

تحلیل نظرسنجی کتاب درسی آزمون متمرکز در دانشگاه جامع علمی کاربردی با رویکرد متن کاوی

مریم ملاباقر^۱، مصطفی یوسفی طرزجان^۲، اسرافیل علاء^۳

(دریافت: ۱۴۰۰/۱/۷ - پذیرش: ۱۴۰۰/۵/۱۹، نوع مقاله: پژوهشی)

چکیده

در سال‌های اخیر داده کاوی آموزشی با توجه به انبوهی از داده‌ها و دشواری تحلیل نتایج در حوزه آموزش عالی بسیار مورد استقبال قرار گرفته است. جمع‌بندی نظرات مدرسان در خصوص منابع درسی نیز یکی از چالش‌های اساسی در تصمیم‌گیری چاپ/عدم چاپ مجدد کتب است. این پژوهش بر آن است که از طریق روش نیمه خودکار استخراج متن کتاب/اخلاق حرفه‌ای اثر ناصر صبحی قراملکی (۱۳۹۸)، برای دانشجویان آزمون متمرکز را ارزیابی نماید. در این مقاله نتایج پرسش‌نامه تکمیل‌شده درس اخلاق حرفه‌ای ۴۷ نفر از مدرسان این کتاب درسی، در خصوص سؤال تشریحی درمورد چاپ/عدم چاپ مجدد با روش متن کاوی مورد تحلیل قرار گرفته و نیز از نرم‌افزار متن کاوی «فارسی‌یار» و کدنویسی پایتون و نیز از روش‌های استخراج کلمات کلیدی متن، خلاصه‌سازی خودکار متن، مدل‌سازی موضوعی (شناسایی موضوعات مجموعه متون) و درنهایت از ترسیم ابرکلمات برای تحلیل نظرات ارائه درج شده، استفاده شده است. یافته‌ها حاکی است که روش مدل‌سازی موضوعی برای شناسایی موضوعات مجموعه به کار رفته است و روش ترسیم ابرکلمات که شکل بصری از نظرات متنی را ارائه می‌دادند نسبت به سایر روش‌ها در این پژوهش مناسب‌تر بودند. توصیه می‌شود با بررسی نظرات مدرسان، با بازنویسی و ویراستاری مجدد این کتاب، تجدید چاپ صورت گیرد.

کلیدواژه‌ها: ارزیابی کتاب، داده کاوی، متن کاوی، آزمون متمرکز، و دانشگاه جامع علمی کاربردی.

۱. عضو هیئت علمی، گروه صنعت، دانشگاه جامع علمی کاربردی، تهران، ایران (mollabagher@gmail.com)، (نویسنده مسئول).

۲. عضو هیئت علمی، گروه صنعت، دانشگاه جامع علمی کاربردی، کرج، تهران.

۳. معاون مرکز آزمون‌های سازمان سنجش آموزش کشور، تهران، ایران.

An Analysis of the Opinion Poll on the Centralized Exam Textbook in the University Applied Science & Technology with a Text-Analytics Approach

Maryam Mollabagher¹, Mostafa Yousefi Tezerjan², Esrafil Ala³

Received: 27/03/2021- Accepted: 10/08/2021

Abstract

Owing to the availability of large amounts of data and the existing difficulty in analyzing the results in the area of higher education, educational data mining has come to gain a lot of importance in recent years. Drawing a conclusion on the views of university teachers about educational sources, too, is one of the main challenges in deciding whether or not to reprint books. The aim of this study is to evaluate the contents of the book “Professional Ethics” by Nasser Sobhi Gharamaleki (2019), written for centralized exam students, by using a semi-automatic text extraction method. In this paper, the views provided by 47 instructors of this textbook, collected through a questionnaire, regarding the descriptive question of whether or not the book should be reprinted were analyzed by resorting to the text mining technique. To analyze the views received the researchers have used the “Farsiyar” text mining software and Python coding, as well as methods of extracting text keywords, automatic text summarization, thematic modeling (identifying the subjects included in the texts), and the method of writing meta-words. The findings of the study indicate that the method used for topic modeling to identify the subjects covered and the method of writing meta-words that provided a visual form of textual comments were more appropriate than other methods used in this study. Keeping in view the views of the lecturers and instructors, it has, therefore, been recommended the future reprints of this book should take place after rewriting and re-editing it.

Keywords

Book Assessment, Data Mining, Text Mining, Centralized Testing, and Comprehensive University of Applied Sciences & Technology

-
1. Faculty Member of the University of Applied Sciences & Technology, Faculty of Industry, Tehran, Iran, (mollabagher@gmail.com) (Corresponding Author).
 2. Faculty Member of the University of Applied Sciences & Technology, Faculty of Industry, Karaj, Iran.
 3. Deputy of Examination Supervisor of the National Organization of Educational Testing, Tehran, Iran.

مقدمه

کتاب درسی یکی از رسانه های مهم بخش درونداد در برنامه درسی هر نظام آموزشی است. به نظر پین گل کتاب درسی نه تنها رسانه ای برای تحقق اهداف آموزشی، بلکه ابزاری مهم در حفظ و انتقال تجارب، مطالعات و تحقیقات و تولیدات فکری اندیشمندان هر جامعه است (کاردان، ۱۳۸۲). برنامه درسی در قلب آموزش وجود دارد؛ بدین معنا که بدون توجه به برنامه درسی، آموزش عبث و بیهوده خواهد بود (Null, 2016).

داده کاوی، استخراج اطلاعات مفید پنهان از یک مجموعه داده ها از طریق تحلیل علمی و روش هایی است که روند داده ها را شناسایی می کند و الگوهای پنهان را در مجموعه داده ها قرار می دهد و به همین ترتیب، داده کاوی را می توان کشف دانش دانست (Hussain et al., 2018). بررسی نظرات استفاده کنندگان یکی از مباحث جدی در مدیریت ارتباط با مشتریان است. در این خصوص ابزارهای زیادی مورد استفاده قرار گرفته است. اما در خصوص نظرات متنی نیاز به ابزارهایی برای استخراج دانش در متن وجود دارد. متن کاوی یکی از ابزارهای دسته بندی و تحلیل نظرات متنی است. (Wu et al., 2021).

با توجه به نقش و اهمیت بالای منابع درسی در موفقیت در این آزمون، این پژوهش بر آن است تا با متن کاوی نظرسنجی صورت گرفته از مدرسان در آزمون متمرکز و تحلیل نتایج به دست آمده، گامی در جهت کیفی سازی این آزمون ها و سنجش عملکرد دانشجویان فراهم آورد و ابزاری در جهت تصمیم سازی سنجش کیفیت در اختیار مدیران ارشد دانشگاه قرار دهد.

