
بسم الله الرحمن الرحیم

وزارت علوم، تحقیقات و فناوری

پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)

طرح پژوهشی

مقایسه روند پژوهش‌ها در حوزه فناوری اطلاعات در پایان‌نامه‌ها و رساله‌های کشور و روند جهانی آن با استفاده از روش تحلیل متون

مجری

آرمان ساجدی‌نژاد

همکار

محمد ربیعی

بهمن‌ماه ۱۳۹۹

وزارت علوم تحقیقات و فناوری
پژوهشگاه علوم و فناوری اطلاعات ایران
ایرانداک

شماره:
تاریخ:
پیوست
۳۳/۹۹۴۵
۱۳۹۹/۱۱/۱۹
پژوهش . مدیریت دانش . آموزش
ندارد

به نام خدا

ایرانداک در نیم‌قرن دوم کار خود همچنان جوان و کوشا، همراه هر پژوهش، پژوهشگر، و سیاست‌گذار
۱۳۹۹ – ۱۳۴۷

گواهی

انجام طرح پژوهشی:

- **عنوان:** مقایسه روند پژوهش‌ها در حوزه فناوری اطلاعات در پایان‌نامه‌ها و رساله‌های کشور و روند جهانی آن با استفاده از روش تحلیل متون
- **مجری:** دکتر آرمان ساجدی‌نژاد
- **همکار:** دکتر محمد ربیعی
- **شمارهٔ قرارداد:** ۳۳/۵۵۹۵
- **تاریخ قرارداد:** ۱۳۹۸/۶/۲
- **تاریخ شروع:** ۱۳۹۸/۶/۲
- **تاریخ پایان:** ۱۳۹۹/۱۰/۱۸
- **ناظر:** دکتر عباس احمدی

در پژوهشگاه علوم و فناوری اطلاعات ایران گواهی می‌شود. گزارش پایانی این طرح، بر پایهٔ داوری در نشست ۴۰۱ (تاریخ ۱۳۹۹/۱۱/۱۲) شورای پژوهشی پژوهشگاه به تصویب رسیده است. لازم می‌دانم مراتب قدردانی خود را به مجری، همکار، و ناظر این طرح تقدیم دارم.

با آرزوی بهروزی
پرویز شهریاری
معاون پژوهش و آموزش

تهران . خیابان انقلاب اسلامی . چهارراه فلسطین . شمارهٔ ۱۰۹۰ صندوق پستی ۱۳۷۱-۱۳۱۸۵
تلفن: ۶۶۴۹۴۹۸۰ - ۶۶۹۵۱۴۳۰ دورنگار: ۶۶۴۶۲۲۵۴ تلفن گویا: ۶۶۴۹۳۹۵۴
irandoc.ac.ir
info@irandoc.ac.ir

مقایسه روند پژوهش‌ها در حوزه فناوری اطلاعات در پایان‌نامه‌ها و رساله‌های کشور و روند جهانی آن با استفاده از روش تحلیل متون

مدیر طرح پژوهشی: دکتر آرمان ساجدی نژاد، پژوهشگاه علوم و فناوری اطلاعات ایران.

نشانی: تهران، خیابان انقلاب، چهارراه فلسطین، شماره ۱۰۹۰، طبقه ۴، اتاق ۴۱۲

صندوق پستی: ۱۳۱۵۷۷۳۳۱۴ تلفن: ۶۶۴۹۴۹۸۰-۰۲۱(داخلی ۴۳۲)

وبگاه: http://irandoc.ac.ir/u/a-sajedinejad

رایانامه: *Sajedinejad@irandoc.ac.ir*

همکار طرح پژوهشی: دکتر محمد ربیعی، پژوهشگاه علوم و فناوری اطلاعات ایران.

چکیده

همواره این سوال مطرح می‌گردد که روند توسعه پژوهش‌ها در یک زمینه خاص علمی در کشور با روندهای جهانی آن زمینه علمی تطابق دارد و یا حوزه‌های مفهومی یک زمینه علمی مانند فناوری اطلاعات کجاست و مرزهای آن در دنیا کدامست؟

بنابراین در این طرح پژوهشی، با یک رویکرد روش شناختی به مدل‌سازی موضوعی فعالیت‌های پژوهشی با ابزارهای متن‌کاوی پرداخته و با پردازش نتایج مدل‌سازی موضوعی در زمینه فناوری اطلاعات، روند پژوهش‌ها را در طول زمان (بازه ۱۰ ساله) کشف می‌کنیم. در این طرح مجموعه‌ای از مجلات مهم و پیشرو در زمینه فناوری اطلاعات در پایگاه WOS[1] را انتخاب و با رویکرد پیشنهادی سعی شده است تا ارتباطات بین واژگان پرکاربرد و تکامل زمانی آنها را تجزیه و تحلیل کنیم و نهایتا با روندهای ملی مقایسه نماییم. لازم به ذکر است ملاک تشخیص پیشرو بودن این مجلات در این طرح از دو بعد سابقه آنها و ضریب تاثیر و همچنین استفاده از نظر خبرگان در نظر گرفته می‌شود. همچنین از سبد کلیدواژه‌های مورد تایید خبرگان برای استخراج هر دو داده‌های ملی و بین‌المللی استفاده شده است تا یکنواختی و امکان مقایسه دو مجموعه فراهم شود.

لازم به ذکر است، در این طرح پژوهشی، پژوهش‌های انجام شده در زمینه فناوری اطلاعات کشور، محدود به پارسا[2]های ثبت شده در پایگاه گنج بوده و همراستایی آنها با روندهای بین‌المللی مورد تحلیل قرار می‌گیرد.

کلیدواژه‌ها: مدل‌سازی موضوعی[3]، متن کاوی، پایگاه گنج، فناوری اطلاعات، ایرانداک

[1] Web of science

[2] – پایان‌نامه و رساله

[3] Topic modeling

فهرست مطالب

عنوان | شمارهٔ صفحه

۱. مقدمه و اهداف ۱
۱-۱. مقدمه ۲
۱-۲. اهداف طرح ۶
۱-۳. پیوند با اساس‌نامه و برنامهٔ استراتژیک ۷
۱-۴. تعریف واژه‌ها و مفاهیم ۹
۲. مدل‌سازی موضوعی حوزه فناوری اطلاعات ۱۰
۲-۱. مقدمه ۱۱
۲-۲. روش مدل‌سازی موضوعی ۱۵
۲-۳. روش پژوهش ۱۷
۲-۴. انتخاب مجلات معتبر ۲۱
۲-۴-۱. جستجوی مجلات برتر در حوزه فناوری اطلاعات ۲۴
۲-۴-۲. انتخاب مجلات با استفاده از نظر تکمیلی خبرگان ۲۷
۲-۵. انتخاب کلیدواژگان منتخب ۳۰
۲-۶. بررسی روند توسعه کلیدواژه‌ها ۳۵
۲-۷. دسته‌بندی حوزه‌ها با بهره‌گیری از فرکانس همزمانی استفاده از کلیدواژگان ۳۸
۲-۷-۱. بررسی چگونگی کاربرد کلیدواژگان ۳۹
۲-۷-۲. بررسی روند کاربرد کلیدواژگان و تمرکز بر حوزه‌های مختلف ۴۴
۲-۷-۳. دسته‌بندی حوزه‌های علمی بر اساس مدل‌سازی موضوعی کلیدواژه‌ها ۵۰
۲-۸. جمع‌بندی و نتیجه‌گیری ۵۶

۳. مدل‌سازی پژوهش‌های ملی در حوزه فناوری اطلاعات و مقایسه با روندهای بین‌المللی ۶۰
۳-۱. مقدمه ۶۱
۳-۲. روش پژوهش ۶۳
۳-۳. انتخاب کلیدواژگان قابل جستجو ۶۵
۳-۴. بررسی روند توسعه کلیدواژه‌ها ۶۷
۳-۴-۱. بررسی روند کاربرد کلیدواژگان و تمرکز بر حوزه‌های مختلف ۶۹
۳-۴-۲. مقایسه روند کلیدواژگان در مطالعات ملی و بین‌المللی ۷۳
۳-۵. دسته‌بندی حوزه‌های علمی بر اساس مدل‌سازی موضوعی کلیدواژه‌ها ۷۹
۳-۵-۱. مقایسه دسته‌بندی‌ها ۸۸
۳-۵-۲. اعتبارسنجی مدل ۹۶
۳-۶. جمع‌بندی و نتیجه‌گیری ۹۷
۴. فهرست منابع ۱۰۲
۴-۱. فهرست منابع انگلیسی ۱۰۳
۴-۲. فهرست منابع فارسی ۱۰۶
پیوست: دسته‌بندی پایگاه WOS از انتشارات علوم ۱۰۷

