

The Analysis of the Distribution and Focus of Keywords in Theses and Dissertations and Compliance with Descriptors, Title, and Abstract

Ashkan Khatir

PhD Candidate in Information Technology Engineering;
Iranian Research Institute for Information Science and Technology
(IranDoc); Iran Email: khatir@students.irandoc.ac.ir

Soheil Ganjefar*

Visiting Lecturer; Iranian Research Institute for Information
Science and Technology (IranDoc); Professor of Electrical
Engineering; Department of Electrical Engineering;
Faculty of Engineering; Bu-Ali Sina University; Iran;
Email: s_ganjefar@basu.ac.ir

Received: 18, Jan. 2017 Accepted: 16, Sep. 2017

Abstract: Index terms provided by authors and professional indexers are used in traditional information retrieval schemes. However, abstracts ideally contain the core message of a document. This can potentially give us the opportunities to use the abstracts to automatically extract index terms. This work is an effort to increase the accuracy of keyword extraction mechanism by adding a temporal weighting to candidate. In addition, this work can be used to research trend analysis and shows where the ongoing research is headed in Iranian Theses and Dissertations (TDs). To achieve the aforementioned objectives, we studied on more than 500 samples in different engineering research area from 50 different universities 1) the correlation between the authors and professional indexers keywords. We observed only 8% similarity between these two indices. 2) We studied the correlation between the index terms and words in abstract and title. We found that 40% of author keywords are extracted from first 20% of the abstract (This figure changes to 45% for professional indexer) and 24% from the second 20% (19% from the next 20%) This finding can be further used to narrow down the input dimensions for the various machine learning schemes for automatic keyword extraction. 3) Using some classification schemes it can be perceived that the most of the ongoing research in Iran is headed toward neural network and optimization.

Keywords: Indexing, Descriptors, Keyword Distribution, Research Focus Area

* Corresponding Author

**Iranian Journal of
Information
Processing and
Management**

**Iranian Research Institute
for Information Science and Technology
(IranDoc)**

ISSN 2251-8223

eISSN 2251-8231

Indexed by SCOPUS, ISC, & LISTA

Vol. 34 | No. 1 | pp. 411-428

Autumn 2018



تحلیل توزیع و تمرکز کلیدواژه‌های پایان‌نامه‌ها و رساله‌ها و میزان تطابق با توصیفگرها، عنوان، و چکیده

اشکان خطیر

دانشجوی دکتری مهندسی فناوری اطلاعات؛
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)؛
khatir@studets.irandoc.ac.ir

سهیل گنج‌فر

دکتری برق کنترل؛ استاد مهمان؛ پژوهشگاه علوم و
فناوری اطلاعات ایران (ایرانداک)؛ استادتمام دانشگاه
بوعلی سینا؛ دانشکده مهندسی؛ گروه برق؛
پدیده‌آور رابط S_ganjefar@basu.ac.ir



دریافت: ۱۳۹۵/۱/۲۹ | پذیرش: ۱۳۹۶/۰۶/۲۵ | مقاله برای اصلاح به مدت ۲۱ روز نزد پدیدآوران بوده است.

چکیده: نمایه‌ها و چکیده‌های یک متن خلاصه‌ای از متن را در اختیار خواننده قرار می‌دهند. پس، می‌توان از آن‌ها برای درک سریع و بازبایی سند استفاده کرد. از آنجا که بخش عمده‌ای از فعالیت‌های علمی فارسی در کشور ایران را پارساها تشکیل می‌دهند، در این پژوهش نمایه‌سازی پارساها از دو دیدگاه نویسنده پارسا و نمایه‌ساز حرفه‌ای مورد بررسی قرار خواهد گرفت. سپس، این نمایه‌ها با عنوان پارساها مورد بررسی قرار می‌گیرند تا میزان انطباق با کلیدواژه‌های عنوانی به دست آید. از سوی دیگر، با بررسی کلی مجموعه‌ای از نمایه‌ها و چکیده‌ها علاوه بر قابلیت بهبود در بازبایی اطلاعات برای محقق، حوزه فعالیت‌هایی که بیشتر اسناد بر روی آن تمرکز کرده‌اند، مشخص می‌شود. علاوه بر آن، در این پژوهش وجود نمایه‌ها و توزیع آن‌ها در چکیده بررسی می‌شود. از توزیع کلیدواژه‌ها در چکیده می‌توان در استخراج خودکار کلیدواژه‌ها از چکیده پارساها در کارهای آتی استفاده نمود. این پژوهش بر روی پارساهای موجود در پایگاه داده «پژوهشگاه علوم و فناوری اطلاعات ایران» که منبع گردآوری پارساهای فارسی است، انجام شده است. روش پژوهش به این صورت است که بعد از گردآوری داده‌ها، پارساهایی که اطلاعات کافی ندارند، پالایش شده و بقیه «پارساها» توسط برنامه‌ای که برای پردازش متن چکیده و نمایه‌های پارساها نوشته‌ایم، مورد تحلیل

فصلنامه | علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
(ایرانداک)

شاپا (چاپی) ۸۲۲۳-۲۲۵۱

شاپا (الکترونیکی) ۸۳۳۱-۲۲۵۱

نمایه در SCOPUS، LISTA، ISC، و

jipm.irandoc.ac.ir

دوره ۳۴ | شماره ۱ | صص ۴۱۱-۴۲۸

پاییز ۱۳۹۷



قرار خواهند گرفت. سپس، اطلاعات به دست آمده با استفاده از آمار توصیفی شرح داده خواهد شد. بررسی انجام شده در این پژوهش نشان داده است که عموماً نمایه‌های انتخاب شده (بیش از ۶۰ درصد) توسط نویسنده و نمایه‌ساز حرفه‌ای از ۴۰ درصد ابتدایی چکیده انتخاب شده‌اند. دیگر تحلیل‌های آماری این پژوهش نشان می‌دهند که میزان انطباق بین توصیفگرها و کلیدواژه‌ها ۸ درصد است. این اختلاف نشان‌دهنده میزان تفاوت نظر زیاد بین نویسندگان پارساها و نمایه‌سازان است. با بهره‌گیری از این اختلاف و با تجمیع کلمات و غنی کردن کلیدواژه‌های سیستم بازیابی اطلاعات می‌توان در بهبود بازیابی اطلاعات نیز استفاده کرد.