دانشگاه جامع علمی کاربردی بزرگ ترین متولی نظام آموزش عالی علمی کاربردی با حدود ۶۰۰ برنامه درسی در حال اجرا، بیش از ۶۰۰ مرکز آموزشی فعال، بیش از ۲۰۰ هزار دانشجو، بیش از ۲ هزار دانش آموخته در سراسر کشور با ۱۷ درصد سهم از کل دانشجویان دانشگاه های ایران است که زیر نظر وزارت علوم، تحقیقات و فناوری اداره می شود. این دانشگاه نسل سوم با ساختار تشکیلاتی حوزه ستاد مرکزی، ۳۲ واحد استانی، بالغ بر ۶۰۰ مرکز آموزش علمی کاربردی تحت نظارت، به دنبال فراهم آوردن موجهی است که مشارکت سازمان ها و دستگاه های اجرایی دولتی و غیردولتی را برای آموزش نیروی انسانی متخصص و مورد نیاز بخش های مختلف اقتصادی، اجتماعی و فرهنگی کشور ممکن سازد.

از آنجا که آزمون ها در تعیین آنچه در محیط آموزشی صورت می گیرد، از لحاظ

موضوع و نحوه تدریس مربیان و یادگیری فراگیران نقش عمده‌ای ایفا می‌کنند و می‌توانند تدریس و یادگیری را تحت تأثیر قرار دهند (Rehmani, 2003) و نیز با توجه به تأکید ریاست دانشگاه مبنی بر اجرای آزمون‌های متمرکز در دروس منتخب در هر مقطع تحصیلی و به منظور حفظ و ارتقای کیفیت آموزش‌ها و نظارت بر حسن اجرای صحیح آن مصوبه اجرای برگزاری آزمون‌های متمرکز پایان نیم‌سال در مرداد ۱۳۹۷ اعلام گردید. این آزمون‌ها در مرحله اول صرفاً در نیم‌سال اول تحصیلی ۱۳۹۷-۱۳۹۸ در استان البرز به صورت آزمایشی انجام شد و در نیم‌سال دوم تحصیلی ۱۳۹۷-۱۳۹۸ نیز در ۶ استان در کل کشور با پیشنهاد و استقبال رئیس واحد استانی مربوط انجام شد. از نیم‌سال دوم تحصیل نیز به صورت سراسری و در کل کشور برگزار شد. از آن پس تاکنون در هر نیم‌سال به صورت کشوری برگزار شده است و هم‌اکنون در بیش از ۲۵ عنوان درسی در پایان هر نیم‌سال و هم‌زمان در سراسر کشور اجرا می‌شود.

کتاب درسی به لحاظ ساختاری شاخص‌هایی دارد که درسی بودن آن را نشان می‌دهد. در این ساختار، هر درس باید حداقل دارای معیارهای اهداف، مقدمه یا پیش‌سازمان‌دهنده، متن، خلاصه، تمرین و در حد مطلوب باشد؛ در غیر آن، سطح کیفی آموزش کاهش می‌یابد. بنابراین، یکی از شیوه‌های اصلی برای ارزیابی کیفیت کتاب‌های درسی، نقد و بررسی آن‌ها به لحاظ شاخص‌های ساختاری است (موسوی و حبیبی، ۱۳۹۹).

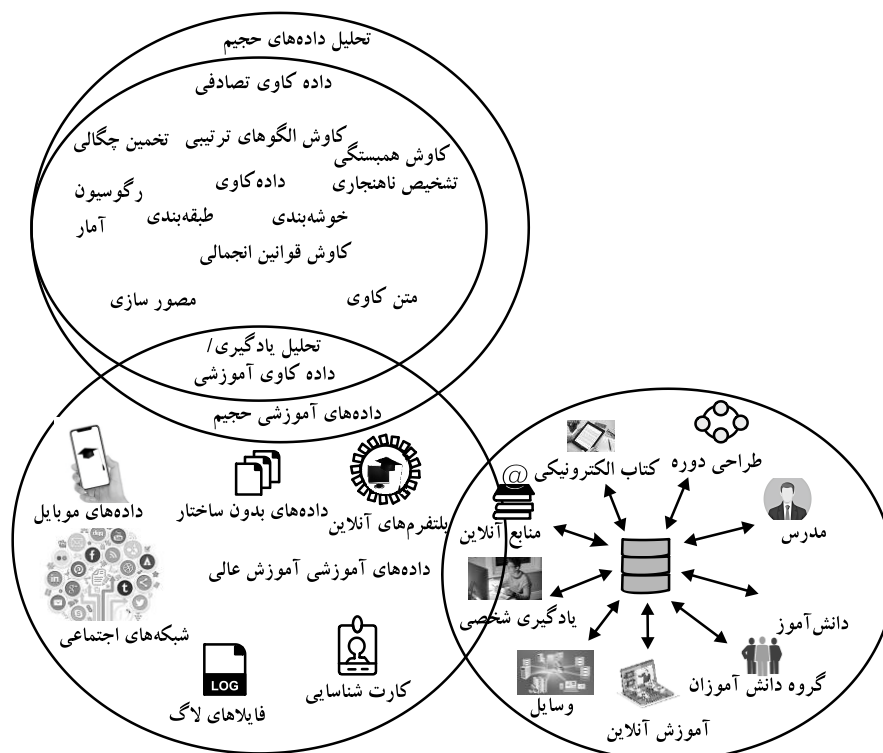
کتاب درسی رسمی‌ترین رسانه در زمینه ارتباط علمی است که به رغم گسترش استفاده از وسایل جدید اطلاع‌رسانی، نقش اساسی و مهمی در یادگیری برای همه سطوح دارد. برخورداری از بهره‌وری در تألیف کتاب‌های درسی نیازمند ارزیابی و نقد مستمر آن‌ها و تلاش برای از بین بردن کاستی‌هاست (رضی، ۱۳۸۸).

لذا با توجه به نقش و جایگاه کتاب درسی در افزایش بهره‌وری و کیفیت آموزشی و نیز اهمیت بالای ارزیابی آن، این مقاله بر آن است که با بررسی و نقد کتاب/خلاق حرفه‌ای (صبحی‌قراملکی، ۱۳۹۸)، گامی در جهت اعتلای این مهم بردارد.

داده‌کاوی در آموزش عالی

پژوهش آلدوها و همکارانش (Aldowah et al., 2019) نشان می‌دهد که ابعاد تحقیقاتی محققان داده‌کاوی آموزشی در ارتباط با فعالیت‌های آموزشی به شکل‌های مختلفی از

جمله: یادگیری مشارکتی رایانه‌ای با هدف کاوش الگوهای مشاکتی در مباحث آموزشی، پشتیبانی از استادان در مدل‌سازی مشارکت فراگیران، ارزیابی منابع آموزشی دانشگاه و بهبود برنامه تحصیلی، شناسایی عوامل مربوط به موفقیت، شکست و قصد ترک تحصیل، راهبردها و برنامه‌ریزی سازمانی، درک پشتیبانی معلمان و تصمیم‌گیری اداری است. در این روش، روش‌های داده‌کاوی مختلف از جمله: طبقه‌بندی، خوشه‌بندی، آمار، رگرسیون، کاوش الگوهای ترتیبی و متن‌کاوی به کار برده می‌شود. استفاده از روش‌های مختلف داده‌کاوی می‌تواند به عنوان مقدمه‌ای برای تغییر اصولی مورد توجه قرار گیرد و به عنوان یک ابزار می‌تواند به مؤسسات آموزش عالی کمک کند تا راه‌حل‌هایی برای مسائل خاص خود بیابند؛ همچنین برای پیشنهاد دادن محیط‌های آموزشی تعاملی‌تر، فرایند یادگیری و یاد دادن مؤثرتر و انطباق دادن بیشتر عملکرد فراگیران و مدرس به کار روند. شکل ۱ تصویری از کاربرد داده‌کاوی در بخش‌های مختلف آموزش عالی را نشان می‌دهد.