فهرست شکل‌ها

عنوان **شمارهٔ صفحه**

شکل ۱- نمونه توسعه چارچوب فناوری اطلاعات (باتیستی و همکاران ۲۰۱۵)..... ۴
شکل ۲- نمایش کلمات پرکاردبرد در حوزه آمار (باتیستی و همکاران ۲۰۱۵)..... ۱۲
شکل ۳-میزان کاربرد یک کلیدواژه (باتیستی و همکاران ۲۰۱۵)..... ۱۲
شکل ۴- مقایسه روش‌های مختلف خوشه‌بندی..... ۱۵
شکل ۵- مراحل پژوهش جاری..... ۱۷
شکل ۶- مدل پژوهش جاری..... ۱۸
شکل ۷- نمایش دسته‌بندی WC پایگاه (WOS Catagories)..... ۱۹
شکل ۸- نمایی از دسته‌بندی پایگاه..... ۲۰
شکل ۹- نمای تلفیق دو جستجو در پایگاه..... ۳۷
شکل ۱۰- میزان استخراج منبع بر حسب سال..... ۳۹
شکل ۱۱- حوزه‌های ترکیبی پرکاربرد..... ۳۹
شکل ۱۲- درصد تمرکز بر حوزه‌ها در مطالعات فناوری اطلاعات..... ۴۱
شکل ۱۳- انتشارات فعال در زمینه فناوری اطلاعات..... ۴۴
شکل ۱۴- کلیدواژه‌های پرکاربرد حوزه فناوری اطلاعات در ده سال..... ۴۴
شکل ۱۵- کاربرد و تکرار کلیدواژه‌ها در سال ۲۰۱۱..... ۴۵
شکل ۱۶ -کاربرد و تکرار کلیدواژه‌ها در سال ۲۰۱۹..... ۴۶
شکل ۱۷- روند رشد ۴ حوزه نوظهور کشف شده در پژوهش..... ۴۶
شکل ۱۸- فلوچارت اجرای الگوریتم LDA..... ۵۲
شکل ۱۹- نمای خروجی مدل‌سازی موضوعی..... ۵۳
شکل ۲۰- میزان انحراف سرگشتگی با توجه به اندازه دسته‌بندی..... ۵۴
شکل ۲۱- تفاوت در تعداد مقالات دسته‌بندی‌ها در ابتدا و انتهای بازه پژوهش جاری..... ۵۹
شکل ۲۲- مراحل پژوهش جاری در فاز دوم..... ۶۳
شکل ۲۳- مدل پژوهش جاری در فاز دوم..... ۶۴
شکل ۲۴- میزان استخراج منبع بر حسب سال..... ۶۹
شکل ۲۵- رشد داده‌های استخراج شده از پایگاه گنج در سال..... ۶۹

شکل ۲۶- کلیدواژه‌های پرکاربرد حوزه فناوری اطلاعات در ده سال ۷۰
شکل ۲۷- کاربرد و تکرار کلیدواژه‌های پایگاه گنج در سال ۱۳۸۹ ..... ۷۰
شکل ۲۸ -کاربرد و تکرار کلیدواژه‌ها در سال ۲۰۱۹ ..... ۷۱
شکل ۲۹- روند رشد ۴ حوزه نوظهور کشف شده در پژوهش ۷۲
شکل ۳۰- مقایسه تعداد کلیدواژه‌های پرکاربرد ۷۳
شکل ۳۱- مقایسه روند مطالعات در کلیدواژه Wireless Network (s) ۷۴
شکل ۳۲- مقایسه روند مطالعات در کلیدواژه Cloud Computing ۷۵
شکل ۳۳- مقایسه تست شکاف زمانی بین مطالعات در کلیدواژه Cloud Computing ۷۵
شکل ۳۴- مقایسه روند مطالعات در کلیدواژه Security ۷۶
شکل ۳۵- مقایسه روند مطالعات در کلیدواژه Internet of Things (IOT) ۷۶
شکل ۳۶- مقایسه روند مطالعات در کلیدواژه Big Data ۷۷
شکل ۳۷- مقایسه روند مطالعات در کلیدواژه‌ی Deep Learning ۷۷
شکل ۳۸- مقایسه تست شکاف زمانی بین مطالعات در کلیدواژه Deep Learning ۷۸
شکل ۳۹- مقایسه روند مطالعات در کلیدواژه Social Media ۷۸
شکل ۴۰- مقایسه روند مطالعات در کلیدواژه‌های Machine Learning ۷۹
شکل ۴۱- میزان انحراف سرگشتگی با توجه به اندازه دسته‌بندی ۸۲
شکل ۴۲- مقایسه تعداد مطالعات در هر دو دسته‌بندی در موضوع اول پایگاه WOS ۹۲
شکل ۴۳- مقایسه تست شکاف زمانی بین مطالعات در موضوع اول پایگاه WOS ۹۳
شکل ۴۴- مقایسه تعداد مطالعات در هر دو دسته‌بندی در موضوع دوم و سوم پایگاه WOS ۹۳
شکل ۴۵- مقایسه تست شکاف زمانی بین مطالعات در موضوع دوم و سوم پایگاه WOS ۹۴
شکل ۴۶- مقایسه تعداد مطالعات در هر دو دسته‌بندی در موضوع چهارم پایگاه WOS ۹۴
شکل ۴۷- مقایسه تعداد مطالعات در هر دو دسته‌بندی در موضوع پنجم پایگاه WOS ۹۵
شکل ۴۸- مقایسه تعداد مطالعات در هر دو دسته‌بندی در موضوع ششم پایگاه WOS ۹۵
شکل ۴۹- مقایسه تعداد مطالعات در هر دو دسته‌بندی در موضوع هفتم پایگاه WOS ۹۶
شکل ۵۰- مقایسه تعداد مطالعات در هر دو دسته‌بندی در موضوع هشتم پایگاه WOS ۹۶

فهرست جدول‌ها

عنوان **شمارهٔ صفحه**

جدول ۱- شرح روش پژوهش ۱۷
جدول ۲- فهرست مجلات دارای ضریب تاثیر بالا در پایگاه WOS ۲۲
جدول ۳- اطلاعات و حوزه پوشش مجلات منتخب ۲۴
جدول ۴- کلمات پرکاربرد در مجلات دارای ضریب تاثیر بالا ۲۵
جدول ۵- تکرار کلیدواژگان پرکاربرد در مجلات مورد تایید خبرگان ۲۷
جدول ۶- انتخاب کلیدواژه از دیدگاه خبرگان برای تشکیل سبد واژگان ۳۰
جدول ۷- نمونه کلیدواژگان مقالات یافت شده در جستجوی نهایی ۳۷
جدول ۸- حوزه‌های پر در مطالعات مرتبط با فناوری اطلاعات ۴۰
جدول ۹- کلیدواژه‌های مورد استفاده هر دسته پایگاه WOS ۴۱
جدول ۱۰- مقایسه کلیدواژه‌های پرکاربرد در دو بازه زمانی ۴۷
جدول ۱۱- خویشاوندی کلیدواژگان ۴۸
جدول ۱۲- دسته‌بندی موضوعات با روش مدلسازی موضوعی ۵۵
جدول ۱۳- مقایسه دسته‌بندی پایگاه WOS و دسته‌بندی انجام شده ۵۶
جدول ۱۴- شرح روش پژوهش ۶۳
جدول ۱۵- کلیدواژه منتخب برای تشکیل سبد واژگان جستجو در پایگاه گنج ۶۵
جدول ۱۶- مقایسه کلیدواژه‌های پرکاربرد در دو بازه زمانی ۷۲
جدول ۱۷- کلیدواژه‌های دسته‌بندی موضوعات با روش مدلسازی موضوعی ۸۳
جدول ۱۸- نمونه محاسبه تابع نزدیکی به هر دسته موضوعی ۸۹
جدول ۱۹- نمونه اختصاص تابع نزدیکی به هر دسته موضوعی دو پایگاه ۹۰
جدول ۲۰- تعداد اختصاص هر پژوهش به هر دسته موضوعی پایگاه WOS ۹۱
جدول ۲۱- تطابق موضوعات پایگاه داده WOS و پایگاه گنج ۹۱
جدول ۲۲- علت فاصله کلیدواژه‌ها از نظر خبرگان خبرگان ۹۹
جدول ۲۳- مقایسه نقص کلیدواژه دسته‌بندی پایگاه WOS و دسته‌بندی پایگاه گنج ۱۰۰

قدردانی

در این قسمت جا دارد از زحمات همکاران خوبم بویژه جناب آقای دکتر علیدوستی رئیس محترم پژوهشگاه و دیگر همکاران گرامی که به عنوان خبرگان در این پژوهش سهمی بزرگی داشته‌اند و در این طرح کمک شایانی نموده‌اند و همچنین همکارانم در پژوهشکده مدیریت فناوری اطلاعات تشکر و قدردانی نمایم. همچنین از همکارم در این طرح پژوهشی جناب آقای دکتر محمد ربیعی تشکر دارم. قدردانی خود را از جناب آقای دکتر عباس احمدی که با نکات دقیق و آموزنده خود این طرح را غنی نموده‌اند ابراز می‌دارم.

از جناب آقای دکتر شهریاری معاونت محترم پژوهش و همچنین سایر اعضای محترم شورای پژوهش برای حمایت‌هایشان، سپاسگزارم.

فصل اول

مقدمه و اهداف

۱-۱. مقدمه

پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) به‌عنوان تنها پژوهشگاه فعال در زمینه فناوری اطلاعات در وزارت عتف یکی از مراجعی است که با ساخت ده‌ها فرهنگ مرتبط جامع[۴]، به عنوان یکی از مراجع مورد استناد پژوهش‌ها قرار گرفته است. با وجود انجام شدن پژوهش‌های فراوان در پژوهشگاه با هدف استخراج مدل‌های توسعه فناوری اطلاعات، به نظر می‌رسد کشف مرزهای حوزه فناوری اطلاعات و روندهای آن (در پژوهش‌های انجام شده در کشور و جهان و مقایسه همراستایی آنها) هنوز توسط این پژوهشگاه صورت نپذیرفته است. از سوی دیگر ایرانداک با تجمیع داده‌های فعالیت‌های علمی کشور ارزش افزوده فراوانی (مانند همانندجویی و رصد وضعیت پژوهش کشور و ...) از آنها ایجاد نموده است، اما روندهای علمی و پژوهشی در حوزه‌های علم در ایرانداک تا کنون مورد توجه قرار نگرفته است. لازم به ذکر است روند هر جستجو درباره یک کلیدواژه خاص در نمودارهای سامانه گنج قابل مشاهده است، اما کشف روندهای یک حوزه علمی مانند فناوری اطلاعات و مقایسه آن با روندهای جهانی فعالیت زیادی را می‌طلبد.

