



# دانشگاه علوم پزشکی کرمان

دانشکده مدیریت و اطلاع‌رسانی پزشکی

پایان‌نامه مقطع کارشناسی ارشد کتابداری و اطلاع‌رسانی پزشکی

عنوان:

متن‌کاوی مقالات کتابداری و اطلاع‌رسانی نمایه شده در پایگاه اطلاعاتی وب‌آوساینس

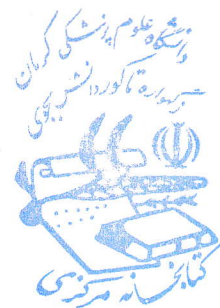
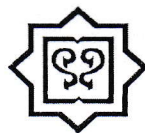
به‌منظور کشف ساختار درونی رشته و اعتبارسنجی آن

توسط: عاطفه ذوالفقارنسب حاجی‌زاده

استاد راهنما: دکتر فاطمه معتمدی

استاد مشاور: دکتر محمد اعظمی

سال تحصیلی ۱۳۹۵-۱۳۹۶



**Kerman University of Medical Sciences**

**Faculty of Management and Information Sciences**

In Partial Fulfillment of the Requirements for the Degree

(MSc)

**Title:**

**Text mining library and information science articles indexed in the Web of Science for discovering internal structure of this field and its validation**

**By: Atefeh Zolfagharnasab Hajizadeh**

**Supervisor:**

**Fatemeh Motamedi (Ph.D)**

**Advisor:**

**Mohammad Azami (Ph.D)**

**Year: 2018**

**مقدمه:** مقالات علمی به دلیل ویژگی‌های خاصی که دارند، از اهمیت بسزایی برخوردارند. به‌منظور بازیابی مطلوب محتوای اطلاعاتی آن‌ها و جلوگیری از بروز مشکلاتی از جمله ریزش کاذب یا بازیابی مدارک نامرتب و همچنین پیشگیری از اتلاف وقت کاربران استفاده از روش‌ها و فنون تحلیلی مناسب جهت کاوش و پردازش متون ذخیره‌شده، بسیار ضروری به نظر می‌رسد. متن کاوی شاخه‌ای از داده‌کاوی است که به دلیل تلاش در جهت به دست آوردن دانش مفید از میان داده‌های متنی، اهمیتی دوچندان یافته است. لذا این مطالعه با هدف متن‌کاوی مقالات کتابداری و اطلاع‌رسانی نمایه شده در پایگاه اطلاعاتی وب آو ساینس به‌منظور کشف ساختار درونی رشته و اعتبارسنجی آن انجام گرفته است.

**روش کار:** این پژوهش توصیفی - مقطعی و از نوع کاربردی است که با استفاده از الگوریتم‌های خوشه‌بندی در متن کاوی صورت گرفته است. ۵۱۹۸۰ رکورد بازیابی شده با استفاده از روش سرشماری از پایگاه «وب آو ساینس» در بازه زمانی سال‌های ۱۹۸۳ تا ۲۰۱۵ با استفاده از نرم‌افزار هیست سایت جمع‌آوری گردید. به‌منظور انجام عملیات متن‌کاوی، در مرحله اول داده‌های جمع‌آوری شده ابتدا مورد پیش‌پردازش قرار گرفتند و در انتهای این مرحله، فرهنگ لغتی از واژگان رشته کتابداری و اطلاع‌رسانی ایجاد گردید. در مرحله بعد با به‌کارگیری نرم‌افزار ریپیدماینر، الگوهای حاوی اطلاعات با استفاده از الگوریتم‌های خوشه‌بندی k-means به دست آمدند و در مرحله پایانی، نتایج اطلاعات استخراج‌شده و ویژگی‌های آن‌ها با استفاده از دو معیار دقت و بازیابی مورد ارزیابی و بررسی قرار گرفتند.

**یافته‌ها:** نتایج حاصل از اعمال الگوریتم خوشه‌بندی k-means در بستر متن‌کاوی ایجاد ۷ خوشه ابتدایی در این حوزه بود که پس از بررسی و ارزیابی لغات تشکیل‌دهنده هر خوشه با عنوان مقاله و کلیدواژه‌های آن، ۵ خوشه با نام‌های «اینترنت و فناوری»، «ژورنال‌یسم و علم‌سنجی»، «بازیابی اطلاعات»، «علم کتابداری و اطلاع‌رسانی» و «خدمات کتابخانه‌ای» تشکیل شد. به‌منظور ارزیابی خوشه‌های تشکیل شده نیز از دو معیار دقت و بازیابی که از معیارهای بازیابی اطلاعات هستند، استفاده شد که معیار دقت ۰,۸۵ و معیار بازیابی ۰,۷۱ به دست آمد که نشان‌دهنده مطلوبیت خوشه‌های به‌دست‌آمده بود.