شکل ۱ کاربرد داده‌کاوی در آموزش عالی

همان‌طور که در شکل دیده می‌شود، داده‌های آموزش عالی می‌توانند از حوزه آموزش آنلاین، شامل کتاب الکترونیک و دوره الکترونیک دانشجویان و وسایل یادگیری آنلاین اخذ شود. از سوی دیگر، با توجه به حجم عظیم داده‌هایی که از داده‌های موبایل، شبکه‌های اجتماعی، پلتفرم‌های آنلاین استخراج می‌شود، می‌توان کار تحلیل داده‌های حجیم را با استفاده از ابزارهای داده کاوی مانند خوشه‌بندی و متن کاوی انجام داد.

متن کاوی

امروزه مقدار زیادی داده، به امید مفید بودن در آینده، از سوی سازمان‌ها جمع‌آوری و ذخیره می‌شود؛ همین امر، مدیریت انبوهی از داده‌ها و استخراج دانش مناسب برای حمایت از تصمیمات را با چالش مواجه می‌کند (Fan & Bifet, 2013).

یکی از انواع داده‌هایی که در گذشته کمتر مورد توجه کسب و کارها و دانشگاه‌ها بوده است، داده‌های متنی است. تمایل به دسترسی به داده‌های شفاف و سختی تحلیل داده‌های متنی باعث شده است که به این داده‌ها توجه کمتری شود (Xu et al., 2016). مفهوم متن کاوی که به دریافت همه اطلاعات مورد نیاز از کوه داده‌های متنی اشاره می‌کند، تقریباً عمری برابر با بازیابی اطلاعات دارد. متن کاوی دارای ویژگی‌های منحصر به فرد و اساسی است که باعث شده بین آن و بازیابی اطلاعات تمیز قائل شوند. هدف متن کاوی به دست آوردن اطلاعات مفیدی از داده‌های متنی است که ذاتاً ساختاریافته، غیرمتشکل و نامنظم هستند (Jusoh & Alfawareh., 2009).

متن کاوی شاخه‌ای از داده کاوی است، یعنی کشف الگوهای ناشناخته از بین داده‌های متنی. این رویکرد به دنبال آن است که اطلاعات مفید را از داده‌های متنی غیر ساختاریافته استخراج کند. متن کاوی شباهت زیادی با داده کاوی دارد؛ با این تفاوت که ابزارهای داده کاوی برای مدیریت داده‌های ساختاریافته طراحی شده است، اما متن کاوی می‌تواند بر مجموعه‌ای از داده‌های غیر ساختاریافته نیز برای استخراج الگوهای ناشناخته جدید تمرکز کند (Miranda Santo et al., 2006). اهداف متن کاوی به روایت پرچمی و همکارانش (۱۳۹۹) در جدول ۱ (Singh et al., 2007) آمده است.

جدول ۱ اهداف متن کاوی

اهداف	تعریف
استخراج ویژگی	مشخص کردن تمایز میان عبارت اسمی، شخص، مکان و یا سازمان از سایر اشیاء است. عملیات مرتبط با این هدف، شامل استخراج کلمات و محاسبه تعداد دفعاتی است که هر کلمه در متن تحلیل شده است.
هدایت مبتنی بر آزمون	استخراج کننده متن را قادر می سازد تا کلمات مرتبط را در متن بیابد و روابط مهم بین آن‌ها را به هم ارتباط دهد.
جست‌وجو و بازاریابی	کاربر را بر آن می‌دارد تا براساس معیارهای جست‌وجوی از پیش تعیین شده به جست‌وجو و بازاریابی اطلاعات بپردازد.
خوشه‌بندی	عملیات گروه‌بندی کلمات کلیدی براساس میزان شباهت‌ها و یا عدم تشابه را انجام می‌دهد.
طبقه‌بندی	عملیات تعریف مجموعه‌ای از کلمات مخصوص دامنه و روابط بین آن‌ها از طریق الگوریتم‌های طبقه‌بندی است.
خلاصه‌سازی	عملیات کاهش مقدار داده‌های متنی با حفظ عناصر اصلی است.
تجزیه و تحلیل روند	فرایندی است که طی آن برای کشف روند از داده‌های متنی وابسته به سازمان استفاده می‌شود.
قواعد انجمنی	مرتبط ساختن الگوی استخراج شده با الگوهایی است که از قبل کشف و استخراج شده‌اند.
مصورسازی	در مصورسازی از استخراج ویژگی و کلمات کلیدی فهرست‌شده برای ایجاد یک نمایش گرافیکی استفاده می‌شود که می‌تواند کاربر را برای شناسایی موضوعات اصلی با مفاهیم مهمشان یاری دهد.

با توجه به این که داده‌های بدون ساختار متنی مانند نظرسنجی که شامل انبوهی از اطلاعات به هم پیوسته هستند که در قالب جداول پایگاه‌های اطلاعاتی رابطه‌ای قرار نمی‌گیرند، بهترین روش برای کشف روابط میان متون آن‌ها، متن کاوی است (Shettar & Shobha, 2007). لذا در این پژوهش از روش متن کاوی برای تحلیل نظرات مدرسان استفاده شده است.

پیشینه پژوهش

اگرچه آزمون در وهله اول و از دید عموم، ابزاری نه چندان خوشایند برای سنجش محفوظات متعلمان در دست استادان تعبیر می‌گردد، چنانچه یک آزمون با دقت و رعایت معیارهای لازم طراحی شود، به ابزاری مناسب برای نظارت بر فعالیت‌های یادگیری و

عاملی مؤثر در ارتقای کمی و کیفی فرایند آموزش تبدیل می گردد. از جمله نکات مهم در ارزشیابی، اندازه گیری دقیق دانش و مهارت هایی است که در چارچوب اهداف یک ماده درسی و یا کل دوره قرار دارد. از این رو، طراحی سؤالات آزمون نباید به گونه ای باشد که به طور مشخص آن دسته از مطالب و آموخته هایی را مورد سنجش قرار دهد که طی فرایند آموزش استاد یا محیط به دانشجویان انتقال یافته است. علی رغم تلاش در جست و جوی اینترنتی در مورد پژوهش های صورت گرفته با موضوع متن کاوی در آموزش عالی، اندکی در این خصوص یافت شد که عبارت اند از:

- کاردان و کیهانی نژاد (۱۳۹۱) در پژوهشی با عنوان «ارائه مدلی برای استخراج اطلاعات از مستندات متنی، مبتنی بر متن کاوی در حوزه یادگیری الکترونیکی»، ضمن بررسی پژوهش های انجام شده در حوزه متن کاوی به ارائه روش ها و کاربردهای آن در آموزش الکترونیکی پرداختند و طی این مطالعات، پژوهش های مرتبط در حوزه آموزش الکترونیکی را طبقه بندی کردند. پس از طبقه بندی پژوهش ها، مسائل و راهکارهای مرتبط با مسائل مطرح شده در آن کارها استخراج شدند.

