول و تمام بخش‌های دیگر سوره‌ها با حالت تصادفی

همان‌طور که در شکل مشاهده می‌شود، تقریباً در تمام سوره‌های قرآنی، شباهت فصل اول با سایر بخش‌ها بسیار بیشتر از حالت تصادفی است. در کمترین حالت که مربوط به سوره قلم است، شباهت بیش از سه برابر حالت تصادفی و در بیشترین حالت که مربوط به سوره مزمل است، این شباهت بالای ۱۰ برابر است. سوره‌های «قیامه»، «حجرات» و «محمد» نیز در رده‌های بعدی قرار دارند.

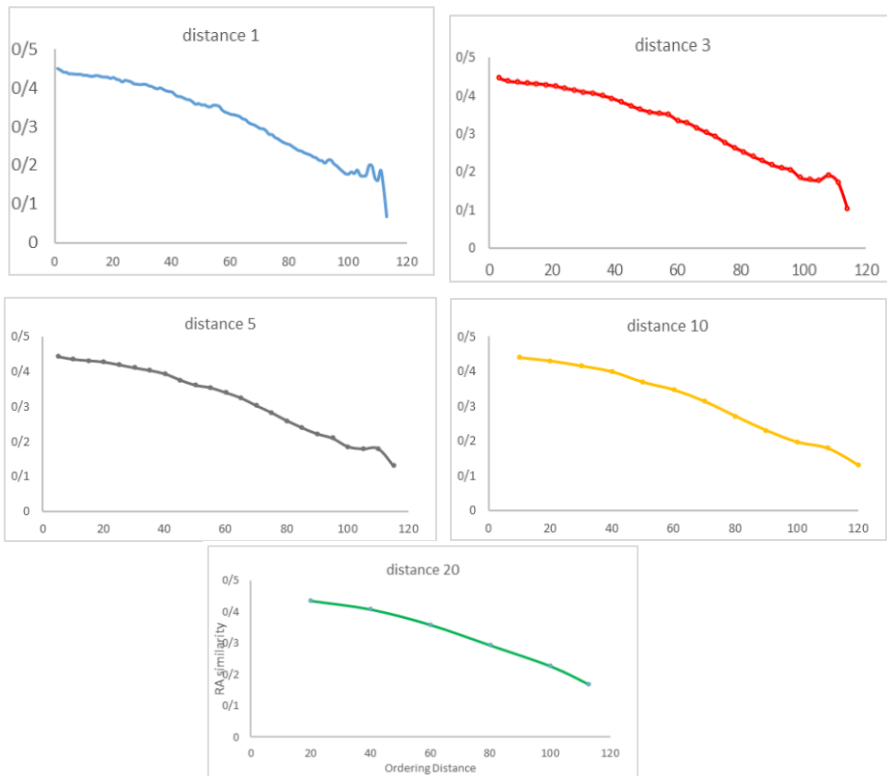
این موضوع نشان می‌دهد که مفاهیم بیان شده در مقدمه سوره‌ها تا حدودی در سرتاسر هر یک از سوره‌ها جاری است و توضیح داده می‌شود. اگرچه این امر در مورد سوره‌های مختلف کم و بیش صادق است، اما نظریه احکام و تفصیل را تقویت می‌کند. برای سوره‌هایی که شباهت کمتری نسبت به سایر سوره‌ها دارند، مطالعه بیشتری لازم است.

۴. نتایج تجربی به ترتیب سوره‌ها

برخی از محققان علوم قرآنی، یاران پیامبر را یکی از عوامل نظم سوره‌ها می‌دانند و بنابراین، ترتیب منطقی یا خاصی را برای سوره‌ها تأیید نمی‌کنند. برخی دیگر بر این باورند که ترتیب سوره‌های قرآنی منطقی یا گاهی وحیانی است. اما سازماندهی ترتیب سوره‌ها مسئله‌ای پیچیده است که برای بررسی آن باید روابط منطقی بین خوشه‌های مجاور سوره‌ها در قرآن با روش‌های مختلف مورد مطالعه قرار گیرد، امری که در این مقاله بررسی نشده است.