مدلسازی موضوعی یک رویکرد خوشه‌بندی شناخته شده برای اطلاعات متنی هستند که به منظور طبقه‌بندی موضوعات موجود در حوزه دانشی قابل استفاده است و با بکارگیری آن امکان ارائه اطلاعات علم‌سنجی مربوط به شاخه‌ها و زیرشاخه‌های یک حوزه دانشی فراهم می‌شود و ارائه یک ابزار عملی برای پژوهش در زمینه توسعه دانش است. به گفته وی و کرافت (۲۰۰۶) با استفاده از روش‌های مبتنی بر خوشه‌بندی و توانایی استفاده از مدل‌سازی موضوعی در تحلیل علم‌سنجی در آینده نزدیک روش‌های جدیدی برای تجزیه و تحلیل داده‌ها توسعه داده خواهد شد.

الگوریتم‌های مدل‌سازی موضوعی شامل تکنیک‌های آماری است که هدف آنها توصیف موضوعاتی است که در اسناد منتشر شده مانند مجلات منتشر شده‌اند. آنها بر اساس تجزیه و تحلیل آماری کلمات

[4] Thesaurus

در آن اسناد، شناسایی خوشه‌هایی از کلمات مشترک، تشخیص موضوعات مورد بحث، روابط بین این موضوعات و تکامل آنها در طول زمان کشف می‌کنند. چندین مدل مانند تجزیه و تحلیل معنایی معکوس[۵] (لاندر و همکاران، ۱۹۹۸)، تجزیه و تحلیل معنایی بالقوه احتمالی[۶] (هافمن 1999) و تخصیص نهفته دیریکله[۷] (بلی و همکاران ۲۰۰۳)؛ و همچنین برخی از مشتقات آنها، مانند پاچینو[۸] (لی و مک‌کالم 2006) و یا مدل‌سازی موضوعی رابطه‌ای (چانگ و بلی 2009) در این زمینه کاربرد دارند.

مدل‌سازی موضوعی نوعی از مدل آماری برای کشف موضوعات "نهفته [۹]" است که در مجموعه‌ها از اسناد از طریق یادگیری ماشین رخ می‌دهد. در حال حاضر، تخصیص نهفته دیریکله یک رویکرد مدل‌سازی کارا و رایج در این زمینه است.

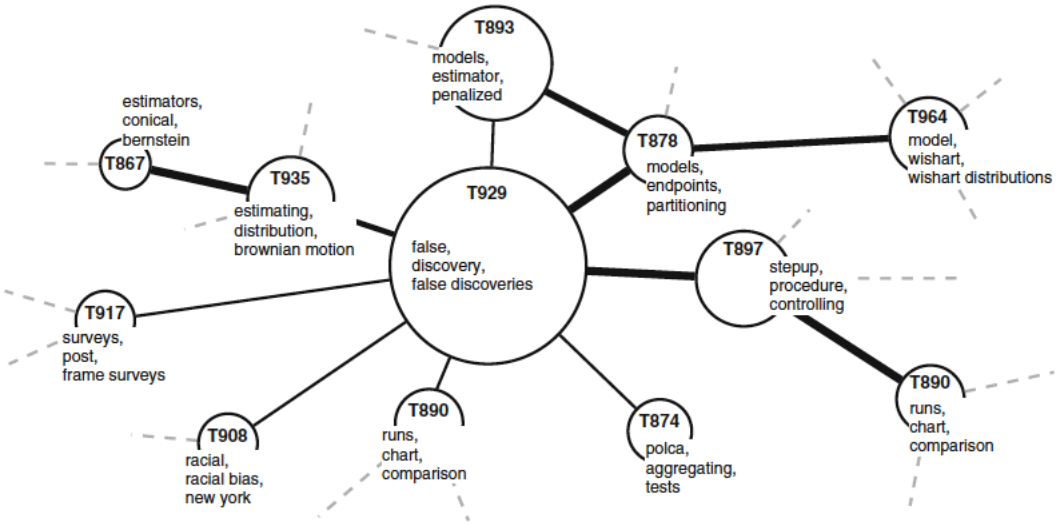
یکی از کاربردهای استفاده از مدلسازی ذکر شده به منظور ارائه سطح بالا از مفاهیم، کلمات کلیدی آن حوزه و زمینه‌هایی که مربوط به موضوع مورد نظر در یک دوره زمانی برای یک زمینه خاص پژوهش مانند فناوری اطلاعات است، ارائه یک نقشه از مجموعه واژگان کلیدی آن حوزه است. بطور مثال می‌توان زمینه‌هایی مربوطه را در ۱۰ سال گذشته محدود به چند مجله پیشرو در حوزه فناوری اطلاعات شناسایی تحلیل نمود. برای به دست آوردن این نتیجه، ابتدا باید هر موضوع با یک توصیفگر مرتبط شده که این توصیف می‌تواند شامل اطلاعات مربوط به کلمات کلیدی مرتبط، برخی اطلاعات آماری مفید درباره انتشار، توزیع مجلات مقالات در هر موضوع، توزیع مقالات موضوعی در هر سال، کشور و تعداد استنادات باشد. در فعالیتی که توسط دکتر جلالی‌منش در ایرانداک (۲۰۱۲) صورت پذیرفته بود، ۶۵۰ عنوان پایان‌نامه در زمینه مهندسی صنایع با الگوریتم‌های متن‌کاوی و تحلیل شبکه اجتماعی با هدف استخراج حوزه مهندسی صنایع، مورد پایش قرار گرفت. در این مطالعه ۱۵۰۰ واژه به‌عنوان واژه‌های مرتبط با مهندسی صنایع انتخاب شدند. در این مطالعه از روشی متفاوت با فعالیت این طرح بهره گرفته شده و در یک مرحله واژگان استخراج شده‌اند (فرایند طراحی شده در این پژوهش ۲ مرحله‌ای است)، ولی می‌توان از تجربیات مرتبط با پاکسازی فهرست لغات استخراج شده و تجربیات مرتبط با مصورسازی نتایج بهره جست.

به منظور ارائه دیدگاه جامع از مباحث علمی مطرح شده در حوزه فناوری اطلاعات توسط مجموعه‌ها از مقالات، مفهوم نقشه موضوعی مطرح می‌شود. نمونه این ارائه در ادبیات مرتبط در حوزه فناوری

---

[5] LSA
[6] PLSA
[7] Dirichlet (LDA)
[8] Pachinco
[9] Latent

توسط نویسندگان این پیشنهاده یافت نشده است. در صورت مشخص بودن حوزه و روند بین‌المللی در ادبیات، این پیشنهاده می‌توانست از نتایج آن استفاده و فقط مقایسه را در راستای پژوهش‌های ملی انجام دهد. اما به‌دلیل فقدان موضوع و همچنین سرعت رشد بالای موضوع فناوری اطلاعات (امکان ایجاد تغییرات در زمان کوتاه)، در این پیشنهاده ملزم به بررسی و استخراج در هر دو حوزه داخلی و بین‌المللی هستیم. یک نقشه موضوعی یک گراف است که گره‌ها، موضوع‌ها را نشان می‌دهند و ارتباطات نمایانگر روابط بین موضوعات می‌باشند. علاوه بر این، هر گره از نقشه موضوعی می‌تواند به صورت گرافیکی توسط یک دایره که حاوی مقادیر مناسب از کلیدواژگان موضوع مربوطه است، نمایان شود. قطر دایره، متناسب با تعداد رخدادهای آن واژگان در موضوع و حوزه و متناسب با تعداد مقالاتی است که در این موضوع ارایه داده شده است. همچنین موقعیت گرافیکی گره‌های مربوطه، طوری مشخص می‌شود تا موضوعات مشابه نزدیک به یکدیگر نشان داده شوند. عرض خط ارتباطی متناسب با میزان شباهت دو موضوع است. به عنوان نمونه می‌توان به شکل زیر اشاره نمود.



شکل ۱- نمونه توسعه چارچوب فناوری اطلاعات (باتیستی و همکاران ۲۰۱۵)

با توجه به توضیحات فوق، در این پژوهش، مدل‌سازی موضوعی بر روی مجموعه‌ای از مجلات بین المللی، با رویکرد تشریح شده انجام پذیرفته و پس از پردازش نتایج توصیفات، عمق فعالیت‌های پژوهش‌های مربوطه از این زمینه در طول سال‌های مختلف مورد بررسی قرار می‌گیرد. این پژوهش روند کلی پژوهش‌های حوزه فناوری اطلاعات در پایان‌نامه‌های و رساله‌های تحصیلات تکمیلی کشور (برگرفته از سامانه گنج) را با روند کلی پژوهش‌های مربوط به فناوری اطلاعات در سطح جهان بررسی نموده و به مقایسه تفکیکی دانشگاه‌های کشور با دیگر تفکیک‌ها پرداخته نمی‌شود. یکی از دلایل انتخاب این مساله دسترسی به این منابع در بعد داخلی و بین‌المللی است. در تعریف یک مساله

عملیاتی، یکی از مهمترین چالش‌ها، دسترسی مناسب به اطلاعات مورد نیاز است. پژوهشگران این پیشنهاده به پایان‌نامه‌های بین‌المللی مناسب در زمینه فناوری اطلاعات دسترسی ندارند حال اینکه به مجلات پایگاه ذکر شده دسترسی دارند. این مسله خللی در اجرای این پژوهش ایجاد نمی‌نماید زیرا یکی از بزرگترین دستاوردهای این پژوهش انجام روشمند این مدل در پژوهشگاه می‌باشد. در صورتی که پژوهشگاه به بانک اطلاعات غنی‌تری در آینده دسترسی داشته باشد می‌تواند این پژوهش را در آن پایگاه تکرار نماید. در بعد خروجی نیز مقایسه بین تحقیقات ملی و مقالات بین‌المللی بسیار ارزشمند است. بطور مثال می‌توان فاصله زمانی (شکاف) میان مطالعات ملی را با پژوهش‌های مجلات پیشرو کشف کنیم.