- عظیمی همت و شمس عزت (۱۳۹۴) در پژوهشی ضمن تأکید بر تحقیقات اندک متن کاوی در زمینه متون فارسی، به دلیل پیچیدگی های ساختاری آن، مروری بر روش های اصلی متن کاوی و کاربرد آن ها در پردازش متون زبان فارسی انجام دادند.

- عبدی نسب و همکارانش (۱۳۹۷) در پژوهشی با عنوان «طراحی الگوی سیستم نرم افزاری اعتباریابی کتاب های دانشگاهی مبتنی بر متن کاوی»، به طراحی و اعتباریابی مقیاس سنجش کتاب های درسی دانشگاهی با استفاده از متن کاوی متون تخصصی و ساماندهی الگوریتم نرم افزاری پرداختند. نمونه آماری آنان مقالات پژوهشی انگلیسی در پایگاه های معتبر جهانی بود و در نهایت چهار شاخص گفتمان علمی، ارزیابی نگارش، نوشتار دانشگاهی، اصول نوشتار دانشگاهی استخراج شد و در پایان با استفاده از ترسیم نقشه علمی مقالات حوزه نوشتار دانشگاهی، شاخص ها و مؤلفه های ارزیابی متون درسی تعیین شدند.

- موسوی و حبیبی (۱۳۹۹) در پژوهشی با عنوان «نقد ساختاری کتاب اخلاق اسلامی؛ مبانی و مفاهیم، براساس مدل های تحلیل محتوای توصیفی و ارتباطی»، به بررسی ساختار کتاب پرداختند. نتایج پژوهش نشان می دهد که کتاب مورد نظر به رغم برتری نسبت به کتاب های هم تراز، به لحاظ ساختاری، دارای اشکالات جدی و نیز فاقد بخش های مهم مقدمه، خلاصه و منابعی برای مطالعه بیشتر است.

روش پژوهش

پژوهش حاضر برای متن کاوی نظرسنجی درمورد کتاب/اخلاق حرفه‌ای در نیم‌سال اول تحصیلی ۱۳۹۹-۱۴۰۰ اجرا شد. روش انجام پژوهش به شرح زیر است:

طراحی پرسش‌نامه برای اطلاع از نظرات مدرسان درس اخلاق حرفه‌ای: ۶ سؤال به صورت طیف لیکرت و یک سؤال به صورت تشریحی چنین است: «درنهایت این کتاب را برای چاپ مجدد چگونه ارزیابی می‌کنید؟ لطفاً پیشنهادات خود را برای اعمال تغییرات در چاپ مجدد مرقوم فرمایید». مراحل به شرح زیر است:

۱. توزیع فایل پرسش‌نامه بین مدرسان درس مربوط و دریافت پاسخ از ۴۷ نفر از مدرسان.

۲. پیش‌پردازش داده‌ها: حذف کلمات بدون بار معنایی که به صورت معمول در سیاهه بازدارنده قرار می‌گیرند (مانند کلماتی چون: برای، همچنین، و این) از میان واژگان حذف شدند (ابوالصدق، ۱۳۹۰).

۳. وزن‌دهی به کلمات متن: پس از پیش‌پردازش داده‌ها، لازم است کلمات شاخص‌گذاری شوند. روش مورد استفاده TF-IDF^۱ است. برای این کار از فرمول شماره ۱ برای محاسبه وزن هر واژه استفاده شده است. فرمول ۱ «فراوانی وزنی کلمه کلیدی» را نشان می‌دهد:

فرمول ۱، فرمول کلی «فراوانی وزنی کلمه کلیدی»:

$$W_{i,j} = tf_{i,j} \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$ تعداد تکرار i از j

df_i تعدادی از متن‌ها که شامل i هستند

N تعداد کل متن‌ها

در یک نگاه، درمی‌یابیم که چطور یک کلمه در یک سند چند بار ظاهر می‌شود. برای محاسبه TF^۲ باید از فرمول ۲ استفاده کرد.

فرمول ۲، محاسبه TF (kw : تعداد کلمات کلیدی و w : تعداد کلمات)

$$TF = \frac{1 + \log(kw)}{\log(w)}$$

1. Term Frequency-Inverse Document Frequency

2. Term Frequency

برای مثال، اگر در یک متن با ۱۰۰۰ لغت کلمه کلیدی مورد نظر ۱۰ بار آمده باشد، چگالی کلمه کلیدی ۱٪ می شود و از طرفی معیار TF ۰/۴۳ می شود (با احتساب لگاریتم در مبنای ۲). حال اگر تعداد تکرار کلمات در صفحه را دو برابر کنیم (یعنی ۲۰ بار) چگالی کلمه کلیدی نیز دو برابر می شود، یعنی ۲٪، اما TF تغییر چندانی نکرده، مقدار آن ۰/۵۳ می شود. همچنین، در فرکانس کلمه، به ندرت از یک کلمه خاص استفاده می شود و در واقع به تنهایی مفید نیست، زیرا به دست آوردن اهمیت کلمات مهم تر از تعداد تکرار استفاده از آن است. برای جلوگیری از این مشکل، از فاکتور IDF^۱ استفاده می شود. این متریک برای درک مقدار واقعی کلمه کلیدی خاص به کمک می آید. IDF نسبت کل اسناد را در مجموعه ای از اسناد که دارای کلمه کلیدی هستند، محاسبه می کند. این متغیر با استفاده از فرمول ۳ محاسبه می شود.