کلیدواژه‌ها: نمایه‌سازی، کلیدواژه، توصیفگر، توزیع کلیدواژه، تمرکز فعالیت پژوهشی

۱. مقدمه

امروزه، به خاطر تولید حجم زیاد مستندات و بزرگ‌شدن پایگاه داده‌های نیمه‌ساخت یافته و غیرساخت یافته، امکان بررسی تک‌تک این اسنادها برای پیدا کردن اطلاعات مورد نیاز توسط کاربر و محقق امری بسیار دشوار، زمان‌بر و در عمل، نشدنی است. حتی در محیط وب، با آن که موتورهای جست‌وجو ابزارهای ارزشمندی را برای بازیابی اطلاعات در اختیار کاربر قرار می‌دهند، اما در بسیاری از موارد ممکن است نتوانند نتایج مؤثری ارائه نمایند (Carpineto et al. 2009). در بعضی موارد موتورهای جست‌وجو افراد را به صفحاتی هدایت می‌کنند که از نظر معنایی ارتباط زیادی با کلمات مورد جست‌وجو ندارند. بنابراین، برای دستیابی به نتایج مؤثر، روشی مورد نیاز است تا کاربر نیاز نداشته باشد تمام صفحه وب، سند و یا بخش بزرگی از آن را مورد مطالعه و بررسی قرار دهد، بلکه بتواند با صرف کمترین زمان، اسنادی را که ارتباط بیشتری با موضوع و محتوای کاری او دارد، پیدا کند. کلیدواژه، کلمه و یا مجموعه کلماتی است که یک مفهوم عمیق و خلاصه‌ای از متن را ارائه می‌کند و نیز منجر به بازیابی اطلاعات خواهد شد (Hulth 2003; Rose et al. 2010). در این مقاله منظور ما از توصیفگر، کلمه یا مجموعه کلماتی است که توسط نمایه‌ساز برای نشان دادن محتوای متن برای نمایه کردن آن متن انتخاب می‌شود. نمایه نیز کلمه یا کلماتی است که توسط نمایه‌ساز و یا نویسنده متن برای نمایه کردن و نشان دادن محتوای سند مورد انتخاب قرار می‌گیرد. بنابراین، می‌توان با انتخاب کلیدواژه‌ها و توصیفگرهای مناسب سیستمی با بازیابی مناسب‌تر در اختیار داشت. در این پژوهش نیز به بررسی و تجزیه و تحلیل آن‌ها خواهیم پرداخت.

انتخاب نمایه‌های مناسب علاوه بر این که در درک خواننده از متن به او کمک می‌کنند، کاربردهای فراوان دیگری نیز دارند. استخراج کلیدواژه‌ها در بعضی روش‌های خلاصه‌سازی متن به منظور استخراج ویژگی‌های متن و وزن‌دهی به کلمات، و برای تعیین جملات پراهمیت مورد استفاده قرار گرفته است (برای مثال نگاه کنید به (Chen, Han & Litvak & Last 2008; Kyoomarsi et al. 2008; Gupta & Lehal 2010; Chen 2002)). از کاربردهای دیگر استخراج کلیدواژه‌ها می‌توان به طبقه‌بندی خودکار متن (برای مثال (Zhang 2008)) و نظر کاوی^۱ به عنوان استراتژی جست‌وجو اشاره کرد (برای مثال (Khan, Baharudin & Khan 2009)).

همان‌طور که ملاحظه کردیم، انتخاب کلیدواژه، کاربردی و حائز اهمیت است. بنابراین، تحلیل و بررسی نمایه‌ها که از دید نویسنده و نمایه‌ساز برای یک سند انتخاب شده، ارزشمند است. برای این که یک سیستم بازیابی اطلاعات مناسبی داشته باشیم، نیاز به درک درست و مناسبی از روابط نمایه‌ها و توزیع آن‌ها در عنوان و چکیده سند هستیم. این اطلاعات می‌تواند به طراحی سیستمی که به‌طور خودکار کلیدواژه‌ها را از متن انتخاب کند، یاری رساند. از طرفی دیگر، از آنجا که انتخاب نامناسب کلیدواژه‌ها مشکلاتی در بازیابی اطلاعات به وجود می‌آورد، توصیه‌هایی مبنی بر عدم انتخاب کلیدواژه‌ها از عنوان نیز شده است تا در بازیابی اطلاعات بیشتر کمک کند. از این رو، در این تحقیق به بررسی وجود یا نبود نمایه‌های مورد استفاده نویسنده و نمایه‌سازان در عنوان پارسا نیز پرداخته خواهد شد.

پژوهش‌هایی نیز در خصوص میزان انطباق سرعنوان‌های موضوعی با عناوین کتاب‌های منتشرشده در «کتابخانه ملی ایران» انجام شده که نشان‌دهنده میزان انطباق ۹۹ درصدی آن‌هاست (اباذری و پالیزوانی ۱۳۹۱). «اسکولتز» در پژوهش خود نشان داده است که میزان انطباق کلیدواژه‌های نویسنده به کلیدواژه‌های کنترل‌شده بسیار بیشتر از کلیدواژه‌هایی هستند که از عنوان استخراج شده‌اند (Schultz, Schultz & Orr 1965). «استریدر» نیز نشان داده است که کلیدواژه‌هایی که توسط نویسنده مشخص می‌شوند اطلاعات بسیار باارزش، اضافه بر نمایه‌ساز در اختیار قرار می‌دهند (Strader 2011). در سال ۲۰۰۵، پژوهشی که توسط «انصاری» بر روی پایان‌نامه‌های پزشکی فارسی انجام شد، نشان می‌دهد که

1. opinion mining

بیش از ۷۰ درصد کلیدواژه‌های عنوان و توصیفگرها با یکدیگر مطابقت دارند (Ansari 2005). پژوهشی که توسط «قنوتی» بر روی مقاله‌های «اریک» و «مندلی» انجام شد، نشان می‌دهد که میزان شباهت بین کلیدواژه‌های نمایه‌ساز و نویسنده حدود ۴ درصد است (Ghanavati et al 2018). در سال ۲۰۱۵، برای بررسی سرعنوان موضوعی پزشکی (MeSH) و تطابق آن با مجله‌های داروسازی که در «مدلاین» نمایه‌سازی شده‌اند، پژوهشی انجام شده است. نتایج این پژوهش نشان داد که ۴/۷ درصد مقاله‌هایی که بازبایی شده بودند از اصطلاحات «مش» استفاده نکرده‌اند و به‌طور کامل نمایه نشده بودند. پژوهش نشان داد که میانگین واژه‌های «مش» به ازای هر مقاله ۰/۹ بوده است. در حدود ۵۲ درصد از میان مقاله‌هایی که در ژورنال‌های داروسازی نمایه شده بودند نیز به‌طور کلی از واژه‌های «مش» استفاده نکرده بودند (Minguet et al. 2015). این بدان معناست که بیش از نیمی از مقاله‌های چاپ‌شده در ۱۰ ژورنال برتر حتی بدون استفاده از یک واژه «مش» نمایه شدند. «کیپ» در پژوهش خود به بررسی نمایه‌سازی محتوای آنلاین از سه دیدگاه گروه خوانندگان، نویسندگان و نمایه‌سازان حرفه‌ای پرداخته است. میزان انطباق بین کلیدواژه نویسنده و برچسب‌های خوانندگان مقاله و کلیدواژه‌ها ۳۳ درصد، میزان انطباق بین توصیفگرها و برچسب‌ها ۱۶ درصد و میزان انطباق بین کلیدواژه‌ها و توصیفگرها ۱۹ درصد بوده است. «مارگارت» در پژوهش خود به مقایسه بین کلیدواژه‌ها، توصیفگرها و برچسب‌هایی که توسط کاربران داده شده، پرداخته است و برچسب‌ها و نمایه‌های پرکاربرد را مشخص نموده و میزان آن‌ها را به ازای هر مقاله نشان داده است (Kipp 2011). از طرفی، وی در پژوهش دیگر خود نشان داده است که اگرچه بسیاری از برچسب‌ها با کلیدواژه‌ها و توصیفگرها شباهت دارند، اما به‌طور کلی با یکدیگر متفاوت هستند (Kipp 2006).