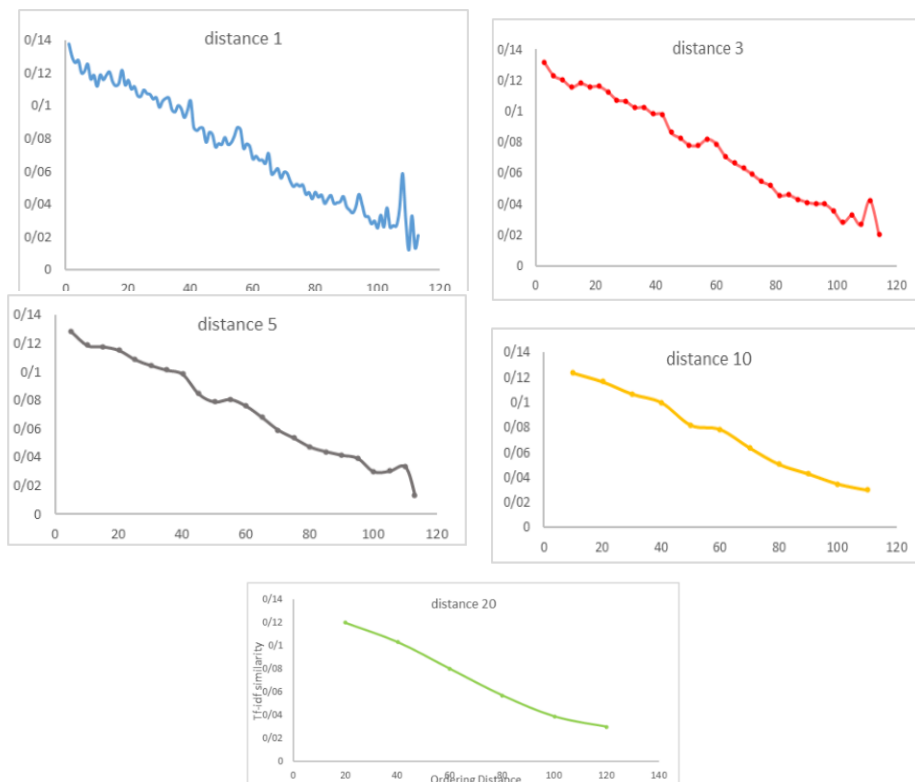
در این بخش به مقایسه شباهت سوره‌های نزدیک به هم طبق ترتیب مصحف با سوره‌های نزدیک از نظر ترتیب نزول، با قطعه‌بندی ۱، ۳، ۵، ۱۰ و ۲۰ می‌پردازیم. منظور از وضوح (I) اندازه پنجره‌ای است که در آن میانگین شباهت سوره‌هایی با فاصله کمتر یا مساوی (I) محاسبه می‌شود. به‌عنوان مثال، وضوح ۱ میانگین شباهت سوره‌های مجاور را محاسبه می‌کند و وضوح ۳ میانگین شباهت سوره‌هایی را که حداکثر فاصله آن‌ها ۳ است محاسبه می‌کند.

شکل (۱۰) شباهت سوره‌های قرآن را در برابر فاصله مکان آن‌ها از یکدیگر برای وضوح‌های مختلف نشان می‌دهد.



مطابق شکل (۱۰)، شباهت سوره‌های مختلف با افزایش فاصله تقریباً به‌صورت خطی کاهش می‌یابد، زیرا میانگین شباهت سوره‌های مجاور ۰.۴۵ و سوره‌هایی که بیشترین فاصله را از یکدیگر دارند ۰.۰۶ است. همین واقعیت در مورد قطعنامه‌های دیگر نیز صادق است.

در شکل (۱۱) نیز نمودار تشابه سوره‌ها بر اساس فاصله tf-idf در مقابل فاصله سوره‌ها ارائه شده است.