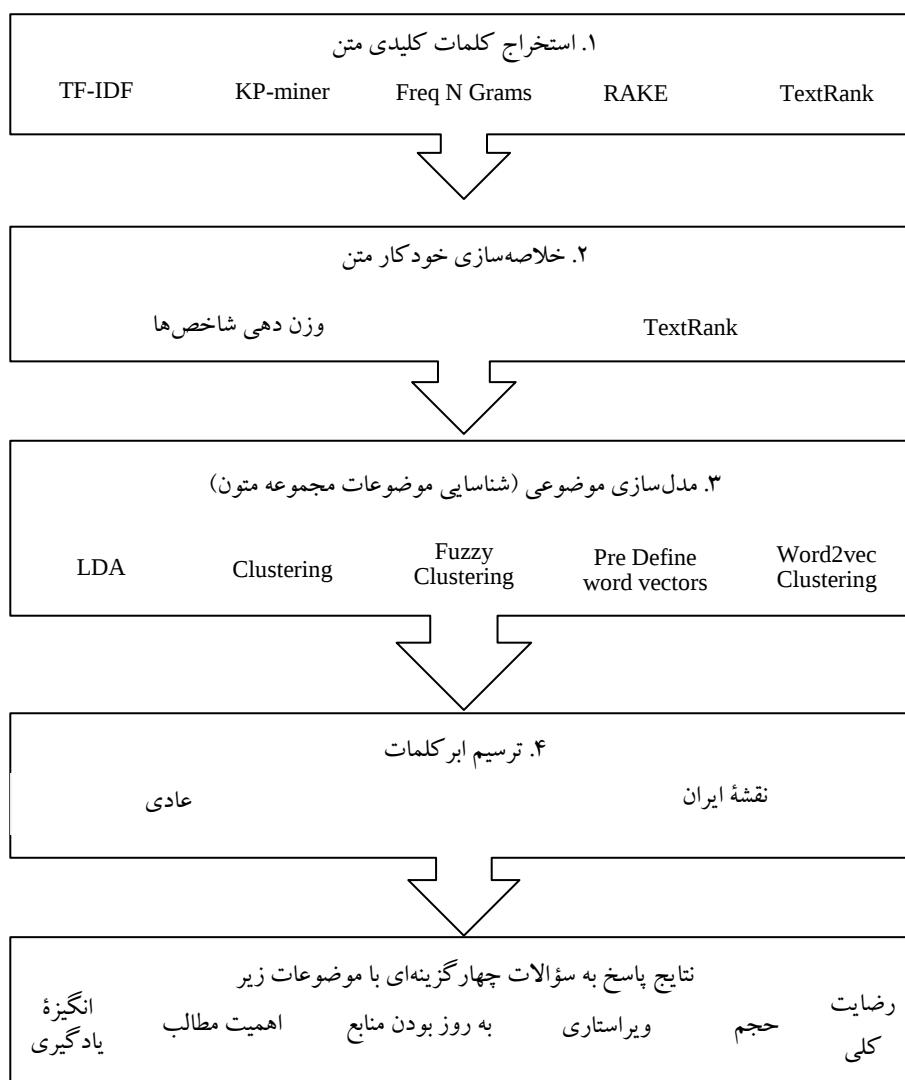
فرمول ۳، محاسبه IDF (td: تعداد کل متن ها و kd: تعداد متن های دارای کلمات کلیدی):

$$IDF = \log\left(1 + \frac{td}{kd}\right)$$

اگر کلمه کلیدی، یک کلمه معمولی باشد، احتمالاً در مقادیر زیادی از اسناد استفاده می شود. در نتیجه، مقدار IDF، آن کم خواهد بود و اگر TF را در آن ضرب کنیم آن مقدار تغییر زیادی نخواهد کرد و برعکس، اگر یک کلمه فقط در یک سند پیدا شود، مقدار IDF آن بزرگ تر خواهد بود که نتیجه کلی TF-IDF هم بزرگ می شود. بر خلاف چگالی کلمه کلیدی که از تکرار کلمه کلیدی خاص در یک صفحه ایجاد شده بود، TF-IDF، به عنوان یک متریک پیشرفته، اهمیت یک کلمه کلیدی خاص در صفحه را نشان می دهد.

یافته های پژوهش

به منظور ارائه دید کلی از پژوهش، مراحل کار به همراه روش های مورد استفاده در شکل ۲ نشان داده شده است. در مرحله اول، استخراج کلمات کلیدی متن انجام می شود؛ سپس متون به صورت خود کار خلاصه می شوند. بعد از خلاصه سازی موضوعات، مجموعه متون با استفاده از مدل سازی موضوعی شناسایی و دسته بندی می شوند و در نهایت ابرکلمات ترسیم می شود. با استفاده از ترسیم ابرکلمات می توان به نکات کلیدی اشاره شده در متن هایی که مدرسان اعلام کرده اند، پرداخت. در نهایت نتایج پاسخ به سؤالات چهارگزینه ای که از مدرسان اخذ شده بود، تحلیل و بررسی می شود.



شکل ۲ مدل مفهومی و روش های مورد استفاده در هر مرحله

۱. استخراج کلمات کلیدی متن

کلمات کلیدی متن از پنج رویکرد TF-IDF، RAKE، KP-Miner، TextRank و FNG استخراج می شوند؛ یافته های آن در جدول ۲ آمده است:

- تعداد کلمات کلیدی در خروجی: ۵

- حداکثر تعداد کلمات هر عبارت کلیدی: تعداد کلمات (اجزاء) هر چندتایی (N-gram) خروجی (بین ۱ تا ۵): ۳
- روش انجام رویکرد استخراج کلمات کلیدی: TF-IDF، KP-miner، Freq N Grams، RAKE و TextRank.

جدول ۲ کلمات کلیدی استخراج شده با استفاده از پنج رویکرد

TextRank	RAKE	Freq N Grams	KP-miner	TF-IDF
کتاب	رابطه اخلاق حرفه‌ای	کتاب	کتاب	کتاب
دانشجو	تدریس دو ترم متوالی	اخلاق	اخلاق	اخلاق
اخلاق	یادگیری مسائل اخلاقی	دانشجو	دانشجو	دانشجو
دانشجویان	چاپ مجدد	مطالب	مطالب	مطالب
نظر	قالب فصلی اضافه	تأثیر گذاری بر اخلاق	تأثیر گذاری بر اخلاق	نظر

۲. خلاصه سازی خود کار متن

این ابزار متن را با چند رویکرد متفاوت به طور خود کار خلاصه می کند. نتیجه حاصل شده در جدول ۳ نشان داده شده است:

- تعداد کلمات متن خلاصه: ۵۰، رویکرد انتخاب جملات مهم: وزن دهی شاخص ها، TextRank.

جدول ۳ نتایج حاصل از خلاصه سازی خود کار متن

TextRank	وزن دهی شاخص ها
کتاب بسیار خوبی است	کتاب بسیار خوبی است
خصوصاً مثال های پیش	مطالعه نیاز به تغییرات زیادی دارد.
سازمان دهنده آن ذهن	مفید است.
دانشجو را برای درس	مطلوب.
دادن قبل از شروع درس	خوب.
باز می کند. فقط بحث	در بند قبل ۸ مورد جهت اصلاح نوشتم.
منشور اخلاقی و	بسیار خوب.
مسئولیت های پنج گانه	خوب است.
بیشتر به آن پرداخته شود	اصلاً صلاح نمی دونم.
و فصل ششم در عوض	باید روان تر بشود.
متعادل تر و روان تر شود.	بعد چاپ گردد.
	باید اصلاحات زیادی صورت گیرد.
	لازم می دانم.