در پژوهش دیگری که «بنی‌اقبال و پیرهادی» بر روی پارساهای فارسی انجام داده، واژه‌های عنوان و چکیده با توصیفگرهای تعیین شده در نمایه «سازمان اسناد و کتابخانه ملی ایران» را مقایسه کرده‌اند. نتایج این تحقیق نشان داد که میزان مطابقت واژه‌های عنوان با توصیفگرهای موجود در نمایه «سازمان» ۴۷ درصد و میزان مطابقت واژه‌های چکیده با توصیفگرها ۵۳/۵ درصد بوده است (۱۳۹۰).

در پژوهش حاضر ابتدا رابطه بین کلیدواژه‌های نویسنده و نمایه‌ساز بررسی خواهد شد. سپس، نمایه‌هایی که بیشترین تکرار را دارند، مشخص می‌شوند و نشان داده می‌شود

که چه میزان از نمایه‌ها در عنوان و چکیده پارساها قرار دارند. در ادامه، پارساهایی که بیشترین نمایه را دارند، مشخص خواهند شد. در انتها نیز توزیع کلیدواژه‌ها و توصیفگرها در داخل چکیده مورد بررسی قرار خواهند گرفت.

۲. هدف و سؤالات پژوهش

انتخاب کلیدواژه و نگارش چکیده برای یک متن علمی می‌تواند متخصصان و پژوهشگران را در بازیابی و درک بهتر متن یاری کند. تجربه نشان داده است که نظرات نویسنده سند و نمایه‌سازان در انتخاب نمایه برای یک سند متفاوت است. بنابراین، می‌توان نتیجه گرفت در صورتی که از یکی از این نوع نمایه‌ها در سیستم بازیابی اطلاعات استفاده شود، کارایی آن سیستم کاهش می‌یابد. بنابراین، مطالعه رابطه بین این دو می‌تواند در بازیابی اطلاعات بسیار مفید باشد. از طرفی دیگر، در نمایه‌سازی خودکار که توسط رایانه صورت می‌گیرد، خصوصیت‌های ظاهری و آماری یک متن اولین ویژگی‌هایی هستند که توسط یک سیستم رایانه‌ای مورد بررسی و تحلیل قرار می‌گیرند. هدف از انجام این پژوهش به دست آوردن، بررسی، و تحلیل این اطلاعات آماری است تا علاوه بر این که معایب در انتخاب نمایه و نگارش چکیده کشف شود، اطلاعاتی نیز در خصوص نحوه انتخاب کلیدواژه‌ها ارائه شود که این اطلاعات می‌توانند در طراحی سیستمی برای استخراج خودکار کلیدواژه‌ها مورد استفاده قرار گیرند. مهم‌ترین سؤالات این پژوهش عبارت‌اند از:

- ◇ رابطه و الگوی کلیدواژه‌های انتخابی نویسنده و نمایه‌ساز چگونه است؟
- ◇ تمرکز تحقیقات بیشتر بر روی چه موضوعاتی است؟
- ◇ چه میزان کلیدواژه‌های انتخابی دانشجو در عنوان پارسا قرار دارند؟
- ◇ توزیع کلیدواژه‌ها در چکیده به چه صورت است؟

۳. روش پژوهش

۳-۱. نحوه انتخاب رساله، تعداد پارساها

برای انجام این تحقیق، از پارساهای فارسی بین سال‌های ۱۳۹۰ تا ۱۳۹۳ که در سامانه «گنج ایرانداک» ثبت شده‌اند، استفاده شده است. یکی از معیارهای انتخاب پارساها وجود کلیدواژه‌های نویسنده پارسا و نمایه‌ساز حرفه‌ای بوده است. از آنجا که رکوردهای

رشته‌های مهندسی در بین این رکوردها بیشتر بودند، از بین پارساهای بازیابی شده تنها پارساهای رشته مهندسی انتخاب شده‌اند. تعداد این پارساها ۷۴۲ عدد هستند. پارساهای انتخاب شده می‌بایست شامل مواردی همچون چکیده، کلیدواژه‌های نویسنده، و توصیفگرها باشند. از آنجا که پارساهای رشته‌های مهندسی در این مجموعه از نظر تعداد با یکدیگر یکسان نبودند، پارساهای گرایش‌های مهندسی که تعداد بیشتری بودند، انتخاب شده‌اند. پارساهای گرایش‌های مهندسی انتخاب شده شامل رشته‌های مهندسی برق، مهندسی مکانیک و مهندسی عمران هستند.

۲-۳. تحلیل داده‌ها

برای انجام پردازش و تحلیل، پارساهای مهندسی را که در بین سال‌های ۱۳۹۰ تا ۱۳۹۳ در سامانه «گنج» به ثبت رسیده‌اند، در یک فایل «اکسل» ذخیره کردیم. هر رکورد از داده‌های ذخیره شده در نهایت، شامل حداقل عنوان، چکیده، کلیدواژه نویسنده، توصیفگر نمایه‌ساز، و فیلد تحصیلی است. برای انجام تحلیل‌های آماری از زبان برنامه‌نویسی #C استفاده شده است. مهم‌ترین تحلیل‌های به کار رفته در این بخش شامل موارد زیر است:

- ◇ میزان اشتراک نمایه‌های نویسنده و نمایه‌ساز؛
- ◇ تعداد نمایه‌های استفاده شده در هر رشته توسط نویسنده و نمایه‌ساز؛
- ◇ نمایه‌های پرتکرار (و به تبع آن موضوعاتی که تمرکز بر روی آنهاست) و عناوین پارساهای دارای نمایه‌های پرتکرار؛
- ◇ توزیع نمایه‌های استفاده شده در عنوان و چکیده.