به طور خاص، چند رویکرد متفاوت در این پژوهش پیاده سازی می‌شود:

(۱) به منظور مطالعه موضوعات فناوری اطلاعات در سطح بین‌المللی در حال حاضر، موضوعات با تحلیل عناوین (و کلیدواژهای) تمامی مقالات چند مجله پیشرو در چند سال اخیر شناسایی می‌شوند.

(۲) در این فاز سبد واژگان شکل می‌گیرد. این سبد شامل کلمات مطرح شده در عنوان مقالات و کلیدواژه‌های مقالات منتخب پایگاه‌های مورد پژوهش در زمینه فناوری اطلاعات است. در مرحله اولیه سبد واژگان شامل کلمات عمومی (و غیر مرتبط با فناوری اطلاعات مانند a , the , an , ...) نیز می‌باشد که باید پاکسازی گردند. تمام کلمات این سبد اولیه واژگان مجددا باید مورد جستجو در پایگاه‌ها قرار گیرد تا سبد کاملتر و غنی‌تر و میزان کاربرد و روند استفاده هر واژه در سبد یافت شود. همانگونه که ذکر شد، سبد واژگان اولیه (که نشان‌دهنده بخش بزرگی از حوزه مربوطه هستند و می‌توانند پژوهش‌های حوزه فناوری اطلاعات را بازیابی نمایند) مجدد در پایگاه‌های جهانی انتخاب شده (مجلات بین‌المللی WOS و کنفرانس‌های معتبر) مورد جستجو قرار می‌گیرد و واژه‌های کلیدی جدید (از موضوع و چکیده وکلیدواژه‌های مطالعات انجام شده) مجدا استخراج شود، زیرا در این پایگاه به‌صورت ماشینی قابلیت استخراج اطلاعات فرادادههای مقالات منتشر شده در طول بازه زمانی چندین سال میسر است ولی پایگاه‌های دیگر این امکان را در اختیار پژوهشگران قرار نداده و استخراج دستی حجم بالای مقالات زمانبر است. با این گام جامعیت علمی واژه‌های کلیدی تضمین و سبد واژگان غنی می‌شود.

(۳) به‌منظور مطالعه تکامل و روند پژوهش‌ها در حوزه فناوری اطلاعات، مقالات استخراج شده با توجه به کلیدواژه‌ها، سال انتشار و ... تحلیل موضوعی و محتوایی (مدلسازی موضوعی) می‌شوند.

(۴) با توجه به ثبت کلید واژه انگلیسی پارساهای کشور در سامانه گنج، سبد واژه‌گان کلیدی نهایی کشف شده در فاز دوم در پژوهش‌های داخل کشور (کلیدواژه‌های انگلیسی ثبت شده، واژه‌های

عنوان انگلیسی و واژه‌های چکیده انگلیسی) مجددا مورد بررسی قرار می‌گیرند و روند آنها در پارساهای ثبت شده در سامانه گنج استخراج می‌شوند (مدلسازی موضوعی).
(۵) مقایسه و تحلیل حوزه و روند میان خروجی فازهای سوم و چهارم صورت می‌پذیرد.

۱-۲. اهداف طرح

از اهداف نهایی طرح جاری می‌توان به موارد زیر اشاره نمود:

- مقایسه روندهای حوزه‌های فناوری در سطح بین‌المللی
- کشف روند حوزه فناوری اطلاعات در پارسا[۱۰]های کشور (محدود به سامانه گنج) با استفاده از کلیدواژه‌ها
- مقایسه روندهای حوزه‌های فناوری در سطح بین‌المللی و پارساهای کشور و شناسایی شکاف

پرسش‌های پژوهش جاری در سه دسته طبقه‌بندی می‌شوند:

۱. سبد واژگان حوزه فناوری اطلاعات در سطح ملی (پارساهای ثبت شده در سامانه گنج) شامل چه واژگانی است و روند پژوهش‌ها چگونه بوده است؟
   - سبد واژگان حوزه فناوری اطلاعات در سطح بین‌المللی چگونه است؟
   - با جستجوی سبد فوق در سطح ملی (جستجو در پارساهای ثبت شده در سامانه گنج) به چه نتایجی در توسعه حوزه فناوری اطلاعات در سطح ملی (پارساهای کشور) خواهیم رسید؟
   - روند کاربرد این واژگان در سطح ملی چگونه بوده است؟
۲. روند پژوهش‌های پارساهای کشور در حوزه فناوری اطلاعات با روند حوزه مربوطه در سطح بین‌المللی چقدر تطابق دارد؟
   - روند کاربرد واژگان در سطح ملی و میزان آنها در مقایسه با سطح بین‌المللی چگونه است؟

---

[۱۰] – پایان‌نامه و رساله

۱-۳. پیوند با اساس‌نامه و برنامهٔ استراتژیک

با توجه به عزم پژوهشگاه بر انجام تحقیقات مرتبط با فعالیت‌ها و اهداف پژوهشگاه به منظور ایجاد یکپارچگی و اثربخشی تحقیقات، همسویی تحقیق جاری با اساسنامه و اسناد بالادستی پژوهشگاه در زیر آمده است:

- ماده ۱ اساس‌نامه:
  - بند ۴: زمینه‌سازی مناسب برای ارتقای فعالیت‌های پژوهشی مرتبط با مدیریت دانش، جامعه اطلاعاتی، علوم و فناوری اطلاعات و نوآوری در موضوعات یادشده؛
- ماده ۲ اساس‌نامه:
  - الف: فعالیت‌های پژوهشی، بند ۲. **بررسی و شناسایی نیازهای پژوهشی در زمینه علوم و فناوری اطلاعات؛**
- در راستای ماموریت و برطبق برنامه استراتژیک ایرانداک:
  - بند ۲: در مدیریت اطلاعات علم و فناوری کشور مشارکت می کند؛
  - بند ۷: **دستاوردها و یافته‌های علمی و فنی کشور و جهان را در زمینهٔ علوم و فناوری اطلاعات منتشر می‌کند؛**
- کار کلیدی ۱ در استراتژی ۲:
  - ارتقای آگاهی‌سیاست‌گذاران
- وظایف **اساسنامه بند ج:** توسعه کاربرد فناوری اطلاعات در سطح وزارت و کشور (ایرانداک ۱۳۹۸)

مفاهیم فناوری اطلاعات و کاربردهای آن در سطح وزارت عتف و کشور توسعه یافته است. به‌عنوان ذکر یک مثال توسعه بخشی از فناوری اطلاعات که در زنجیره مدیریت اطلاعات شامل فراهم‌آوری، سازماندهی و آرشیو، تحلیل اطلاعات و تولید دانش، اشاعه و شبکه‌سازی پررنگ است. در پژوهشگاه نشان‌دهنده توسعه متوازن این حوزه در کسب و کار مدیریت اطلاعات علمی پژوهشگاه است و به‌عنوان نمونه، جهت‌گیری مناسب این توسعه در طرح‌های پژوهشی زیر مشاهده می‌شود:

- طراحی نظام کدگذاری موجودیت‌های دیجیتال
- فرایند کنترل کیفیت دیجیتال‌سازی (اسکن)
- الگوی تولید نسخهٔ الکترونیکی مدارک
- نیازسنجی و مدلسازی نیازهای اطلاعاتی

- طراحی فرایند تولید پایگاه‌های اطلاعات علمی
- بررسی یکپارچه‌سازی معنایی اطلاعات در کتابخانه‌های دیجیتال و ارائه الگویی برای کتابخانه‌های دیجیتال ایران
- فرایند کنترل کیفیت نمایه‌سازی
- طراحی فرایند تولید پایگاه‌های اطلاعات علمی
- طراحی یک معماری مفهومی برای پردازش داده‌ها برمبنای مدل‌های رایانش ابری
- تدوین الگوی مفهومی سامانه ملی مکانی امایش منابع اطلاعات علمی و پژوهشی کشور
- سیستم مدیریت خدمات پایگاه‌های اطلاعات

این همراستایی در توسعه فناوری اطلاعات دیگر سازمان‌ها به ندرت به چشم می‌خورد. بنابراین ارائه چارچوبی که بتوان حوزه فناوری اطلاعات را شناسایی نمود ضروری می‌نماید. این هدف کاربرد فناوری اطلاعات در سطح وزارت و کشور و سازمان‌ها و شرکت‌های مختلف کشور را در پی خواهد داشت.

این پیشنهاده با طرح‌های پژوهشی پیشین پیشنهاددهنده

- طرح پژوهشی بررسی کارکردهای مدیریت اطلاعات علمی و فناورانه در بافت دولت الکترونیکی و دولت همراه، آرمان پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) ، آرمان ساجدی نژاد، ۱۳۹۶
- طراحی چارچوب مفهومی بومی تعالی فناوری اطلاعات سازمانی، آرمان ساجدی نژاد، ۱۳۹۷
- بازطراحی کارکرد مدیریت اطلاعات علمی و فناورانهٔ پژوهشگاه علوم و فناوری اطلاعات ایران در بافت دولت الکترونیکی، آرمان ساجدی نژاد، ۱۳۹۵

و همچنین با تعدادی از پژوهش‌های پیشین مانند

- محبی، آ. (۱۳۹۶). طرح پژوهشی تدوین نقشه راه فناوری اطلاعات پژوهشگاه با تمرکز بر پیشرفتهای نوین در حوزه تجزیه و تحلیل اطلاعات. پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)
- نامداریان، ل. (۱۳۹۵). طرح پژوهشی بررسی و تبیین جایگاه پژوهشگاه علوم و فناوری اطلاعات در نظام سیاستگذاری علم و فناوری کشور. پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)

ارتباط و پیوند خواهد داشت.