۳. مدل سازی موضوعی (شناسایی موضوعات مجموعه متون)

این ابزار با چند رویکرد متفاوت کلمات کلیدی مربوط به موضوعات مختلف درون مجموعه متون (نظرات، پیام‌ها، اخبار یا ...) را استخراج می‌کند. نتایج مدل سازی موضوعی در جدول ۴ نشان داده شده است.

- تعداد موضوع‌ها: ۳

- رویکرد تشخیص موضوع‌ها: Pre Define word vectors, Fuzzy Clustering, Clustering, LDA Word2vec Clustering.

جدول ۴ نتایج مدل سازی موضوعی با رویکردهای مختلف

LDA	[[«کتاب»، «سؤال»، «فصل»، «سرفصل»، «پیشنهاد»، «ارائه»، «دانشجویان»، «بازنگری»، «نظر»، «تعاریف»], [«کتاب»، «مطالب»، «چاپ»، «نیاز»، «سطح»، «اصلاح»، «کاربردی تر»، «مجدد»، «اضافه»، «جدیدتر»], [«کتاب»، «اخلاق»، «دانشجو»، «مطرح»، «اخلاقی»، «درس»، «مثال‌های»، «لازم»، «تأثیر گذاری»]]
Clustering	[[«کتاب»، «اخلاق»، «دانشجو»، «مطالب»، «مطرح»، «چاپ»، «نیاز»، «پیشنهاد»، «سؤال»، «فصل»], [«می‌رسد»، «مصادیق»، «راحت تر»، «درس»، «نظر»، «دانشجویان»، «عملی»، «جامعه»، «قالب»، «کاربردهای»، «تحقیقاتی»، «نقد»، «حوزه»]]
Fuzzy Clustering	[[«دانشجویان»، «درس»، «نظر»، «اخلاق»، «عملی»، «جامعه»، «قالب»، «کاربردی تر»، «کاربردهای»، «تحقیقاتی»], [«کتاب»، «دانشجو»، «فصل»، «مطالب»، «بازنگری»، «تدریس»، «بررسی»، «پیشنهاد»، «چاپ»، «نیاز»], [«نگارشی»، «نوبسی»، «احترامی»، «نویسنده»، «قائل»، «صادقانه»، «خدمتتان»، «سابقه»، «ده»، «سال»]]
Pre Define word vectors	[[«باحث»، «سرفصل»، «کاربردی تر»، «ویژه»، «مثال‌های»، «جدیدتر»، «دو نرم»، «روان تر»، «روان تر»، «نگارشی»], [«نظر»، «مطالب»، «مطرح»، «نیاز»، «پیشنهاد»، «سطح»، «سؤال»، «ارائه»، «اصلاح»، «اخلاقی»], [«کتاب»، «اخلاق»، «دانشجو»، «درس»، «چاپ»، «دانشجویان»، «فصل»، «تدریس»، «آزمون»، «مطالعه»]]
Word2vec Clustering	[[«درس»، «سطح»، «سؤال»، «گذاری»، «مسائل»، «قابل»، «آزمون»], [«کتاب»، «دانشجو»، «اخلاق»، «مطالب»، «چاپ»، «نیاز»، «نظر»، «پیشنهاد»، «مطرح»، «اصلاح»، «سرفصل»]]

۴. ترسیم ابرکلمات

به منظور ارائه شکل بصری از نظرات متنی و بالابردن قابلیت درک پژوهش، شکل ۳ ابرکلمات را که با زبان پایتون پیاده سازی شده است برای کتاب درسی اخلاق حرفه‌ای نشان می‌دهد.

اگر کتاب روانتر شود و مسائل اخلاقی کسب و کار بیشتر
روان تر و به زود تر شود بهتر است

برای چاپ مجدد کتاب نیاز به اضافه نمودن جدیدیت

خبر تکلام ما دل و جان و قل و لم ی

در بند قیل ۸ مورد جهت اصلاح نوشته‌م چاپ مجدد باید

پیشنهاد می کنم برای طراحی مستقل و کتاب از یک تیم ر

کتاب بسیار خوبی است خصوصا مثال های پیش سازمان دهند

کتاب باید ساده تر و قابل فهم تر نوشته شود

کتاب خوبی و مثال های کاربردی مناسبی برای درک دان

اگر رابطه اخلاق حرفه ای با سایر حوزه های دانش هم م

این سوال در چند خط نمیتوان جواب داد بنده در حل نو

متوسط و کسل کننده و غیر کاربردی بودن آن با وضعیت ک

خوب است

بسیار خوب

مطالب مطرح شده در این کتاب توالی لازم را نداشته و

با احترامی که به نویسنده کتاب قائل هستم ولی باید ص

اصلاح اساسی در متن کتاب گری بسیار پیشنهاد است

ویراستاری و نمونه سوال بعد هر فصل و مثال های کاربر

نیاز به تغییرات اساسی دارد

به نظر می رسد اگر مصادیق بیشتر شود فهم راحت تر است

از ارائه آمار و اطلاعات جامعه قائل شده باید در کتا

و از کثرت پردازی خودداری نمود چراکه بزرگان دین ما

عرض کردم مطالب مفیدی ساده تر و کم حجم تر شود مای

این کتاب را پیشنهاد نمی کنم

ضرورت بازنگری کلی و تجدید نظر در سرفصل ها برای ایجا

اصلاح پیشنهاد

اگرچه این کتاب برای آزمون تشریحی ایده ال است ؛ وقت

نیاز به تغییرات زیادی دارد

نیاز به تغییرات زیادی دارد

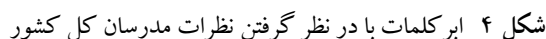
خوب

بسیار خوب

شکل ۳ ابرکلمات پیاده‌سازی شده با زبان پایتون

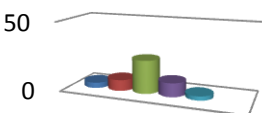
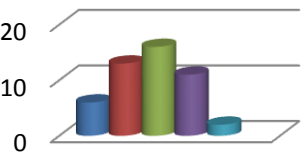
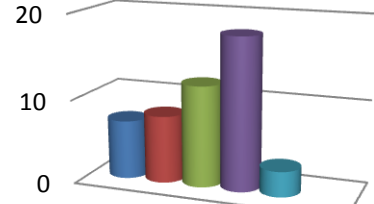
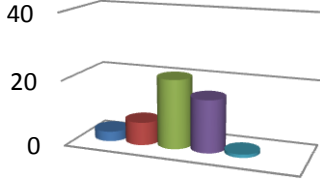
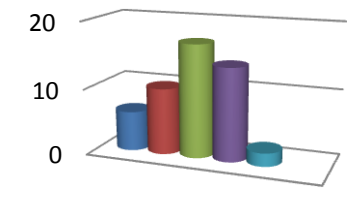
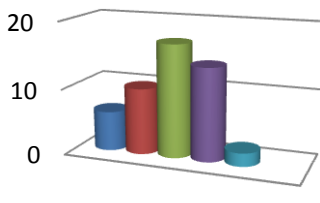
شکل ۳ ابرکلمات پیاده سازی شده با زبان پایتون را نشان می دهد. در این شکل براساس ضوابط معرفی شده در بخش های قبل، کلماتی دارای بیشترین تکرار و تأکید از سوی مدرسان تدریس کننده این کتابها مشخص شدند. کلمات دارای تکرار و تأکید بیشتر با فونت درشت تر و کلمات دارای تأکید کمتر با فونت ریزتر نوشته شده اند. با استفاده از این تصویر می توان با یک نگاه متوجه نکات کلیدی و مهم از نظر مدرسان شد. درواقع، خلاصه سازی دیدگاه های مفصل مدرسان را می توان با ابرکلمات انجام داد.