۴. یافته‌ها

در این بخش یافته‌های پژوهش بر اساس ترتیب پرسش‌های پژوهش ارائه می‌شود. از بین پارساهایی که در سال‌های ۱۳۹۰ تا ۱۳۹۳ در پایگاه «گنج» ثبت شده‌اند، تنها ۱۹۸۴ پارسا دارای هم کلیدواژه و هم توصیفگر هستند. از بین این ۱۹۸۴ پارسا، ۷۴۲ پارسا مهندسی بوده و از بین آنها ۵۲۷ پارسا در رشته‌های مهندسی برق، مهندسی مکانیک، و مهندسی عمران هستند. جدول ۱، این اطلاعات را به صورتی جزئی تر نشان می‌دهد:

جدول ۱. پایگاه داده موجود

| کل رشته‌ها | رشته‌های مهندسی | مهندسی برق | مهندسی مکانیک | مهندسی عمران | سه رشته |
|---------------|-----------------|------------|---------------|--------------|---------|
| تعداد رساله | ۱۹۸۴ | ۷۴۲ | ۲۰۴ | ۱۳۶ | ۱۸۷ |
| تعداد دانشگاه | ۹۸ | ۶۷ | ۲۸ | ۳۳ | ۴۱ |
| | | | | | ۵۲۷ |
| | | | | | ۵۰ |

در مجموع، از ۵۲۷ پارسا که از سه رشته مهندسی عمران، مکانیک، و برق در جامعه آماری انتخاب شده‌اند، نتایجی به شرح زیر حاصل گردید: در سؤال اول مطرح شده است که رابطه و الگوی کلیدواژه‌های انتخابی نویسنده و نمایه‌ساز چگونه است؟ نتایج به‌دست آمده نشان داده است که به‌طور متوسط برای هر پارسا ۳/۲ کلیدواژه توسط نویسنده و ۷/۳ توصیفگر توسط نمایه‌ساز انتخاب شده است. آمار تعداد کلیدواژه‌ها و توصیفگرها در جدول ۲، آورده شده است:

جدول ۲. میزان کل نمایه‌هایی که توسط نویسنده و نمایه‌ساز مورد استفاده قرار گرفته‌اند

| میزان اشتراک | کل کلمات | نمایه‌ساز | نویسنده | |
|-------------------|----------|-----------|---------|--------------------------------|
| ۳۰۰ کلمه = ۸ درصد | ۳۸۳۳ | ۲۵۶۰ | ۱۵۷۳ | کلمات یکتا |
| - | ۵۵۸۳ | ۳۸۶۴ | ۱۷۱۹ | کل کلمات |
| - | ۷,۲۷ | ۴,۸۶ | ۲,۹۸ | میانگین کلمات یکتا در هر رساله |
| - | ۱۰,۵۹ | ۷,۳۳ | ۳,۲۶ | میانگین کل کلمات در هر رساله |

دیگر نتایجی که در راستای این تحقیق به‌دست آمده، نشان می‌دهد که از بین پنج پارسایی که بیشترین نمایه‌ها را دارند، سه پارسا متعلق به رشته عمران هستند. از طرف دیگر، از بین پنج پارسایی که بیشترین توصیفگر را دارند، سه پارسا متعلق به رشته عمران و پارساهای عمران بیشترین کلیدواژه‌ها را (که توسط نویسنده متن تخصیص داده شده‌اند) به خود اختصاص داده‌اند. جداول مربوط به این نتایج در پیوست آمده است.

دیگر نتایج به‌دست آمده نشان می‌دهد که رشته عمران بیشترین اشتراک بین کلمات کلیدی و توصیفگرها را داشته است؛ به‌صورتی که از بین ده پارسایی که بیشترین اشتراک بین کلیدواژه‌ها و توصیفگرها را دارند شش پارسای آن متعلق به رشته عمران است. این امر می‌تواند نشان‌دهنده همفکری بین نویسندگان و نمایه‌سازان باشد. جدول مربوط به این موضوع در پیوست آمده است. بیشترین اشتراک بین نمایه‌ساز و نویسنده

۷۵ درصد است.

در سؤال سوم مطرح شده است که تمرکز تحقیقات، بیشتر بر روی چه موضوعاتی است؟ همان‌طور که می‌دانیم، کلیدواژه‌ها نمایندهٔ سند هستند. جدول ۳، پرتکرارترین نمایه‌ها را برای این پارساها نشان می‌دهد که از میان ۱۵۷۳ کلیدواژه نویسنده و ۲۵۶۰ توصیفگر بیشترین کلماتی که توسط نویسنده، توصیفگر و به‌طور کلی، مورد استفاده قرار گرفته‌اند، به‌صورت زیر است:

جدول ۳. پرتکرارترین نمایه‌ها

| تکرار | کلیدواژه‌های پرتکرار | تکرار | توصیفگرهای پرتکرار | تکرار | نمایه‌های پرتکرار |
|-------|--------------------------------|-------|--------------------|-------|-------------------|
| ۱۵ | بهینه‌سازی | ۲۷ | شبکهٔ عصبی | ۴۰ | شبکهٔ عصبی |
| ۱۳ | شبکهٔ عصبی (شبکهٔ عصبی مصنوعی) | ۲۳ | روش اجزای محدود | ۳۶ | بهینه‌سازی |
| ۸ | الگوریتم ژنتیک | ۲۱ | بهینه‌سازی | ۲۵ | روش اجزای محدود |
| ۸ | نانوسیال | ۱۸ | شبیه‌سازی | ۲۰ | شبیه‌سازی |
| ۵ | انتقال حرارت | ۱۷ | انتقال گرما | ۱۷ | الگوریتم ژنتیک |
| ۴ | شبیه‌سازی | ۱۵ | بتن مسلح | ۱۷ | انتقال گرما |

در بین این پارساها نویسندگان در ۱۰۲ پارسا تنها یک کلیدواژه و در ۶۲ پارسا تنها دو کلیدواژه به پارساهای خود اختصاص داده‌اند. نمایه‌ساز نیز برای هیچ پارسایی کمتر از ۴ کلیدواژه اختصاص نداده است و به‌طور کلی، تقریباً همهٔ پارساها بیش از ۶ نمایه را دارند. جدول ۴، تعداد پارساها را برای هر تعداد نمایه نشان می‌دهد.