۱-٤. تعریف واژه‌ها و مفاهیم

- مجلات پیشرو: مجلاتی که از لحاظ ضریب تاثیر و یا سابقه در حوزه علمی فناوری اطلاعات دارای بالاترین رتبه هستند
- کلیدواژه: کلید واژه‌های مقالات منتخب در مجلات پیشرو
- سبد واژگان: مجموعه واژگان استخراج شده از عناوین و کلیدواژه مقالات مورد مطالعه در حوزه فناوری اطلاعات
- پارسا: پایان نامه و رساله‌ی ثبت شده در زمینه فناوری اطلاعات در سامانه گنج و دارای کلیدواژه‌های انگلیسی
- مدل‌سازی موضوعی: مدل‌سازی کلمات در متون علمی با کشف درجه احتمال پراکندگی

فصل دوم

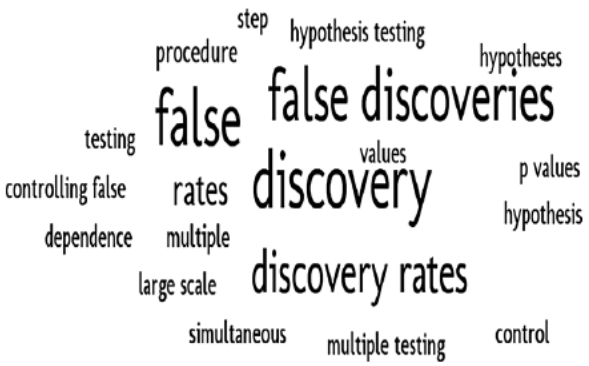
مدل‌سازی موضوعی حوزه فناوری اطلاعات

۲-۱. مقدمه

همانطور که در فصل قبل ذکر گردید، در مدل‌سازی موضوعی، ارتباط میان کلمات و زمینه‌های معنایی نهفته در اسناد استخراج می‌شود (بلی و لفرتی 2007). در ادبیات (هال و همکاران 2008 ؛ بلی 2012) همواره بین نیاز به تعداد زیادی از واژگان برای پوشش همه زمینه‌های یک حوزه علمی در مجموعه اسناد موضوعی چالش وجود دارد. (گرین و هورنیک ۲۰۱۱). همانطور که در همه رشته‌ها رایج است، مجلات زیادی فناوری اطلاعات را از دیدگاه‌های مختلف بررسی می‌کنند. رایج ترین معیارهای بخش‌بندی ادبیات و مجلات عبارتند از: تعریفی یا نظری بودن مطالعات آنها، ارتباط نام مجله با موضوعات، مرزها و زیر شاخه‌های زمینه مورد نظر (بروک 2007 ، بکلند ۲۰۱۲ و فرنر 2015). بسیاری از مطالعات منتشر شده در مجلات دامنه محدودی دارند، که بر منابع خاص، موضوع یا تخصصی ویژه متمرکز شده‌اند. بنابراین، بسیاری از پژوهش‌ها در زمینه کشف حوزه‌های فعالیت بر روی یک مجله متمرکز شده است (به‌عنوان نمونه مطالعه هارت و هوتن (۱۹۹۲) و تأکید وابستگی‌های نهادی نویسندگان یا (سوین و همکاران 2012)). در عین حال، برخی دیگر از مطالعات شناخت حوزه‌های مفهومی (مانند فناوری اطلاعات) بر مجموعه‌ای از کنفرانس‌ها تمرکز می‌کنند (مانند مقاله اسمتون و همکاران (۲۰۰۲) در مورد بازیابی اطلاعات، و یا مدل‌سازی موضوعی با استفاده از تکنیک‌های تجزیه و تحلیل شبکه اجتماعی (SNA) توسط ساگیموتو و مک‌کین 2010)). گاهی اوقات، نویسندگان مجبورند در مدل‌سازی موضوعی فقط به نوعی از مجلات یا یک مجله خاص متمرکز شوند، بدون اینکه تلاش کنند تا کل زمینه علمی را به طور جامع پوشش دهند. چنین موردی در مورد مطالعه موخرجی (2009) با تجزیه و تحلیل ۱۷ مجله دسترسی آزاد در حوزه مربوطه صورت پذیرفت. در موارد دیگر تمرکز بر روی نوع پژوهش است؛ بطور مثال پژوهش کیم و جونگ (۲۰۰۶) در مورد پژوهش‌های نظری است. یکی دیگر از دیدگاه‌های کمی متمایز ارائه شده در مطالعات،

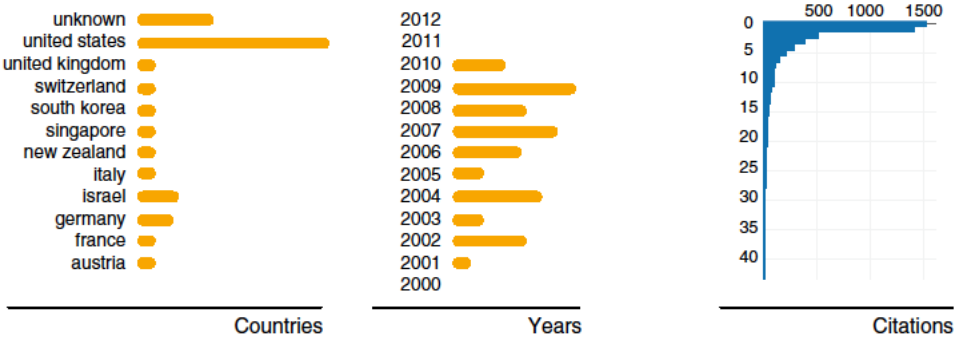
استفاده از روش‌های تجزیه و تحلیل محتوا است. این روشها بر تجزیه و تحلیل کلمات کلیدی متمرکز هستند و کلمات در عنوان یا چکیده و یا حتی متن کامل را مورد تحلیل قرار می‌دهند (مانند فعالیت علمی آستروم (2002)). اما همانطور که ذکر گردید، فعالیت مشابه در زمینه شناسایی حوزه فناوری اطلاعات در ادبیات انجام در این طرح پژوهشی یافت نشده است.

به منظور ارائه نمونه‌ها از توصیف موضوع، به ارایه نمونه‌ای از مدل‌سازی موضوعی در حوزه آمار اشاره می‌نماییم. با توجه به رویکرد توصیف شده در بخش‌های قبلی این طرح، در این پژوهش توصیفگرهای موضوع با تعیین کلیدی‌ترین کلمات کلیدی و منبع انتشار، سال و توزیع درکشورها، محاسبه شده‌اند. دو مجله مورد بررسی این پژوهش در بازه ۱۰ ساله به طور خاص، سالنامه آمار و JASA بوده‌اند. توصیفگرها همچنین می‌توانند برای ارائه یک نمایش گرافیکی از یک موضوع، با توجه به رویکرد بیان شده استفاده شوند. یک موضوع به عنوان یک متن نمایش داده می‌شودکه بزرگی و قطر آن متناسب با تعدادی از مقالات شامل آن موضوع است (شکل زیر).



شکل ۲- نمایش کلمات پرکاربرد در حوزه آمار (باتیستی و همکاران ۲۰۱۵)

کلمه گسترش و کشف (Discoveries) در طی سال‌ها نوسان داشته و در طول زمان با افزایش اوج خود در سال ۲۰۰۹ افزایش یافته است. بیشترین فعالیت در این زمینه کشور ایالات متحده است.



شکل ۳-میزان کاربرد یک کلیدواژه (باتیستی و همکاران ۲۰۱۵)

با استفاده روزافزون از پایگاه داده‌های متنی ساختار یافته، رویکردهای تازه در زمینه اندازه‌گیری علم بیشتر بر روش‌های کمی تکیه می‌کنند که اغلب به عنوان کتابشناختی نامیده می‌شوند (برگمن و فرنر 2002). در این پژوهش‌ها و مطالعات از ابزارهای متدولوژیکی مختلف برای استخراج اطلاعات از پایگاه‌های داده استفاده شده و پژوهشگران تلاش دارند تا ساختارهای زیرین را درون مجموعه‌ای از داده‌ها کشف کنند (دایم و همکاران 2006). مطالعات با استفاده از روش‌های مختلف ابتدایی (سومینن 2013) تا با بهره‌گیری از روش‌های مختلف برای تعیین تعداد کشورها، نویسندگان یا فناوری‌های ذکر شده در یک سال توسعه یافته و این متغیرها با روش‌های متن‌کاوری تحلیل می‌شوند (گلیسون و همکاران 2005) .

رویکرد استخراج متن به دنبال یافتن کلمات یا عباراتی است که می‌تواند محتوا و ساختار (روابط) موجود در داده‌ها را توضیح دهد. شناسایی مشترکات در متن، متمرکز بر تجزیه و تحلیل داده‌ها و توزیع آنها در متن و ارتباطات با استفاده از رویکردهای مختلف خوشه‌بندی است (فلدمن و سانجر 2006). در فعالیت هافمن (1999) ادعا شده است که مدل‌سازی موضوعی ابزار مفید برای استخراج اطلاعات از داده‌های متنی است. بعدها، وی و کرافت (2006) نشان دادند که در بازیابی اطلاعات، مدل‌سازی‌های موضوعی بیشتر از روش‌های سنتی، مبتنی بر خوشه‌ها عمل می‌کنند.