شکل ۴ ابرکلمات را با در نظر گرفتن نظرات مدرسان کل کشور نشان می دهد. با توجه به اینکه در این شکل از روش های دقیق تر و داده های بیشتر استفاده شده و نظرات مدرسان کل کشور در خصوص کتاب/اخلاق حرفه ای مورد بررسی قرار گرفته شده است، از دقت بالاتری برخوردار است و کار خلاصه سازی را به صورت مفیدتری انجام داده است. ابرکلمات در قالب شکل نمایش داده می شود. در این شکل برای نشان دادن اینکه نظرات همهٔ مدرسان کل کشور اخذ شده است، شکل در قالب نقشه ارائه شده است. البته می توان آن را در شکل مستطیلی نیز نشان داد و در نتیجه کار تغییری حاصل نمی شود.



با توجه به این که پرسش‌نامه اجرا شده دارای یک سری سؤالات چهارگزینه‌ای و سپس سؤال متنی بود، برای روشن‌تر شدن ابعاد یافته‌های پژوهش، نمودارهای ترسیم‌شده از نتایج بررسی این سؤالات ارائه شده است.

شکل ۵ پاسخ به سؤالات چهارگزینه‌ای را نشان می‌دهد؛ یعنی در پاسخ به سؤال میزان ایجاد انگیزه و شوق یادگیری و سؤال میزان اهمیت و بدیع بودن یافته‌های کتاب، نظر اکثر کاربران در حد متوسط است و پاسخ‌ها حالت نرمالی دارد. اما در سایر سؤالات یعنی به روز بودن منابع در کتاب، ویراستاری کتاب و میزان رضایت و میزان حجم کتاب نظرات حالت نرمال کمتری دارند و بیشتر در حد متوسط و زیاد هستند. از لحاظ ویراستاری و نگارش و روان بودن ارائه مطالب، نظر اکثر مدرسان بر وضعیت مناسب بودن زیاد است. همان‌گونه که در نمودارهای شکل ۵ نیز مشخص است، پاسخ سؤالات چهارگزینه‌ای، کتاب را در کلیه بخش‌ها، متوسط ارزیابی نموده، صرفاً «ویراستاری، نگارش و روان بودن ارائه مطالب» را در حد بالایی می‌داند.

 50 0 ۱ خیلی کم ۲ کم ۳ متوسط ۴ زیاد ۵ خیلی زیاد (blank)	 20 10 0 Total ۱ خیلی کم ۲ کم ۳ متوسط ۴ زیاد ۵ خیلی زیاد
پاسخ سؤال «میزان اهمیت و بدیع بودن یافته‌های کتاب»	پاسخ سؤال «میزان ایجاد انگیزه و شوق یادگیری»
 20 10 0 ۱ خیلی کم ۲ کم ۳ متوسط ۴ زیاد ۵ خیلی زیاد	 40 20 0 ۱ خیلی کم ۲ کم ۳ متوسط ۴ زیاد ۵ خیلی زیاد
پاسخ سؤال «ویراستاری، نگارش و روان بودن ارائه مطالب»	پاسخ سؤال «به روز بودن منابع در کتاب را چگونه دیدید؟»
 20 10 0 ۱ خیلی کم ۲ کم ۳ متوسط ۴ زیاد ۵ خیلی زیاد	 20 10 0 ۱ خیلی کم ۲ کم ۳ متوسط ۴ زیاد ۵ خیلی زیاد
پاسخ سؤال «میزان رضایت در مجموع»	پاسخ سؤال «میزان حجم کتاب جهت پوشش سرفصل»

شکل ۵ پاسخ سؤالات چهارگزینه‌ای به تفکیک سؤال

۶. راستی آزمایی یافته‌ها

پس از متن کاوی و خوشه‌بندی، لازم است روش به کار گرفته شده برای متن کاوی مورد ارزیابی قرار گیرد. مهم‌ترین بخش متن کاوی، اعتبارسنجی به منظور بازیابی اطلاعات و داده‌های متنی است که می‌توان آن را هدف نهایی این کار دانست. هدف اصلی در بازیابی اطلاعات، جست‌وجوی متون نزدیک به پرس‌وجوی کاربر است. هدف متنی است که می‌توان آن را هدف نهایی این کار دانست. بازیابی اطلاعات و داده‌های متنی، مهم‌ترین قسمت فرایند متن کاوی است و می‌توان آن را هدف نهایی از انجام همه مراحل قبلی دانست (ابوالصدق، ۱۳۹۰):

گفتنی است در این پژوهش برای بخش تجزیه و تحلیل داده‌ها، از نرم‌افزار متن کاوی فارسی‌یار و کدنویسی پایتون استفاده شده است. برای صحت‌سنجی نیز، پس از ارائه نظرات به خبرگان این حوزه، نتایج مورد تأیید قرار گرفت.

بحث و نتیجه‌گیری

در این پژوهش برای بررسی و نظرات متنی مدرسان در خصوص کتاب/اخلاق حرفه‌ای (صبحی قراملکی، ۱۳۹۸) با روش‌های مختلف کلمات کلیدی متن‌ها استخراج گردید و ابرکلمات کلیدی تشکیل شد. یافته‌ها حاکی از آن است که روش مدل‌سازی موضوعی که جهت شناسایی موضوعات مجموعه متون مورد استفاده قرار گرفت و روش ترسیم ابرکلمات که شکل بصری از نظرات متنی را ارائه می‌دادند نسبت به سایر روش‌ها در این پژوهش مناسب‌تر بودند. طبق نتایج این تحقیق در زمان داشتن نتایج خروجی سؤالات به صورت متون، برای تحلیل، خلاصه‌سازی و دسته‌بندی سریع و دقیق آن‌ها می‌توانیم از ابزارهای متن کاوی استفاده کنیم. نمایش خلاصه نظرات به صورت متن می‌تواند ابزاری برای تصمیم‌گیری صحیح مدیران باشد. در این مقاله روش‌های متعدد متن کاوی مورد بررسی قرار گرفت و کار برای تحلیل نظرات مدرسانی که کتاب/اخلاق حرفه‌ای را تدریس نموده‌اند، پیاده‌سازی شد. از این روش می‌توان برای تحلیل نظرات سایر کتاب‌ها یا تحلیل نظرات متن‌های موجود در گروه‌های شبکه‌های مجازی نیز استفاده کرد. با استفاده از نتایج این پژوهش و تکرار آن برای سایر کتاب‌های آزمون متمرکز، انتشارات دانشگاه می‌تواند در تصمیم‌گیری مناسب برای تغییر کتاب، تجدید چاپ کتاب با بازنویسی یا

تجدید چاپ کتاب بدون تغییرات اقدام نماید. درخصوص کتاب *اخلاق حرفه‌ای* پیشنهاد تجدید چاپ با بازنویسی و ویراستاری مجدد است. نیز نتایج نشان می‌دهد که مطالب کتاب نسبتاً خوب است، اما لازم است مورد بازنگری و ویراستاری قرار گیرد و توالی مطالب اصلاح شده و منابع جدیدی به آن اضافه و به صورت روان‌تر و با مثال‌های کاربردی بیشتری بازنویسی شود.