جدول ۴. تعداد پارساها به ازای کلیدواژه‌ها و توصیفگرها

| تعداد واژگان نمایه (کلیدواژه، توصیفگر یا کل نمایه‌ها) که به یک رساله تخصیص داده شده است. | تعداد پارساها (نمایه‌شده توسط نویسنده) کلیدواژه | تعداد پارساها (نمایه‌شده توسط نمایه‌ساز) توصیفگر | مجموعه نمایه‌ها (مجموع کلیدواژه‌ها و توصیفگرها) |
|--|---|--|---|
| ۱ | ۱۰۲ | ۰ | ۰ |
| ۲ | ۶۲ | ۰ | ۰ |
| ۳ | ۶۵ | ۰ | ۰ |
| ۴ | ۲۲۳ | ۵ | ۰ |

| تعداد واژگان نمایه (کلیدواژه، توصیفگر یا کل نمایه‌ها) که به یک رساله تخصیص داده شده است. | تعداد پارساها (نمایه‌شده توسط نویسنده) کلیدواژه | تعداد پارساها (نمایه‌شده توسط نمایه‌ساز) توصیفگر | مجموعه نمایه‌ها (مجموع کلیدواژه‌ها و توصیفگرها) |
|---|--|---|--|
| ۵ | ۵۳ | ۴۴ | ۱ |
| ۶ | ۱۶ | ۱۰۶ | ۱۶ |
| ۷ | ۱ | ۱۵۴ | ۳۹ |
| ۸ | ۵ | ۱۱۶ | ۷۷ |
| ۹ | ۰ | ۵۹ | ۹۲ |
| ۱۰ | ۰ | ۳۱ | ۱۰۲ |
| ۱۱ | ۰ | ۸ | ۸۲ |
| ۱۲ | ۰ | ۲ | ۴۹ |
| ۱۳ | ۰ | ۰ | ۳۵ |
| ۱۴ | ۰ | ۱ | ۴ |
| ۱۵ | ۰ | ۱ | ۲۴ |
| بیش از ۱۵ | ۰ | ۰ | ۶ |

در جدول ۴، ۵ پارسا بیشترین تعداد کلیدواژگان را دارند و در ۱۰۲ پارسا، نویسنده تنها یک کلیدواژه به آن اختصاص داده است. از جدول ۴، می‌توان مشاهده نمود که ۶۵ درصد از پارساها بین ۳ تا ۵ کلیدواژه دارند. در جدول ۷، عنوان‌های این پارساها آورده شده است. این جدول نشان می‌دهد که بیشترین کلیدواژه‌ها برای پارساهای رشته برق هستند. معیار دیگری که برای بازیابی بهتر اسناد در برخی انتشارات مانند «الزویر»، «پرکست» و «امرالد» برای انتخاب کلیدواژگان پیشنهاد شده است، انتخاب کلماتی خارج از عنوان برای نمایه است. برای سنجش این پارامتر کلیدواژگان که توسط نویسنده و نمایه‌ساز انتخاب شده‌اند، با کلمات عنوان پارسا مقایسه شده‌اند.

سؤال سوم پژوهش مطرح می‌کند که چه میزان کلیدواژه انتخابی دانشجو در عنوان پارسا قرار دارند؟ جدول ۵، نشان می‌دهد که از بین ۵۲۷ پارسای مورد بررسی حدوداً کلیدواژه‌های ۵۳ پارسا (۱۰ درصد پارساها) با عنوان انطباق کامل داشته‌اند و در ۲۳۲ پارسا، کلیدواژه‌های آن‌ها از عنوان استخراج نشده است. از طرف دیگر، در ۱۳۸ پارسا هیچ توصیفگری از عنوان استخراج نشده و توصیفگرهای هیچ‌یک از پارساها کلاً از عنوان استخراج نشده است (جدول ۶).

جدول ۵. میزان اشتراک کلیدواژه‌های نویسنده با عنوان

| میزان اشتراک با عنوان | ۰ درصد | بین ۲۵ تا ۵۰ درصد | بین ۵۰ تا ۷۵ درصد | بین ۷۵ تا ۱۰۰ درصد | تعداد پارساها |
|-----------------------|--------|-------------------|-------------------|--------------------|---------------|
| | ۲۳۲ | ۱۰۰ | ۱۱۲ | ۳۰ | ۵۳ |

جدول ۶. میزان اشتراک توصیفگرها با عنوان

| میزان اشتراک با عنوان | ۰ درصد | بین ۲۵ تا ۵۰ درصد | بین ۵۰ تا ۷۵ درصد | بین ۷۵ تا ۱۰۰ درصد | تعداد پارساها |
|-----------------------|--------|-------------------|-------------------|--------------------|---------------|
| | ۱۳۸ | ۲۵۳ | ۱۳۰ | ۶ | ۰ |

برای تحلیل دقیق‌تر، میزان انطباق کلیدواژه‌ها و توصیفگرها با عنوان پارسا و چکیده پارسا مورد بررسی قرار گرفته‌اند که مشخص می‌کند نویسنده و نمایه‌ساز چه میزان از کلمات داخل چکیده و عنوان استفاده کرده‌اند. بررسی انجام‌شده نشان می‌دهد که در ۱۶۰ پارسا کل کلیدواژه‌ها از چکیده و عنوان استخراج شده است (حدود ۳۰ درصد پارساها). از طرفی، جدول ۷، نشان می‌دهد که در ۵۰ درصد از پارساها بیشتر کلیدواژه‌ها از داخل چکیده و عنوان استخراج شده‌اند. در جدول ۸، نشان داده شده است که میزان انطباق توصیفگرها، هم با چکیده و هم با عنوان یک درصد است و ۶ درصد از پارساها کلاً توصیفگرهایشان در چکیده و در عنوان نبوده است.

جدول ۷. میزان اشتراک کلیدواژه‌های نویسنده با عنوان و چکیده

| میزان اشتراک | ۰ درصد | بین ۲۵ تا ۵۰ درصد | بین ۵۰ تا ۷۵ درصد | بین ۷۵ تا ۱۰۰ درصد | تعداد پارساها |
|--------------|--------|-------------------|-------------------|--------------------|---------------|
| | ۱۳۸ | ۲۵۳ | ۱۳۰ | ۶ | ۰ |

جدول ۸. میزان اشتراک توصیفگرها با عنوان و چکیده

| میزان اشتراک | ۰ درصد | بین ۲۵ تا ۵۰ درصد | بین ۵۰ تا ۷۵ درصد | بین ۷۵ تا ۱۰۰ درصد | تعداد پارساها |
|--------------|--------|-------------------|-------------------|--------------------|---------------|
| | ۳۳ | ۱۵۴ | ۲۴۷ | ۸۷ | ۶ |

از طرف دیگر، رابطه همبستگی بین تعداد کلیدواژه‌های انتخابی توسط دانشجو و طول چکیده و عنوان پارسا مورد بررسی قرار گرفت. این بررسی نشان داد که هیچ رابطه معناداری بین طول چکیده و طول عنوان با تعداد کلیدواژه‌های انتخاب‌شده توسط

دانشجو وجود ندارد و همبستگی آن‌ها برابر صفر است. برای محاسبه این همبستگی ابتدا بردارهای کلیدواژه و طول عنوان را به برنامه «متلب» داده و همبستگی آن‌ها را محاسبه کردیم که بسیار نزدیک به صفر شد. در محاسبه همبستگی در صورتی که مقدار دو بردار برابر ۱ شود، یعنی آن دو بردار با یکدیگر رابطه مستقیم، و در صورتی که برابر ۱- شود، بدان معناست که آن دو رابطه معکوس دارند. در صورتی که همبستگی دو بردار برابر صفر شود، بدان معناست که هیچ رابطه‌ای بین آن دو بردار وجود ندارد. همبستگی بین بردار تعداد کلیدواژه‌ها و طول عنوان نیز برابر صفر شد.