اگر چه بخشی از ادبیات در حال حاضر در زمینه مدلسازی موضوع توسعه یافته است، اما این ادبیات عمدتا بر متدولوژی و توسعه آن متمرکز است. پژوهشگران بازیابی اطلاعات در اواخر دهه ۱۹۹۰، اولین مدل احتمالی با نام احتمال نامگذاری مستقل معنایی معکوس (pLSI) را پیشنهاد دادند. مدل‌های شاخص نامتقارن احتمالی، هر کلمه در یک سند را به عنوان نمونه‌ای از یک مدل بزرگ مطرح می‌کند که اجزای آن متغیرهای تصادفی چندجمله‌ای هستند که می‌توانند به عنوان نمایه‌ای از موضوعات در نظر گرفته شوند. مشکل در الگوریتم pLSI این است که تعداد پارامترها به صورت خطی با تعداد موضوع رشد می‌کنند.

برای حل این مشکل، یک مدل بهبود یافته، تخصیص نهایی دیریکله[1] پیشنهاد شده است (بلی و همکاران، ۲۰۰۳). این رویکرد یک مدل سه بعدی است که نقاط ضعف pLSI را برطرف می‌کند و در حال حاضر به طور کاربردی گسترده‌ای در مسایل مختلف مانند استخراج متن، بیوانفورماتیک و پردازش تصویر مورد استفاده قرار می‌گیرد. علاوه بر این، مدلسازی‌های موضوعی ارتباطی (CTM)، تخصیص سلسله مراتبی نهفته دیریکله (LDA سلسله مراتبی) و فرایند سلسله مراتبی دیریکله (HDP) در سال‌های اخیر مطرح و توسعه داده شده است.

[1] - Dirichlet (LDA)

مدل‌سازی موضوعی نیز برای رسیدن به سری زمانی در متن (یان و همکاران 2012؛ استیورز و همکاران 2004) گسترش یافته است و روش‌های عملی برای تحلیل ساختارهای متون علمی ارایه داده‌اند. نی و همکاران (۲۰۱۲)، مدعی شده‌اندکه کارایی این روش‌ها به شدت به توانایی الگوریتم مدلسازی موضوعی برای تمایز بین موضوعات مختلف بستگی دارد. اخیرا شاهد آن هستیم که مدل‌سازی موضوعی در حال تبدیل شدن به روشی متداول است و پژوهشگران بسیاری از این الگوریتم‌ها استفاده می‌کنند.

در مقاله‌ای دیگر از یاو و همکاران (۲۰۱۴)، روش‌هایی از جمله LDA برای جداسازی مجموعه‌ای از انتشارات علمی به چندین خوشه بررسی می‌شود. برای ارزیابی نتایج، مجموعه‌ای از اسناد که حاوی مقالات علمی از چندین حوزه مختلف است، ایجاد شده و بررسی می‌شود که آیا مقالات در همان حوزه با یکدیگر همخوانی خواهند داشت. برنامه‌های کاربردی علم‌سنجی از جمله توانایی تجزیه و تحلیل متن در این پژوهش مورد بررسی قرار می‌گیرد. در این مقاله، مسئله جداسازی یک مجموعه سند علمی به چند دسته بر اساس محتویات آنها بر اساس مدل‌سازی موضوع در نظر گرفته شده است. برای این منظور، دو نمونه از داده‌ها توسط پژوهشگر ساخته شده است. هدف این است که از مدل‌سازی موضوعی به عنوان یک روش نسبتا خودکار استفاده شود که خوشه‌های مستندات حاوی اسناد مشابه ایجاد کند و سپس نتایج با ساختار واقعی داده‌ها مقایسه شود. با انجام این کار، توانایی روش مدل‌سازی موضوعی برای تشخیص ساختارها از متن و سنجش موفقیت آن با دو عامل دقت و فراخوانی ارزیابی می‌شود. در زمینه یادگیری ماشین، روش ساده‌ای برای جمع‌آوری داده‌ها، استفاده از الگوریتم‌های طبقه بندی شده است. با این حال، از آنجا که نشانه‌گذاری مجموعه‌ای از اسناد می‌تواند زمان‌بر و ناکارآمد باشد، در این مقاله ترجیح داده شده است تا مدل‌سازی موضوعی برای حل این مشکل استفاده شود. داده‌ها قبل از مدل‌سازی موضوعی پردازش می‌شوند. در ابتدا، داده‌ها تمیز می‌شود تا فقط از عنوان، کلیدواژه و چکیده‌ی هر سند استفاده شود. از دو فرایند مختلف برای تمیز کردن داده‌ها استفاده می‌شود:

(۱) داده‌ها به صورت نشانه‌های متشکل از کلمات استفاده شده، یا

(۲) استخراج نشانه با استفاده از پردازش زبان طبیعی (NLP).

بر این اساس ابتدا هر کلمه را به عنوان یک نشانه در نظر گرفته و سپس کلمات در لیست توقف نوشتاری ارائه شده در نرم افزار MALLET (مک‌کالم 2002) پالایش می‌شوند (نیومن و همکاران ۲۰۱۲). مزیت این روش پیش پردازش این است که می‌تواند شرایط معنی دارتری را تولید کند، به عنوان مثال، سلول خورشیدی نباید به یک واژه مجزای خورشیدی و سلول تقسیم شود و باید به عنوان

یک واژه‌ی دو کلمه‌ای مورد استفاده قرار گیرد. نتایج هر دو فرایند به طور جداگانه با الگوریتم‌های مدل‌سازی موضوعی مورد تجزیه و تحلیل قرار می‌گیرد. همانطور که نتایج پژوهش‌ها نشان می‌دهد، خوشه‌بندی K-means نیز مانند الگوریتم مدل‌سازی موضوع عمل می‌نماید.

با این حال، مدل‌سازی موضوعی نتایج با ثبات‌تری تولید می‌کند و در نتیجه بهتر عمل می‌کند. این نتایج در شکل زیر خلاصه شده است (یاو و همکاران ۲۰۱۵).

	F score average	*F* score st. dev.
K-means	0.66	0.30
LDA	0.71	0.04
CTM	0.72	0.07
hLDA	Not computed	Not computed
HDP	0.90	0.04

شکل ٤- مقایسه روش‌های مختلف خوشه‌بندی

نتایج نشان می‌دهد که مدل‌سازی موضوعی در مرحله فعلی بلوغ، نیاز به توسعه دارد تا دقت را بهبود بخشد (یاو و همکاران ۲۰۱۵).

در پژوهش انجام شده توسط فیگورلا و همکاران (۲۰۱۷) یک مرور کلی از مطالعه کتابشناختی حوزه کتابخانه و اطلاع رسانی (LIS) با هدف ارائه دیدگاه چند رشته‌ای از مرزهای موضوعی و زمینه‌های اصلی و گرایش‌های پژوهش ارائه شد. بر اساس پژوهش گذشته نگر و انتخابی، در سال‌های ۱۹۷۸ تا ۲۰۱۴ میلادی، کتابشناسی (عنوان و چکیده) تولید علمی در پایگاه داده LIS منجر به استخراج ۹۲۷۰۵ سند شد. در زمینه تکنیک آماری مدل‌سازی موضوعی، از طریق تکنیک تخصیص دریکله به منظور شناسایی موضوعات نهفته و دسته‌های اصلی در مجموعه اسناد مورد استفاده و تجزیه و تحلیل قرار گرفت. نتایج کمی وجود ۱۹ موضوع مهم که می‌تواند به چهار بخش اصلی فرایندها، فناوری اطلاعات، کتابخانه و برنامه‌های اطلاعاتی تقسیم شود را نشان داد.

۲-۲. روش مدل‌سازی موضوعی

مدل‌سازی موضوعی بر اساس این ایده تشکیل شده است که اسناد، ترکیبی از مباحث هستند و موضوع (عنوان و کلیدواژه یا فقط عنوان) آنها با یک توزیع احتمالی بیش از دیگر کلمات مورد استفاده قرار داده می‌شود. در این مدل‌ها موضوعات (و توزیع آنها) به عنوان ساختار نهفته یا متغیرهای نهفته در نظر گرفته می‌شوند. الگوریتم‌های مدل‌سازی موضوعی روش‌های آماری هستند که کلمات اصلی اسناد را تجزیه و تحلیل می‌کنند تا مباحثی را که از طریق آنها ارایه می‌شود تحلیل نموده و چگونه اتصال این موضوعات را به یکدیگر و چگونگی تغییر آنها را با گذر زمان کشف نمایند.

یک موضوع به عنوان یک توزیع در یک واژگان ثابت تعریف شده است. به عنوان مثال موضوع اینترنت اشیا دارای کلمات در حوزه فناوری اطلاعات با احتمال بالا است. این مدل فرض می‌کند که موضوعات قبل از اسناد (مقالات) تولید می‌شوند. برای هر سند (عنوان و/یا کلیدواژه)، کلمات در یک فرایند دومرحله‌ای تولید می‌شوند:

- به طور تصادفی یک توزیع برای موضوع انتخاب می‌شود (توزیع دریکله)؛
- برای هر کلمه آن موضوع، ابتدا به صورت تصادفی یک موضوع از توزیع موضوعات فاز قبل انتخاب شده و سپس به طور تصادفی یک کلمه از توزیع متناظر را در فهرست لغات انتخاب می‌شود.

مشکل اصلی بهره از مدل‌سازی موضوعی، نیاز به بالا بودن تعداد رکوردها و کلیدواژه‌ها برای بالاتر بودن صحت مدل است. بدین‌منظور در پژوهش جاری به استفاده از اسناد ۱۰ سال مقالات ۵ مجله پیشرو (ذکر شده در پیشنهاده) برای به دست آوردن متغیرهای نهفته محدود نشده و مجلات بیشتری با توجه به نظر خبرگان انتخاب شده‌اند. خروجی مدلسازی موضوعی مدل‌های احتمالی هستند که در آن داده‌ها یک نتیجه از یک فرایند آماری هستند که شامل متغیرهایی نهفته است. این فرایند توزیع احتمالی مشترک را در هر دو متغیر تصادفی مشاهده شده و نهفته تعریف می‌کند.