**نتیجه‌گیری:** بررسی‌ها بر روی خوشه‌های تشکیل شده، نشان‌دهنده این بود که حوزه‌های مرتبط با فناوری و اینترنت بیشتر مورد توجه و حوزه‌های مرتبط با دانش نظری این رشته در حاشیه قرار گرفته‌اند. همچنین کلماتی در هر زیر گروه حوزه رشته کتابداری و اطلاع‌رسانی، کاربرد بیشتری دارند، مشخص شده‌اند و با توجه به همین موضوع، پایگاه‌های اطلاعاتی بزرگ، با بهره‌گیری از این کلمات، می‌توانند کاربران را در دستیابی به اطلاعات دقیق و مورد نیاز در مدت زمانی

بسیار کوتاه، یاری کنند. علاوه بر این، سیاست‌گذاران و محققین حوزه کتابداری و اطلاع‌رسانی ضمن شناسایی مباحث  
پرکاربرد و همچنین حوزه‌های کمتر مورد توجه در این رشته، می‌توانند در اعتلا و ارتقای این حوزه تأثیرگذار باشند.

**کلمات کلیدی:** متن‌کاوی، خوشه‌بندی، الگوریتم k-means، داده‌کاوی، مقالات کتابداری و اطلاع‌رسانی، بازیابی

اطلاعات، اعتبارسنجی، پایگاه وب آو ساینس



## **Abstract**

**Introduction:** Scientific articles have great importance due to their certain features. In order to optimal retrieve of the information content and prevent the occurrence of problems including false drop or irrelevant documents retrieval as well as preventing users from wasting time using appropriate analytical techniques to explore and process the stored texts, seems very necessary. The text mining is a branch of data mining because of the effort to gain useful knowledge has taken a great deal of importance. Therefore, this study was conducted with the aim of text mining of library articles and information indexed in the Web of Science database in order to validate it and discover the internal structure of the field.

**Methods:** This is a descriptive cross-sectional and applied research which have been taken by use of clustering algorithms in text mining. 51980 records were retrieved using the census method from the Web of Science database during the period 1983 to 2015 by use of HistCite software collected. In order to perform text mining operations, in the first step, the collected data initially were preprocessing and at the end of this stage, the dictionary of library and information field was created. In the next step, using the RapidMiner software, patterns containing information were obtained by use of k-means clustering algorithms and in the final stage, the results of the extracted information and their features were evaluated by use of two precision and recovery criteria.

**Results:** The results of applying the k-means clustering algorithm in the context of text mining creating the 7 primitive clusters in this domain after examining and evaluating the vocabulary of each cluster with the title of the article and its keywords, five clusters were named "Internet and Technology", "Journalism and Scientometrics", "Information Retrieval", "Library and Information Science" and "Library Services" were formed. In order to evaluate the formed clusters, two precision and recovery criteria, which are the criteria for recovery data were used, with a precision rating of 0.85 and a recall criterion of 0.71, indicating the desirability of the clusters obtained.

**Conclusion:** The studies on the clusters showed that the fields of Internet and Technology were more focused and the fields related to theoretical knowledge of this field were marginalized. Also, the words in each subsection of the field of Library and Information Science were more applicable and regarding this subject, large databases, by use of these words, can help users achieve detailed information and requirements in a very short period of time. In addition, policy makers and researchers in the field of library and information, while identifying the most widely used topics and as well as less attention areas in this field, can influence the promote and upgrade of this area.

**Keywords:** Text mining, Clustering, k-means algorithm, Data mining, Library and Information articles, Information retrieval, Web of Science database



دانشگاه علوم پزشکی کرمان  
تحصیلات تکمیلی دانشگاه

بسمه تعالی

صور تجلسه دفاع از پایان نامه

تاریخ .....

شماره .....

پیوست .....

جلسه دفاعیه پایان نامه تحصیلی خانم عاطفه ذوالفقار نسب حاجی زاده دانشجوی کارشناسی ارشد رشته کتابداری و اطلاع رسانی پزشکی

دانشگاه علوم پزشکی کرمان تحت عنوان: متن کاوی مقالات کتابداری و اطلاع رسانی نمایه شده در پایگاه وب آوساینس به منظور کشف ساختار درونی رشته و اعتبارسنجی آن ساعت ۹:۱۵ روز شنبه مورخ ۹۶/۱۲/۰۵ با حضور اعضای محترم هیات داوران متشکل از:

امضا	نام و نام خانوادگی	سمت
	خانم دکتر معتمدی	الف: استاد(ان) راهنما
	آقای دکتر اعظمی	ب: استاد(ان) مشاور
	آقای طالبیان	ج: عضو هیات داوران (داخلی)
	خانم دکتر صابر ماهانی	د: عضو هیات داوران (خارجی)
	آقای دکتر یزدی فیض آبادی	ه: نماینده تحصیلات تکمیلی

تشکیل گردید و ضمن ارزیابی به شرح پیوست با درجه بسیار خوب و نمره ۱۷/۸۸ مورد تأیید قرار گرفت.

دانشگاه علوم پزشکی کرمان  
معاونت آموزش  
معاون آموزشی  
معاون امور آموزشی