در سؤال آخر پژوهش به دنبال این هستیم که توزیع کلیدواژه‌ها در چکیده به چه صورت است؟ برای بررسی توزیع نمایه‌ها در پارساها ابتدا نمایه‌هایی که در داخل چکیده قرار دارند، جدا گردید. سپس، مکان فیزیکی آن‌ها بر حسب کلمه داخل چکیده پیدا شد و به تعداد کل کلمات آن چکیده تقسیم و در عدد ۱۰۰ ضرب گردید تا مشخص شود که آن کلمه در کدام بخش از چکیده قرار دارد. نتیجه این توزیع می‌تواند در طراحی یک سیستم نمایه‌سازی ماشینی برای به دست آوردن احتمال بیشتر وجود نمایه در مکان‌های مختلف چکیده مورد استفاده قرار بگیرد. از طرفی دیگر، طبق استاندارد «انسی / میزو» چکیده شامل ۵ بخش است که به ترتیب عبارت‌اند از: هدف چکیده، روش‌شناسی، نتایج، جمع‌بندی، و اطلاعات اضافی (National Information Standards Organization (NISO) 2010). بنابراین، در صورتی که برای هر بخش تقریباً یک سهم مساوی در چکیده در نظر گرفته شود، با بررسی توزیع کلیدواژه‌ها در چکیده می‌توان حدوداً به میزان اهمیت هر بخش پی برد. همان‌طور که در جدول ۹، ملاحظه می‌شود، بیش از ۶۰ درصد از نمایه‌هایی که در داخل چکیده وجود دارند، در ۴۰ درصد ابتدایی چکیده واقع شده‌اند.

جدول ۹. توزیع نمایه‌ها در چکیده

| درصد توصیفگرها | درصد کلیدواژه‌ها | |
|----------------|------------------|-----------------------------------|
| ۴۵ درصد | ۴۰ درصد | در ۲۰ درصد اول چکیده قرار دارند |
| ۱۹ درصد | ۲۴ درصد | در ۲۰ درصد دوم چکیده قرار دارند |
| ۱۸ درصد | ۱۸ درصد | در ۲۰ درصد سوم چکیده قرار دارند |
| ۱۰ درصد | ۱۱ درصد | در ۲۰ درصد چهارم چکیده قرار دارند |
| ۷ درصد | ۷ درصد | در ۲۰ درصد پنجم چکیده قرار دارند |

۵. تجزیه و تحلیل یافته‌ها و جمع‌بندی

بیشتر پژوهش‌های پیشین در زمینه مقایسه کلیدواژه‌ها و توصیفگرها بر روی مقالات انجام شده است. در این پژوهش به بررسی میزان تشابه کلیدواژه‌ها و توصیفگرها پرداخته و نشان داده شده است که شباهت بین کلیدواژه‌های نویسنده و توصیفگرهای نمایه‌ساز کمتر از ۸ درصد است. اختلاف این دو نشان‌دهنده اختلاف نظر زیاد در انتخاب کلیدواژه برای متن بین نویسنده و نمایه‌ساز است. بنابراین، نمی‌توان به تنهایی از یکی در سیستم بازیابی اطلاعات مؤثر استفاده کرد. از طرفی، پژوهشی که «قنوتی» بر روی مقاله‌های پایگاه اطلاعاتی «اریک» و «مندلی» انجام داده، نشان داده است که شباهت بین توصیفگرها و کلیدواژه‌ها ۴ درصد است.

همان‌طور که در جدول ۲، مشاهده گردید، تفاوت کلیدواژه‌های یکتا و کل که توسط نویسنده انتخاب شده، بسیار اندک است و این بدان معناست که نویسنده‌های مقاله دایره لغات بسیار متفاوتی از یکدیگر دارند. می‌توان نتیجه گرفت که استفاده تنها از توصیفگرهای نمایه‌ساز نمی‌تواند کافی باشد. بنابراین، استفاده هر یک به تنهایی می‌تواند باعث عدم کارایی لازم در استفاده از کلیدواژه‌های نویسنده در بازیابی اطلاعات شود. از طرفی، از آنجا که کلیدواژه‌ها در بازیابی اطلاعات نقش کلیدی دارند، می‌توان با جمع کلمات و غنی کردن کلیدواژه‌ها سیستم بازیابی اطلاعات کارآمدتری به دست آورد. از طرفی دیگر، نشان داده شده است که در ۶۰ درصد رساله‌ها کلیدواژه‌های کمتر از ۲۵ درصد آن‌ها از عنوان انتخاب شده‌اند. نتایج در جدول ۷ نشان می‌دهد که حدود ۵۰ درصد کلیدواژه‌های نویسنده عیناً از داخل چکیده استخراج شده‌اند. در پژوهش Strader (2011) نیز نتیجه مشابه ۵۴ درصدی به دست آمده است.

در ادامه، میزان اشتراک نمایه‌های استفاده‌شده توسط نویسنده و نمایه‌ساز با عنوان و چکیده مورد بررسی قرار گرفت. بررسی انجام شده نشان می‌دهد که کلیدواژه‌هایی که نویسنده از آن‌ها استفاده کرده در ۶۲ درصد از پارساها، کمتر از ۲۵ درصد با عنوان انتطابق دارند (طبق جدول ۷ حدود ۵۶ درصد از کلیدواژه‌های نویسنده در داخل عنوان وجود دارند) که این مقدار اندکی کمتر از یافته «انصاری» در پژوهش خود است (انتخاب ۷۰ درصد کلیدواژه‌ها از عنوان). از طرف دیگر، ۵۱ درصد از کلیدواژه‌های پارساهای بررسی شده بیش از ۷۵ درصد با چکیده و عنوان شباهت دارند. در پژوهش Strader (2011) نیز نتیجه ۵۴ درصدی به دست آمده است. این نتایج نشان می‌دهند که نویسندگان عموماً

اکثر کلیدواژه‌های خود را عیناً از چکیده انتخاب می‌کنند.