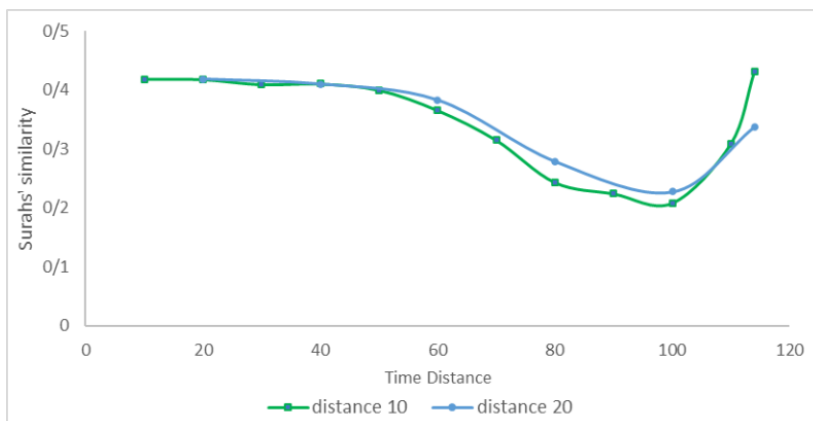
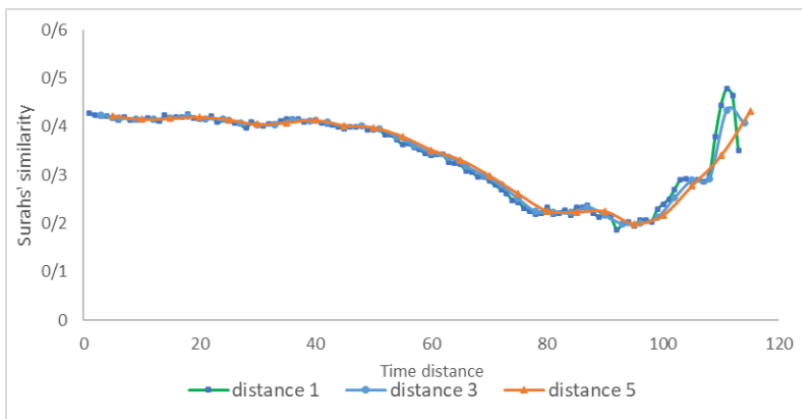


شکل (۱۱): تشابه سوره‌های مختلف در برابر فاصله آنها از یکدیگر بر اساس شباهت (محور عمودی: شباهت Tf-idf؛ محور افقی: فاصله ترتیبی)

بر اساس این شکل، تشابه سوره با tf-idf نیز با افزایش فاصله ترتیبی روند نزولی را نشان می‌دهد. از شکل‌های ۱۰ و ۱۱ می‌توان نتیجه گرفت که فاصله سوره‌ها و تشابه آن‌ها حداقل در سطح کلان با هم همبستگی دارند، به طوری که هر چه فاصله سوره‌ها بیشتر باشد، میانگین تشابه آن‌ها کمتر خواهد بود. این یافته به نوعی سازماندهی کلان سوره‌ها را نشان می‌دهد که در کنار یکدیگر قرار گرفته‌اند. به دلیل فضای کم، تحلیل میکروسکوپی تشابه سوره‌ها را به مقاله دیگری موکول می‌کنیم و تنها ذکر این نکته را کافی می‌دانیم که بر اساس این نتیجه دو فرضیه قابل تصور است: ۱. اینکه بیشتر سوره‌های مجاور از نظر مفهومی با هم مرتبط هستند و سوره‌های قرآنی دارای پل ارتباطی بین سوره‌ای هستند؛