شمارنده توزیع شرطی توزیع مشترک تمام متغیرهای تصادفی است که می‌تواند به راحتی محاسبه شود و از لحاظ تئوری، می‌توان آن را با تجمیع توزیع مشترک تمامی ساختارهای نهفته ، محاسبه کرد. عملا، به دلیل اینکه تعداد ساختارهای موضوع احتمالا زیاد است، این محاسبه دشوار و زمان‌بر است.

الگوریتم‌های مدل‌سازی موضوعی به دو دسته تقسیم می‌شوند که توزیع‌های جایگزین دیگری را پیشنهاد می‌دهند تا تقریبا الگوریتم‌های مبتنی بر نمونه‌برداری را بهبود بخشند:

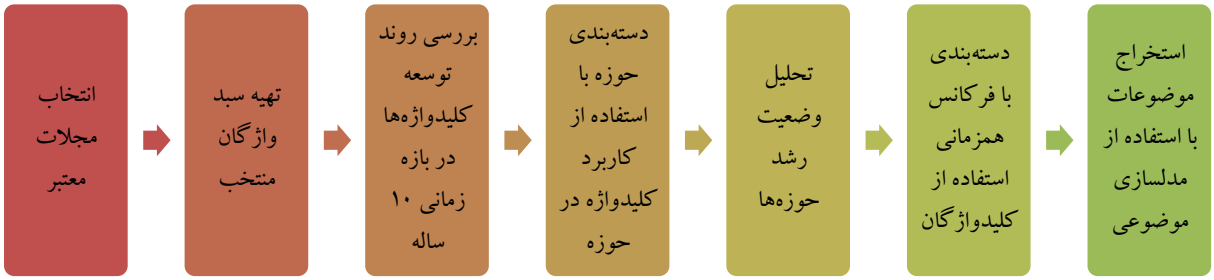
- گروه اول ارایه یک زنجیره مارکوف (یک توالی از متغیرهای تصادفی)،
- گروه دوم، الگوریتم‌های مبتنی بر نمونه‌گیری (VEM)[1].

با توجه به نوع داده‌های انتخاب شده (عنوان مقالات و/ویا کلیدواژگان) در این طرح، الگوریتم VEM انتخاب خواهد شد، زیرا الگوریتم VEM برای هر مقاله، همپوشانی بین موضوعات مهم را ارایه می‌دهد و کشف همبستگی‌های موضوعی یکی از اهداف اصلی این طرح (کشف مرزهای حوزه فناوری اطلاعات) است.

[1] - Variational expectation-maximization (VEM) algorithm

۲-۳. روش پژوهش

روش پژوهش جاری در فاز اول به مراحل زیر انجام خواهد پذیرفت:



شکل ۵- مراحل پژوهش جاری

در جدول زیر مراحل مختلف شکل فوق تشریح می‌شوند:

جدول ۱- شرح روش پژوهش

مرحله	روش
انتخاب مجلات معتبر	جستجوی مجلات برتر در حوزه فناوری اطلاعات
	انتخاب مجلات به نحوی که تمامی حوزه‌های تعریف شده را پوشش دهند
	استفاده از نظر تکمیلی خبرگان برای معرفی مجلات دیگر
تهیه سبد واژگان منتخب	بررسی کلیدواژگان مجموعه مجلات و فرکانس استفاده در بازه زمانی کوتاه
	انتخاب کلیدواژه‌های منتخب توسط خبرگان
بررسی روند توسعه کلیدواژه‌ها در بازه زمانی ۱۰ ساله	جستجو در تمامی پایگاه بدون محدودیت حوزه فناوری اطلاعات در باز زمانی ۱۰ ساله
	بررسی چگونگی گزینش متون خروجی جستجو
دسته‌بندی حوزه با استفاده از کاربرد کلیدواژه در حوزه، دسته‌بندی و استخراج موضوعات	بررسی روند کاربرد کلیدواژگان/ ارتباطات کلیدواژگان
	دسته‌بندی کلیدواژه‌ها در حوزه‌های مفهومی فناوری اطلاعات با بهره‌گیری از فرکانس استفاده از کلیدواژگان در هر حوزه
	دسته‌بندی کلیدواژه‌ها در حوزه‌های مفهومی فناوری اطلاعات با بهره‌گیری از مدلسازی موضوعی

مدل پژوهش جاری در شکل زیر نمایش داده شده است:

شکل ٦- مدل پژوهش جاری

لازم به ذکر است در این پژوهش از دسته‌بندی‌های فناوری اطلاعات مختلف استفاده شده است. در این بخش به مطالعه‌ی دسته‌بندی‌های مختلف وبسایت‌های علم‌سنجی و پایگاه‌های دانش از فناوری اطلاعات می‌پردازیم. استفاده از این دسته‌بندی‌ها در ادامه این طرح پژوهشی مورد استفاده قرار خواهد گرفت. با توجه به پیشنهاده طرح پژوهشی جاری و بررسی روند تولید و انتشار علم در پایگاه WOS، یکی از دسته‌بندی‌های مورد نظر در این طرح، دسته‌بندی این پایگاه است. این دسته‌بندی باعث می‌شود تا در ادامه بتوانیم روندهای توسعه و تمرکز مطالعات را که از الگوریتم‌های متن‌کاوی به دست خواهند آمد دسته‌بندی کنیم. با توجه به اینکه هر یک از مطالعات فارغ از کلیدواژه و عنوان آن

دارای برچسبی برای مشخص شدن این دسته‌بندی هستند، می‌توان در ادامه طرح، روند میزان تمرکز مطالعات بر هر دسته را نیز مشخص نمود.

این دسته‌بندی با اختصار WC در مشخصات هر رکورد ثبت می‌شود. در شکل زیر نمونه مشخصات این دسته‌بندی برای چند رکورد آمده است:

DE	PU	PY	WC
Internet of Things; Ubiquitous sensing; Cloud computing; Wireless sensor networks; RFID; Smart environments	ELSEVIER SCIENCE BV	2013	Computer Science, Theory & Methods
Deep learning; Supervised learning; Unsupervised learning; Reinforcement learning; Evolutionary computation	PERGAMON-ELSEVIER SCIENCE LTD	2015	Computer Science, Artificial Intelligence; Neurosciences
Python; supervised learning; unsupervised learning; model selection	MICROTOME PUBL	2011	Automation & Control Systems; Computer Science, Artificial Intelligence
neural networks; regularization; model combination; deep learning	MICROTOME PUBL	2014	Automation & Control Systems; Computer Science, Artificial Intelligence
Dataset; Large-scale; Benchmark; Object recognition; Object detection	SPRINGER	2015	Computer Science, Artificial Intelligence
Algorithms; Performance; Experimentation; Classification LIBSVM optimization regression support vector machines SVM	ASSOC COMPUTING MACHINERY	2011	Computer Science, Artificial Intelligence; Computer Science, Information Systems
Physical layer security; Information-theoretic security; wiretap channel; secrecy; artificial noise; cooperative jamming; sec	IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC	2014	Computer Science, Information Systems; Telecommunications
Big Data; data mining; heterogeneity; autonomous sources; complex and evolving associations	IEEE COMPUTER SOC	2014	Computer Science, Artificial Intelligence; Computer Science, Information Systems; Engineering, Electrical & Electronic
Super-resolution; deep convolutional neural networks; sparse coding	IEEE COMPUTER SOC	2016	Computer Science, Artificial Intelligence; Engineering, Electrical & Electronic
Open Source; Computer Vision; Neural Networks; Parallel Computation; Machine Learning	ASSOC COMPUTING MACHINERY	2014	Computer Science, Theory & Methods; Engineering, Electrical & Electronic
Salient object detection; visual attention; saliency map; unsupervised segmentation; image retrieval	IEEE COMPUTER SOC	2015	Computer Science, Artificial Intelligence; Engineering, Electrical & Electronic
Cloud computing; Data privacy; Data protection; Security; Virtualization	ACADEMIC PRESS LTD- ELSEVIER SCIENCE LTD	2011	Computer Science, Hardware & Architecture; Computer Science, Interdisciplinary Applications; Computer Science, Software Engineering
Algorithms; Theory; Principal components; robustness vis-a-vis outliers; nuclear-norm minimization; l(1)-norm minimizatio	ASSOC COMPUTING MACHINERY	2011	Computer Science, Hardware & Architecture; Computer Science, Information Systems; Computer Science, Software Engineering; Compu

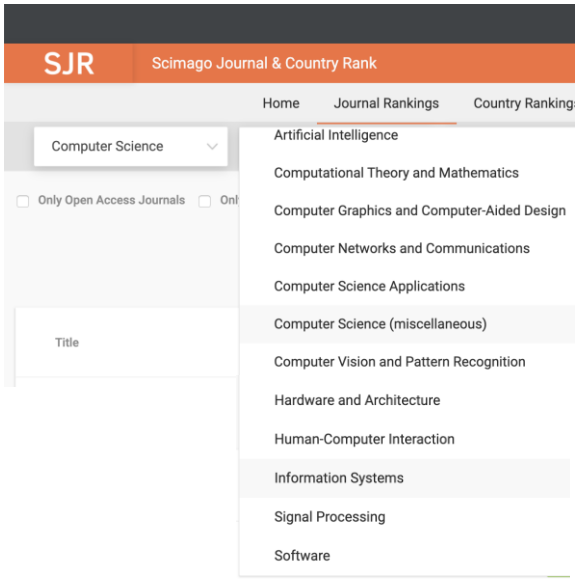
شکل ۷- نمایش دسته‌بندی WC پایگاه (WOS Catagories)

دسته‌بندی فناوری اطلاعات به‌صورت زیر در این پایگاه شکل گرفته است:

- Computer Science, Artificial Intelligence
- Computer Science, Cybernetics
- Computer Science, Hardware & Architecture
- Computer Science, Information Systems
- Computer Science, Interdisciplinary Applications
- Computer Science, Software Engineering
- Computer Science, Theory & Methods
- Information Science & Library Science
- Medical Informatics
- Telecommunications

باقی دسته‌بندی‌ها (علوم دیگر) در پیوست این طرح آمده است. این دسته‌بندی به‌منظور جستجوی مفیدتر و بررسی روند دقیق‌تر در فصل ۲-۷ مورد استفاده قرار می‌گیرد.