تشکر و قدردانی. در پایان از حمایت‌های بی‌دریغ و همکاری تنگاتنگ ریاست محترم دانشگاه (آقای دکتر محمدحسین امید) و نیز معاون نظارت و سنجش دانشگاه (آقای دکتر محسن محمدزاده) در تمامی مراحل انجام این پژوهش، نهایت تشکر را دارد.

منابع

- ابوالصدق، سارا (۱۳۹۰). کاربرد متن کاوی در بررسی ادبیات مهندسی صنایع، یزد: دانشگاه یزد.
- پرچمی جلال، مجید، شکیبا زاهد، هادی و محمدامین قادری (۱۳۹۹). «مقایسه تطبیقی ابزار پرسش‌نامه و ابزار متن کاوی در شناسایی معیارهای مؤثر بر شرکت/عدم شرکت مشاوران در مناقصات بین‌المللی صنعت آب - نمونه موردی شرکت مهندسی مشاور»، *نشریه آب‌کشی و زهکشی ایران*، شماره ۲، جلد ۱۴، ص ۴۹۸-۵۱۰.
- رضی، احمد (۱۳۸۸). «شاخص ارزیابی و نقد کتاب‌های درسی دانشگاهی»، *عیار*، سال چهاردهم، دوره ۱۴، شماره ۲۱، ص ۲۱-۳۰.
- صبحی قراملکی، ناصر (۱۳۹۸). *اخلاق حرفه‌ای*، تهران: نشر دانشگاهی - علمی کاربردی.
- عباسی، محمدرضا و علی اصغر شیره‌پزآرانی (۱۳۹۰). «سیاست‌گذاری کیفیت آموزش عالی: چالش‌ها و چشم‌اندازها»، *پنجمین همایش ارزیابی کیفیت در نظام دانشگاهی*، تهران: دانشگاه تهران (پردیس دانشکده‌های فنی).
- عبدی نسب، علیرضا، مؤمنی، عصمت و سیدمهدی طاهری (۱۳۹۷). «طراحی الگوی سیستم نرم‌افزاری اعتباریابی کتاب‌های دانشگاهی مبتنی بر متن کاوی»، *پژوهش و نگارش کتب دانشگاهی*، سال بیست و دوم، شماره ۴۲، ص ۳۸-۷۷.
- عظیمی همت، منیره و فاطمه شمس عزت (۱۳۹۴). «مروری بر متن کاوی متون فارسی»، *دومین کنفرانس بین‌المللی و سومین همایش ملی کاربرد فناوری‌های نوین در علوم مهندسی*، مشهد: دانشگاه تربت حیدریه، دانشگاه فردوسی مشهد.
- کاردان، احمد آقا و مینا کیهانی‌نژاد (۱۳۹۱). «ارائه مدلی برای استخراج اطلاعات از مستندات متنی، مبتنی بر متن کاوی در حوزه یادگیری الکترونیکی»، *فناوری اطلاعات و ارتباطات ایران*، ۴ (۱۲-۱۱)، ص ۴۷-۵۷.
- کاردان، علی محمد (۱۳۸۲). «نقش کتاب‌های دانشگاهی و ویژگی‌های آن»، *سخن سمت*، دوره ۸، شماره ۱، ص ۱۳-۲۰.

موسوی، سمیه سادات و رضا حبیبی (۱۳۹۹). «نقد ساختاری کتاب اخلاق اسلامی؛ مبانی و مفاهیم، براساس مدل‌های تحلیل محتوای توصیفی و ارتباطی»، پژوهش و نگارش کتب دانشگاهی، دوره ۲۴، شماره ۴۶، ص ۱۴۶-۱۶۵.

- Aldowah H., Al-Samarraie H. & W. M. Fauzy (2019). "Educational data mining and learning analytics for 21st century higher education: A review and synthesis", *Telematics and Informatics Elsevier Ltd*, 37: 13-49, 01-Apr.
- Fan, W., & A. Bifet (2013). "Mining big data: Current status, and forecast to the future", *ACM SIGKDD Explorations Newsletter*, 14(2): 1-5.
- Hussain, S., Dahan, N.A., Ba-Alwib, F.M. & N. Ribata (2018). "Educational data mining and analysis of students 'academic performance using WEKA", *Indonesian Journal of Electrical Engineering and computer science*, 9 (2).
- Jusoh, S. & H. M. Alfawareh (2009). "Agent-based knowledge mining architecture", *Proceedings of the 2009 International Conference on Computer Engineering and Applications, IACSIT*, 602-606.
- Miranda Santo, Marcio de, Gilda Massari Coelho, Dalci Maria dos Santos & Lélío Fellows Filho (2006). "Text mining as a valuable tool in foresight exercises: A study on nanotechnology", *Technological Forecasting and Social Change*, 73 (8): 1013-27.
- Null, W. (2016). *Curriculum: From Theory to Practice*, Rowman & Littlefield.
- Ojala, T., Cruz, E., Costa, J. & J. Korhonen (2008). "Mobile lecture interaction: Making technology and learning click", *IADIS International Conference Mobile Learning*, 7: 8-15.
- Rehmani, A. (2003). "Impact of public examination system on teaching and learning in pakistan. retrieved december 24", 2010 from <http://www.aku.edu/AKUEB/pdfs/pubexam.pdf>.
- Shettar R, Shobha G. (2007). "Survey on mining in semi-structured data", *International Journal of Computer Science and Network Security*, 7(8): 226-231.
- Singh, Neha, Clark Hu, & W. S. Roehl (2007). "Text mining a decade of progress in hospitality human resource management research: Identifying emerging thematic development", *International Journal of Hospitality Management*, 26 (1): 131-47.
- Wu, C., Mai, F., & X. Li (2021). "The effect of content depth and deviation on online review helpfulness: Evidence from double-hurdle model", *Information and Management*, 58(2), doi: 10.1016/j.im.2020.103408
- Xu, Z., Frankwick, G. L., & E. Ramirez (2016). "Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective". *Journal of Business Research*, 69(5): 1562-1566.