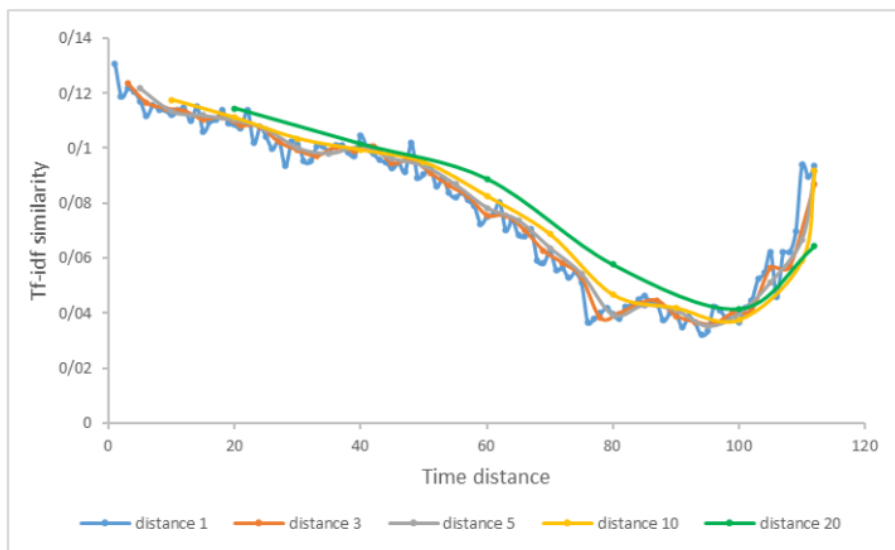
۲. سوره‌های قرآن در خوشه‌های سوره‌ای به هم مرتبط هستند که این خوشه‌ها در کنار هم چیده شده‌اند. هر کدام از این دو فرضیه مصادیق زیادی دارد. به‌عنوان مثال، بعد از سوره فاتحه‌الکتاب که در ب ورود به کتاب است، با آیه ﴿ذَلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ﴾ در سوره بقره مواجه می‌شویم. همچنین، سوره‌های شماره ۱۰ تا ۱۵ همگی با «الر» شروع می‌شوند و عنوان همگی آن‌ها در ابتدای سوره به کتاب یا آیات کتاب مرتبط است!

برای مطالعه دقیق‌تر، نتایج به‌دست آمده را با شباهت سوره‌های قرآنی با فاصله زمانی ترتیب نزول آن‌ها مقایسه کردیم. شکل ۱۲ شباهت سوره‌ها را با فاصله زمانی نزول نشان می‌دهد.



شکل (۱۲): تشابه سوره‌های مختلف در مقایسه با ترتیب نزول آنها از یکدیگر بر اساس شباهت RS (محور عمودی: شباهت سوره‌ها؛ محور افقی: فاصله زمانی)

این نمودار برای تشابه tf-idf به صورت زیر خواهد بود.



شکل (۱۳): تشابه سوره‌های مختلف در برابر ترتیب نزول آنها از یکدیگر بر اساس شباهت RS همان‌طور که مشاهده شد، تشابه سوره‌ها نسبت به ترتیب نزول آنها، به‌ویژه بر اساس شباهت RA، روند خاصی را دنبال نمی‌کند. در روش RA، تشابه در ابتدا با افزایش فاصله ترتیبی نزول دو سوره تغییر چندانی نمی‌کند. شباهت برای فواصل بین ۶۰ و ۱۰۰ کاهش یافته و سپس افزایش می‌یابد. همین توصیف با شدت کمتر برای tf-idf نیز صادق است.

بر این اساس، می‌توان نتیجه گرفت که ترتیب فاصله سوره‌ها در قرآن با شباهت آنها مرتبط است، به طوری که سوره‌های نزدیک به یکدیگر از نظر ترتیب نیز به احتمال نسبتاً بالایی شبیه‌تر هستند. برخلاف فاصله ترتیبی سوره‌ها، چنین نظمی برای فاصله زمانی نزول سوره‌ها صادق نیست. بنابراین، می‌توان نتیجه گرفت که ترتیب سوره‌ها در قرآن از نظمی منطقی پیروی می‌کند که مستلزم بررسی دقیق‌تر است.