در ادامه، نمایه‌های پر کاربرد در رشته‌های عمران، برق و مکانیک که در پارساهای فارسی مورد استفاده قرار گرفته‌اند، ارائه شده است. نتایج نشان داد که در هر سه رشته عمران، مکانیک و برق از نمایه‌های شبیه‌سازی، بهینه‌سازی، شبکه‌های عصبی و الگوریتم ژنتیک بسیار استفاده شده است. از آنجا که کلیدواژه‌ها نماینده یک سند هستند، استفاده زیاد از این لغات به عنوان کلمات می‌تواند بیانگر این باشد که تمایل نویسندگان بیشتر بر روی چه موضوعی بوده است (Zhang 2008; Rose et al. 2010). بنابراین، از جدول ۳، می‌توان نتیجه گرفت که تمایل و تمرکز پارساهای نوشته‌شده در سال‌های ۱۳۹۰ تا ۱۳۹۳ بیشتر بر روی بهینه‌سازی با استفاده از الگوریتم‌های تکاملی بوده است. از این رو، می‌توان به صورت تقریبی محدوده پژوهش‌ها را مورد شناسایی قرار داده و با استفاده از این شناخت پژوهش‌های انجام‌شده در سطح دانشگاهی کشور را با نیازهای صنعت همسوتر ساخت. در انتها، به بررسی وضعیت توزیع نمایه‌ها در داخل چکیده پرداخته و نشان داده شد که ۶۴ درصد نمایه‌هایی که در داخل چکیده وجود دارند، در ۴۰ درصد بخش نخست هستند. این آمار نشان می‌دهد که نویسنده بیشتر مطالب خود را در هدف و روش پژوهش بیان می‌کند و با استفاده از این ویژگی، سیستم‌های استخراج خودکار کلیدواژه و یا خلاصه‌سازهای متن می‌توانند به کلمات کاندیدی که به عنوان کلیدواژه از این بخش‌ها استخراج می‌کنند، وزن بیشتری دهند.

فهرست منابع

اباذری، زهرا، سعیده پالیزوانی. ۱۳۹۱. بررسی میزان همخوانی و روزآمدی سرعنوان‌های موضوعی فارسی با عناوین کتاب‌های منتشرشده در پایگاه کتابخانه ملی ایران در سال‌های ۱۳۸۴-۱۳۸۸. *نظام‌ها و خدمات اطلاعاتی* ۲(۱): ۲۷-۴۲.

بنی‌اقبال، ناهید، فریبرز خسروی، و صدیقه پیرهادی. ۱۳۹۰. مقایسه واژه‌های عنوان و چکیده پایان‌نامه‌ها با توصیفگرهای تعیین‌شده در نمایه سازمان اسناد و کتابخانه ملی ایران. *مطالعات ملی کتابداری و سازمان‌دهی اطلاعات* ۸۶: ۱۴۷-۱۳۴.

References

- Ansari, Mariam. 2005. Matching between Assigned Descriptors and Title Keywords in Medical Theses. *Library Review* 54 (7): 410-414.
- Carpinetto, Claudio, Stanislaw Osiński, Giovanni Romano, and Dawid Weiss. 2009. A Survey of Web Clustering Engines. *ACM Computing Surveys (CSUR)* 41 (3): 17.
- Chen, Fang, Kesong Han, and Guilin Chen. 2002. An Approach to Sentence-Selection-Based Text

- Summarization. Paper presented at the TENCON'02 Proceedings. IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering 2002 Beijing, China.
- Ghanavati, Maryam, Alireza Noruzi, Maryam Nakhoda, and Ashkan Khatir. Consistency between descriptors, author-supported keywords and tags in the ERIC and Mendeley databases. *Iranian Journal of Information Processing and Management* 33, no. 4 (2018): 1745-1766.
- Gupta, Vishal, and Gurpreet Singh Lehal. 2010. A Survey of Text Summarization Extractive Techniques. *Journal of Emerging Technologies in Web Intelligence* 2 (3): 258-68.
- Hulth, Anette. Improved Automatic Keyword Extraction Given More Linguistic Knowledge. 2003. Paper presented at the Proceedings of the 2003 conference on Empirical methods in natural language processing, 2003 Stroudsburg, PA, USA.
- Khan, Khairullah, Baharum B Baharudin, and Aurangzeb Khan. 2009. Mining Opinion from Text Documents: A Survey. Paper presented at the Digital Ecosystems and Technologies, 2009. DEST'09. 3rd IEEE International Conference on, 2009 New York, NY, USA.
- Kipp, Margaret El. 2006. Complementary or Discrete Contexts in Online Indexing: A Comparison of User, Creator and Intermediary Keywords. *E-Print in Journal of library and Information Science* 35 17-48 :(1).
- _____. 2011. Tagging of Biomedical Articles on Citeulike: A Comparison of User, Author and Professional Indexing. *Knowledge Organization* 38 (3): 245-261.
- Kyoomarsi, Farshad, Hamid Khosravi, Esfandiar Eslami, Pooya Khosravyan Dehkordy, and Asghar Tajoddin. 2008. Optimizing Text Summarization Based on Fuzzy Logic. Paper presented at the Seventh IEEE/ ACIS International Conference on Computer and Information Science, Washington, DC, USA.
- Litvak, Marina, and Mark Last. 2008. Graph-Based Keyword Extraction for Single-Document Summarization. Paper presented at the Proceedings of the workshop on Multi-source Multilingual Information Extraction and Summarization, August 23 – 23, 2008. Manchester, United Kingdom.
- Minguet, Fernando, Teresa M Salgado, Lucienne van den Boogerd, and Fernando Fernandez-Llimos. 2015. Quality of Pharmacy-Specific Medical Subject Headings (Mesh) Assignment in Pharmacy Journals Indexed in Medline. *Research in Social and Administrative Pharmacy* 11 (5): 686-95.
- National Information Standards Organization (NISO). 2010. *Ansi/ Niso Z39.14-1997 (R2015) "Guidelines for Abstracts*. Baltimore, Maryland: National Information Standards Organization
- Rose, Stuart, Dave Engel, Nick Cramer, and Wendy Cowley. 2010. Automatic Keyword Extraction from Individual Documents. *Text Mining* 1: 1-20.
- Schultz, Claire K, Wallace L Schultz, and Richard H Orr. 1965. Comparative Indexing: Terms Supplied by Biomedical Authors and by Document Titles. *American Documentation* 16 (4): 299-312.
- Strader, C Rockelle. 2011. Author-Assigned Keywords Versus Library of Congress Subject Headings. *Library resources & technical services* 53 (4): 243-50.
- Zhang, Chengzhi. 2008. Automatic Keyword Extraction from Documents Using Conditional Random Fields. *Journal of Computational Information Systems* 4 (3): 1169-80.