علاوه بر این، بر این اساس می‌توان تشخیص داد که چرا سوره‌های قرآن به ترتیب نزول خود ترتیب داده نشده‌اند. با بررسی دقیق‌تر این موضوع، نتایج حاشیه‌ای متعددی به دست می‌آید، از جمله اینکه چرا اولین سوره در ابتدای قرآن به نام فاتحه، به معنای «کتاب بازکننده»، است یا اینکه چه رابطه‌ای بین دسته‌های نزدیک سوره‌ها وجود دارد.

نتیجه‌گیری

این تحقیق با دو هدف انجام شد: بررسی ساختارمندی درونی هر سوره بر اساس نظریه‌های «وحدت موضوعی» و «احکام و تفصیل»، و همچنین بررسی «ترتیب سوره‌ها» در کل قرآن. در این راستا، ابتدا داده‌های قرآنی تهیه و پاکسازی شد. سپس با به‌کارگیری روش‌های پردازش زبان طبیعی word2vec، tf-idf و همچنین باهم‌آیی ریشه‌ها در آیات، تشابه ریشه‌های قرآنی به‌دست آمد.

با تشابهات محاسبه‌شده، در گام نخست پیوند عنوان سوره با محتوای بدنه محاسبه شد. سپس وحدت موضوعی هر سوره با محاسبه شباهت مفاهیم درونی هر سوره مورد مطالعه قرار گرفت. همچنین، وجود ساختار احکام و تفصیل در سوره‌های قرآنی مورد سنجش قرار گرفت. شباهت بین مفاهیم درون‌سوره‌ای و عنوان سوره و همچنین مفاهیم درون‌سوره‌ای در مقایسه با نتایج حالت تصادفی نشان داد که شباهت موضوع هر سوره به مفاهیم بدنه و هم شباهت بین مفاهیم به یکدیگر، بسیار بیشتر از شباهت در حالت تصادفی است. این بدان معناست که سوره‌های قرآنی عمدتاً حول یک موضوع واحد شکل گرفته‌اند.

همچنین، با محاسبه شباهت بین فصل اول و آخر و نیز بین فصل اول و سایر فصل‌های سوره‌های مختلف، سازمان‌دهی سوره‌ها بر اساس ساختار «احکام و تفصیل» بررسی شد. در نهایت، بر اساس بررسی همبستگی بین ترتیب سوره‌ها در کل قرآن و ترتیب نزول آن‌ها با شباهت سوره‌ها، به این نتیجه رسیدیم که ترتیب سوره‌ها در کل قرآن نیز نسبتاً سازمان‌یافته است.

کارهای آینده

از آنجایی که این مقاله اولین کار در زمینه مطالعه الگوریتمیک ساختارمندی سوره‌ها، بالخصوص ساختارمندی ترتیب چینی سوره‌های قرآن است، منطقی است که بر روی هر سوره به‌صورت جداگانه، کارهای مشابهی با جزئیات بیشتری انجام شود. علاوه بر این، مطالعه تشابه ساختار سوره‌ها و همچنین خوشه‌های سوره‌ای با استفاده از روش‌های پردازش زبان طبیعی پر اهمیت به نظر می‌رسد. از سوی دیگر، از آنجایی که فصل‌بندی علامه طباطبایی از منظر خاصی صورت پذیرفته است، در کارهای آتی بسیار مهم است که فصل‌بندی دستی یا خودکار به‌عنوان پیش‌نیاز مطالعه دقیق ساختار سوره‌ها باشد.

علاوه بر این، با مشارکت متخصصان انسانی به جای الگوریتم‌های شباهت پردازش زبان طبیعی، نتایج دقیق‌تری در مطالعه ساختار سوره به‌دست می‌آید. در نهایت، سایر کارهای احتمالی آینده بر این اساس شامل کارهایی است که بر اساس روش‌های دیگر محاسبه تشابه، مقایسه ساختار قرآن با کتاب‌های دیگر و بررسی سازماندهی خوشه‌های قرآنی انجام می‌شود.