پیوست

جدول ۱. بیشترین نمایه‌ها و عنوان رساله

| عنوان رساله | رشته | تعداد نمایه‌ها |
|--|--------|----------------|
| بررسی رفتار دینامیکی غیرخطی اتصال صلب، با مقطع کاهش یافته تیر (RBS) با روش اجزاء عمران محدود | عمران | ۱۹ |
| تحلیل عددی بارگذاری دینامیکی بر روی یک سازه تحت محیط‌های واسط متفاوت | مکانیک | ۱۹ |
| شبیه‌سازی دوبعدی جریان در لوله‌ها | عمران | ۱۸ |
| آنالیز حساسیت شبکه GPS شمال تهران به فعالیت‌های احتمالی آتش‌فشان دماوند | عمران | ۱۷ |
| ارزیابی کفایت سیستم توزیع شامل واحدهای تولید پراکنده تجدیدپذیر و خودروهای برقی برق | عمران | ۱۶ |

جدول ۲. پارساهایی که بیشترین توصیفگرها را دارند

| تعداد توصیفگر | رشته | عناوین پر توصیفگر نمایه‌ساز |
|---------------|--------|---|
| ۱۵ | مکانیک | تحلیل عددی بارگذاری دینامیکی بر روی یک سازه تحت محیط‌های واسط متفاوت |
| ۱۴ | عمران | بررسی رفتار دینامیکی غیرخطی اتصال صلب، با مقطع کاهش یافته تیر (RBS) با روش اجزاء محدود |
| ۱۲ | برق | کنترل غیرخطی پدیده آشوب در دینامیک وضعیت ماهواره |
| ۱۲ | عمران | بررسی عملکرد لرزه‌ای قاب‌های با دهانه بلند و ارتفاع کم با عدم رعایت ضابطه تیر ضعیف-ستون قوی |
| ۱۱ | عمران | تعیین فرکانس ارتعاشی صفحات ترک‌دار با استفاده از روش‌های بدون شبکه |

جدول ۳. پارساهایی که بیشترین کلیدواژه‌ها را دارند

| تعداد کلیدواژه | رشته | عناوین |
|----------------|-------|--|
| ۸ | برق | ردیابی شیء متحرک در شبکه‌های حسگر بی‌سیم با انرژی مصرفی کم |
| ۸ | برق | ارزیابی کفایت سیستم توزیع شامل واحدهای تولید پراکنده تجدیدپذیر و خودروهای برقی |
| ۸ | عمران | آنالیز حساسیت شبکه GPS شمال تهران به فعالیت‌های احتمالی آتش‌فشان دماوند |
| ۸ | برق | طراحی و ساخت یک حسگر گاز بر مبنای ساختار فیبرنوری |
| ۸ | عمران | شبیه‌سازی دوبعدی جریان در لوله‌ها |

جدول ۴. عناوین و تعداد کلمات کلیدی پارساهایی که کلمات کلیدی و توصیفگرهای آن‌ها بیشترین میزان اشتراک را با یکدیگر دارند

| عنوان رساله | رشته | درصد کلمات مشترک با توصیفگرها | درصد کلمات مشترک با نویسنده | درصد اشتراک با کل نمایه‌ها | تعداد کل کلمات | تعداد کلیدواژه | تعداد توصیفگر |
|---|--------|-------------------------------|-----------------------------|----------------------------|----------------|----------------|---------------|
| بهبود کیفیت خدمات در سیستم‌های رادیوشناختی با استفاده از بهینه‌سازی برخی کمیت‌ها | برق | ۵۰ درصد | ۷۵ درصد | ۴۳ درصد | ۷ | ۴ | ۶ |
| تحلیل لرزه‌ای توربین‌های بادی با در نظر گرفتن اثرات اندرکنش خاک و سازه | عمران | ۵۰ درصد | ۷۵ درصد | ۴۳ درصد | ۷ | ۴ | ۶ |
| شناسایی و پایش پدیده گرد و غبار از داده‌های سنجش از دوری با استفاده از عمق اپتیکی | عمران | ۵۰ درصد | ۷۵ درصد | ۴۳ درصد | ۷ | ۴ | ۶ |
| بررسی خوردگی و نفوذ یون کلر در بتن خودتراکم حاوی زئولیت و متاکائولین | عمران | ۵۰ درصد | ۷۵ درصد | ۴۳ درصد | ۷ | ۴ | ۶ |
| تأثیر خصوصیات هیدرولیکی مصالح سدهای غیرهمگن بر پایداری آن | عمران | ۶۰ درصد | ۵۰ درصد | ۳۸ درصد | ۸ | ۶ | ۵ |
| تحلیل ویسکوالاستیک تنش تورق در اتصالات چسبی | مکانیک | ۴۳ درصد | ۷۵ درصد | ۳۸ درصد | ۸ | ۴ | ۷ |
| تخصیص منابع در شبکه‌های چند کاربره رله‌ای بی‌سیم رقابتی با رهیافت نظریه بازی | برق | ۴۰ درصد | ۶۷ درصد | ۳۴ درصد | ۶ | ۳ | ۵ |
| بررسی فرایند ادغام قطرات معلق از طریق شبیه‌سازی عددی با روش‌های حجم سیال و شبکه‌ی بولتزمن | مکانیک | ۳۸ درصد | ۷۵ درصد | ۳۴ درصد | ۹ | ۴ | ۸ |

| عنوان رساله | رشته | درصد کلمات مشترک با توصیفگرها | درصد کلمات مشترک با نویسنده | درصد اشتراک با کل نمایه‌ها | تعداد کل کلمات | تعداد کلیدواژه | تعداد توصیفگر |
|--|-------|-------------------------------|-----------------------------|----------------------------|----------------|----------------|---------------|
| بررسی اثرات ساختگاه و تخمین خصوصیات جنبش نیرومند زمین با استفاده از داده‌های میکروترمور و مقایسه نتایج آن با مدل‌سازی پروفیل خاک (مطالعه موردی: ناحیه غربی شهر بابل) | عمران | ۳۸ درصد | ۷۵ درصد | ۳۴ درصد | ۹ | ۴ | ۸ |
| بررسی روش آزمایشگاهی پاکسازی آلودگی نفتی در محیط‌های آبی (آب‌های زیرزمینی) | عمران | ۳۸ درصد | ۷۵ درصد | ۳۴ درصد | ۹ | ۴ | ۸ |

اشکان خطیر

متولد ۱۳۶۴ دانشجوی دکتری در رشته مهندسی فناوری اطلاعات در پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) است. تحلیل روند، متن کاوی و داده کاوی از جمله علایق پژوهشی وی است.



سهیل گنجه‌فر

متولد ۱۳۴۹ دارای مدرک تحصیلی دکتری در رشته برق کنترل است. ایشان هم‌اکنون استاد تمام دانشگاه بوعلی سینا، گروه برق است. از جمله علاقمندی‌های ایشان، یادگیری ماشین و متن کاوی است.