فهرست منابع

- آرام، محمد رضا، لایقی، فاطمه (۱۳۹۵)، «بررسی ساختاری سوره مائده با رویکرد ساختار درختی»، پژوهشنامه قرآن و حدیث، ش ۱۹، ص ۵۵-۷۷.
- جیگاره، مینا، صادقی زهرا (۱۳۹۵)، «بررسی و تحلیل سوره انفطار با تکیه بر نظریه ساختارگرایی»، نشریه پژوهشنامه نقد ادب عربی، ش ۱۳، ص ۵۰-۷۴.
- خامه‌گر محمد (۱۳۸۱)، «درآمدی بر تفسیر ساختاری قرآن»، پژوهش‌های قرآنی، ش ۲۹-۳۰، ص ۲۰۶-۲۷۱، (ویژه نامه علامه طبرسی).
- (۱۳۸۵)، «نگاهی به نخستین ترجمه ساختاری قرآن کریم»، بینات، ش ۴۹ و ۵۰، ص ۲۷۸-۲۹۱.
- (۱۳۸۳)، «عبارت‌پردازی داستان‌های قرآنی و مقاصد سوره‌ها (به فارسی)»، گلستان قرآن، ش ۱۷۹، ص ۱۳-۱۷.
- (۱۳۸۶)، «نظریه هدفمندی سوره‌ها: مبانی و پیشینه»، پژوهش‌های قرآنی، دوره ۱۳، ش ۵۱ و ۵۲، ص ۱۸۲-۲۱۳.
- (۲۰۰۲)، «ساختار هندسی سوره‌های قرآن، درآمدی بر تفسیر ساختاری قرآن»، گلستان قرآن، ش ۱۳۸، ص ۹-۱۳.
- دهقانی فارسانی، یونس (۱۳۹۷)، «ساختار سوره انشقاق» بلاغ مبین، ش ۱۴، ص ۳-۱۴.
- طباطبایی، محمدحسین (۱۳۷۴)، المیزان فی تفسیر القرآن، انتشارات اسلامی، وابسته به جامعه مدرسین حوزه علمیه قم، جلد ۲.
- فتاحی‌زاده، فتحیه، ذاکری، مرضیه (۱۳۹۵)، «رویکرد ساختارشناسانه به سوره کهف»، مطالعات تفسیری، ش ۲۵، ص ۱۰۱-۱۲۰.
- A. Alfaifi and E. Atwell, (2016), "Comparative evaluation of tools for Arabic corpora search and analysis," International Journal of Speech Technology, vol. 19, pp. 347-357.
- A. B. M. Sharaf and E. Atwell, (2012), "QurSim: A corpus for evaluation of relatedness in short texts," in LREC, pp. 2295-2302.

- A. Clauset, C. R. Shalizi, and M. E. Newman, (2009), "Power-law distributions in empirical data," SIAM review, vol. 51, pp. 661-703.
- A. J. Arberry, (1996), The Koran interpreted: A translation: Simon and Schuster.
- A. R. Yauri, R. A. Kadir, A. Azman, and M. A. A. Murad, (2013), "Quranic verse extraction base on concepts using OWL-DL ontology," Research Journal of Applied Sciences, Engineering and Technology, vol. 6, pp. 4492-4498.
- A.-B. M. Sharaf and E. Atwell, (2012), "QurAna: Corpus of the Quran annotated with Pronominal Anaphora," in LREC, pp. 130-137.
- E. Atwell, (2009), "A Corpus-based Computational Model for Knowledge Representation of the Quran," in Proceedings of the Corpus Linguistics Conference 2009 (CL2009), p. 169.
- H. U. Khan, S. M. Saqlain, M. Shoaib, and M. Sher, (2013), "Ontology based semantic search in Holy Quran," International Journal of Future Computer and Communication, vol. 2, p. 570.
- K. Dukes and T. Buckwalter, (2010), "A dependency treebank of the Quran using traditional Arabic grammar," in Informatics and Systems (INFOS), 2010 The 7th International Conference on, pp. 1-7.
- M. A. Sherif and A.-C. Ngonga Ngomo, (2015), "Semantic Quran," Semantic Web, vol. 6, pp. 339-345.
- M. Alhawarat, (2015), "Extracting Topics from the Holy Quran Using Generative Models," INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS, vol. 6, pp. 288-294.
- M. H. Tabatabai, (1996), "al-Mizan fi Tafsir al-Qur'an," Qom: Islamic Publications Society of Seminary Teachers of Qom, vol. 2.
- M. Mitzenmacher, (2004), "A brief history of generative models for power law and lognormal distributions," Internet mathematics, vol. 1, pp. 226-251.
- M. Shoaib, M. N. Yasin, U. K. Hikmat, M. I. Saeed, and M. S. H. Khiyal, (2009), "Relational WordNet model for semantic search in Holy Quran," in Emerging Technologies, 2009. ICET 2009. International Conference on, pp. 29-34.

- P. Soucy and G. W. Mineau, (2005), "Beyond TFIDF weighting for text categorization in the vector space model," in IJCAI, pp. 1130-1135.
- Q. V. Le and T. Mikolov, (2014), "Distributed Representations of Sentences and Documents," in ICML, pp. 1188-1196.
- R. Bell, (1953), "Introduction to the Quran."
- R. Iqbal, A. Mustapha, and Z. Mohd. Yusoff, (2013), "An experience of developing Quran ontology with contextual information support," *Multicultural Education & Technology Journal*, vol. 7, pp. 333-343.
- R. Ismail, Z. A. Bakar, and N. A. Rahman, (2015), "Ontology Learning Framework for Quran".
- R. R. Larson, (2010), "Introduction to information retrieval," *Journal of the American Society for Information Science and Technology*, vol. 61, pp. 852-853.
- S. A. Pitchay and F. Ridzuan, (2016), "A Systematic Review Analysis for Quran Verses Retrieval," *Journal of Engineering and Applied Sciences*, vol. 100, pp. 629-634.
- S. K. Hamed and M. J. Ab Aziz, (2016), "A question answering system on Holy Quran translation based on question expansion technique and Neural Network classification," *Journal of Computer Science*, vol. 12, pp. 169-177.
- T. Mikolov, K. Chen, G. Corrado, and J. Dean, (2013), "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781.
- Y. Zhang, R. Jin, and Z.-H. Zhou, (2010), "Understanding bag-of-words model: a statistical framework," *International Journal of Machine Learning and Cybernetics*, vol. 1, pp. 43-52.

References

- Alfaifi, A., & Atwell, E. (2016). Comparative evaluation of tools for Arabic corpora search and analysis. *International Journal of Speech Technology*, 19, 347-357.
- Alhawarat, M. (2015). Extracting topics from the Holy Quran using generative models. *International Journal of Advanced Computer Science and Applications*, 6, 288-294.
- Aram, M. R., & Laiqi, F. (2016). Structural analysis of Surah Al-Ma'idah with a tree-structure approach. *Quran and Hadith Research Journal*, (19), 55-77.
- Arberry, A. J. (1996). *The Koran interpreted: A translation*. Simon and Schuster.
- Atwell, E. (2009). A corpus-based computational model for knowledge representation of the Quran. In *Proceedings of the Corpus Linguistics Conference 2009 (CL2009)* (p. 169).
- Bell, R. (1953). *Introduction to the Quran*. Edinburgh University Press.
- Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Review*, 51, 661-703.
- Dehghani Farsani, Y. (2008). The structure of Surah Al-Inshiqaq. *Balagh Mubeen*, (14), 3-14.
- Dukes, K., & Buckwalter, T. (2010). A dependency treebank of the Quran using traditional Arabic grammar. In *Informatics and Systems (INFOS), 2010 The 7th International Conference on* (pp. 1-7).
- Fattahi Zadeh, F., & Zakeri, M. (2016). A structuralist approach to Surah Al-Kahf. *Interpretive Studies*, 7(25), 101-120.
- Hamed, S. K., & Ab Aziz, M. J. (2016). A question answering system on Holy Quran translation based on question expansion technique and neural network classification. *Journal of Computer Science*, 12, 169-177.
- Iqbal, R., Mustapha, A., & Yusoff, Z. M. (2013). An experience of developing Quran ontology with contextual information support. *Multicultural Education & Technology Journal*, 7, 333-343.
- Jigareh, M., & Sadeghi, Z. (2015). Review and analysis of Surah Al-Infitar based on the theory of structuralism. *Journal of Arabic Literary Criticism*, 7(13), 50-74.
- Khan, H. U., Saqlain, S. M., Shoaib, M., & Sher, M. (2013). Ontology based semantic search in Holy Quran. *International Journal of Future Computer and Communication*, 2, 570.

- Khamegar, M. (2002). The geometric structure of the Quranic surahs: An introduction to the structural interpretation of the Quran. *Golestan-e-Quran*, (138), 9-13.
- Khamegar, M. (2003). An introduction to the structural interpretation of the Quran. *Quranic Studies*, (29-30), 206-271.
- Khamegar, M. (2004). Phrasing of Quranic stories and the objectives of the surahs (in Persian). *Golestan-e-Quran*, (179), 13-17.
- Khamegar, M. (2006). A look at the first structural translation of the Holy Quran. *Beynat*, 49(50), 278-291.
- Khamegar, M. (2008). The theory of the purposeful surahs: Foundations and background. *Quranic Studies*, 13(51-52), 182-213.
- Larson, R. R. (2010). Introduction to information retrieval. *Journal of the American Society for Information Science and Technology*, 61, 852-853.
- Le, Q. V., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *ICML* (pp. 1188-1196).
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1, 226-251.
- Pitchay, S. A., & Ridzuan, F. (2016). A systematic review analysis for Quran verses retrieval. *Journal of Engineering and Applied Sciences*, 100, 629-634.
- Sharaf, A.-B. M., & Atwell, E. (2012). QurAna: Corpus of the Quran annotated with pronominal anaphora. In *LREC* (pp. 130-137).
- Sharaf, A.-B. M., & Atwell, E. (2012). QurSim: A corpus for evaluation of relatedness in short texts. In *LREC* (pp. 2295-2302).
- Sherif, M. A., & Ngonga Ngomo, A.-C. (2015). Semantic Quran. *Semantic Web*, 6, 339-345.
- Shoaib, M., Yasin, M. N., Hikmat, U. K., Saeed, M. I., & Khiyal, M. S. H. (2009). Relational WordNet model for semantic search in Holy Quran. In *Emerging Technologies, 2009. ICET 2009. International Conference on* (pp. 29-34).
- Soucy, P., & Mineau, G. W. (2005). Beyond TFIDF weighting for text categorization in the vector space model. In *IJCAI* (pp. 1130-1135).
- Tabataba'i, M. H. (1995). *Al-Mizan fi Tafsir al-Qur'an* (Vol. 2). Islamic Publications affiliated with the Society of Seminary Teachers of Qom.

- Tabataba'i, M. H. (1996). *Al-Mizan fi Tafsir al-Qur'an* (Vol. 2). Qom: Islamic Publications Society of Seminary Teachers of Qom.
- Yauri, A. R., Kadir, R. A., Azman, A., & Murad, M. A. A. (2013). Quranic verse extraction based on concepts using OWL-DL ontology. *Research Journal of Applied Sciences, Engineering and Technology*, 6, 4492-4498.
- Zhang, Y., Jin, R., & Zhou, Z.-H. (2010). Understanding bag-of-words model: A statistical framework. *International Journal of Machine Learning and Cybernetics*, 1, 43-52.