دسته‌بندی دیگری که در این پژوهش مور استفاده قرار می‌گیرد دسته‌بندی وبسایت scimagojr است. دلیل استفاده از این دسته‌بندی، جستجو در این پایگاه به منظور یافتن مجلاتی است که از رتبه بالایی از ضریب تاثیر یا سابقه یا دیگر عواملی که از نظر خبرگان به‌عنوان مجله پیشرو در زمینه فناوری اطلاعات برخوردارند. نمای دسته‌بندی در این پایگاه در شکل ۸ آمده است:



شکل ۸- نمایی از دسته‌بندی پایگاه

لازم به ذکر است این دسته‌بندی در بررسی روند فناوری اطلاعات در ادامه فصل، مورد استفاده قرار نمی‌گیرد و صرفا به‌منظور یافتن مجلات پیشرو است. دلیل این مساله استفاده از پایگاه WOS برای یافتن پژوهش‌ها در این طرح است و لازم به ذکر است که در ادامه طرح، دسته‌بندی برآمده از مدلسازی موضوعی که از نتایج طرح جاری است ارائه خواهد شد.

بر اساس دسته‌بندی انجام شده، تمام مطالعات فناوری اطلاعات در زمینه‌های زیر توسط پایگاه مورد بررسی تقسیم شده‌اند:

- هوش مصنوعی
- نظریه محاسباتی و ریاضیات در فناوری اطلاعات
- گرافیک و طراحی به کمک رایانه
- شبکه رایانه و ارتباطات
- کاربرد علوم کامپیوتر
- علوم کامپیوتر (متفرقه)
- بینایی کامپیوتر و تشخیص الگو
- سخت افزار و معماری
- تعامل انسان و رایانه
- سیستم‌های اطلاعاتی
- پردازش سیگنال
- نرم افزار

از دیگر دسته‌بندی‌های معتبر در پایگاه‌های علمی دنیا می‌توان به موارد زیر اشاره نمود که به‌دلیل عدم استفاده در این طرح پژوهشی به آنها پرداخته نمی‌شود:

- «ان.دی.ال.تی.دی.»[۱]
- «او.ای.تی.دی.»[۲]
- «سی.آر.ال»[۳]
- «سایبرتسیس»[۴]
- «اریک»[۵]
- «پی.کیو.دی.تی. اوپن» متعلق به «پروکوئست»[۶]
- «دارت یوروپ»[۷]
- «دیوا پورتال»[۸]
- «داتاد»[۹]

۲-٤. انتخاب مجلات معتبر

در فاز اول از برنامه پژوهش تشریح شده در بخش قبل، انتخاب مجلات معتبر برای استخراج اولیه سبد واژگان انجام می‌پذیرد. در طرحواره این پژوهش استفاده از کلیدواژگان مورد استفاده در مجلات برتر در حوزه فناوری اطلاعات به‌عنوان روشی برای تشکیل سبد اولیه در نظر گرفته شد. شاخص‌های متعددی برای انتخاب این مجلات پیش روی محققین قرار دارد. یکی از انواع شاخص‌ها، استفاده از شاخص‌های فعالیت مجلات مانند اچ‌ایندکس، سابقه، میزان ضریب تاثیر و ... است. نوع دیگر انتخاب این مجلات، استفاده از دانش خبرگی متخصصین حوزه فناوری اطلاعات است. ما در این پژوهش از هر دو روش بهره‌گیری نموده‌ایم.

مجلات معتبر در هر حوزه ذکر شده در بخش قبل (۲-۴) در جدول زیر آمده است:

---

1 - http://search.ndltd.org/
2 - https://oatd.org/
3 - http://catalog.crl.edu/
4 - http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/key-organizations/latin-america-and-the-caribbean/cybertesis
5 - https://eric.ed.gov/
6 - pqdtopen.proquest.com
7 - http://www.dart-europe.eu/basic-search.php
8 - http://diva-portal.org/
9 - http://datad.aau.org/

جدول ۲- فهرست مجلات دارای ضریب تاثیر بالا در پایگاه WOS

تکرار	Total Cites (3years)	Total Docs. (3years)	H index	عنوان	رتبه در H	زمینه
4	11286	577	326	IEEE Transactions on Pattern Analysis and Machine Intelligence	1	AI
4	7777	1061	180	Pattern Recognition	2	
2	3440	726	173	Journal of Machine Learning Research	3	
2	4345	278	172	International Journal of Computer Vision	4	
2	4245	457	170	IEEE Transactions on Fuzzy Systems	5	
2	14186	2405	335	Bioinformatics	1	Computional theory and mathematics
4	11286	577	326	IEEE Transactions on Pattern Analysis and Machine Intelligence	2	
2	4245	457	170	IEEE Transactions on Fuzzy Systems	3	
1	2247	183	154	IEEE Transactions on Evolutionary Computation	4	
2	4124	691	148	IEEE Transactions on Knowledge and Data Engineering	5	
2	13081	1371	242	IEEE Transactions on Image Processing	1	Computer graphics and computer aided design
1	5251	734	173	ACM Transactions on Graphics	2	
1	2874	623	118	IEEE Transactions on Visualization and Computer Graphics	3	
1	3208	368	113	Medical Image Analysis	4	
1	1495	433	110	Journal of Mechanical Design - Transactions of the ASME	5	
2	12402	1106	213	IEEE Communications Magazine	1	Computer network and communication
1	8083	741	211	IEEE Journal on Selected Areas in Communications	2	
1	3446	707	157	IEEE/ACM Transactions on Networking	3	
1	905	192	152	Computer Communication Review	4	
1	14929	2388	146	IEEE Transactions on Vehicular Technology	5	
2	14186	2405	335	Bioinformatics	1	Computer science application
1	9981	1495	260	IEEE Transactions on Automatic Control	2	
2	5123	1391	249	IEEE Transactions on Information Theory	3	
1	25178	2627	236	IEEE Transactions on Industrial Electronics	4	
2	12402	1106	213	IEEE Communications Magazine	5	
1	10379	2088	226	European Journal of Operational Research	1	Computer science (miscellaneous)
1	2212	946	189	Communications of the ACM	2	
1	4208	548	149	Computers and Education	3	
1	3119	685	133	Computers and Operations Research	4	
1	3048	275	132	ACM Computing Surveys	5	
4	11286	577	326	IEEE Transactions on Pattern Analysis and Machine Intelligence	1	Computer vision and pattern
4	7777	1061	180	Pattern Recognition	2	

تکرار	Total Cites (3years)	Total Docs. (3years)	H index	عنوان	رتبه در H	زمینه
2	4345	278	172	International Journal of Computer Vision	3	recognition
1	1767	892	144	Journal of the Optical Society of America A: Optics and Image Science, and Vision	4	
2	3110	836	139	Pattern Recognition Letters	5	
1	3522	874	154	Computer Physics Communications	1	Hardware and architecture
1	4268	824	120	IEEE Transactions on Parallel and Distributed Systems	2	
1	561	155	117	Journal of the ACM	3	
1	2342	262	111	IEEE Network	4	
1	3496	772	110	IEEE Transactions on Computers	5	
1	13748	2250	137	Computers in Human Behavior	1	Human computer interaction
1	1245	401	119	Cyberpsychology, Behavior, and Social Networking	2	
1	4643	735	111	IEEE Transactions on Systems, Man, and Cybernetics: Systems	3	
1	1058	249	109	International Journal of Human Computer Studies	4	
1	1206	267	104	IEEE Transactions on Human-Machine Systems	5	
2	5123	1391	249	IEEE Transactions on Information Theory	1	Information systems
1	1750	148	195	MIS Quarterly: Management Information Systems	2	
2	4124	691	148	IEEE Transactions on Knowledge and Data Engineering	3	
1	955	151	145	Information Systems Research	4	
1	1650	239	142	Information and Management	5	
1	9099	1459	233	IEEE Transactions on Signal Processing	1	Signal processing
4	7777	1061	180	Pattern Recognition	2	
1	1969	282	155	IEEE Signal Processing Magazine	3	
2	3110	836	139	Pattern Recognition Letters	4	
1	7738	1271	134	Mechanical Systems and Signal Processing	5	
4	11286	577	326	IEEE Transactions on Pattern Analysis and Machine Intelligence	1	Software
2	13081	1371	242	IEEE Transactions on Image Processing	2	
1	5845	686	195	IEEE Transactions on Medical Imaging	3	
4	7777	1061	180	Pattern Recognition	4	
2	3440	726	173	Journal of Machine Learning Research	5	

واژه فرکانس نشان‌دهنده تعداد حضور هر یک از مجلات در بین ۵ مجله برتر (همین جدول) در هر حوزه است.

۲-٤-۱. جستجوی مجلات برتر در حوزه فناوری اطلاعات

بر اساس جدول ۲، مجلات پر کاربرد در هر دسته انتخاب بر حسب شاخص اچ مرتب‌سازی شده و توسط محققین انتخاب شدند. این انتخاب به نحوی صورت گرفت که مجلاتی که در چند حوزه در فهرست ۵ مجله برتر قرار گرفته‌اند انتخاب گردند. همچنین همه دسته‌ها حداقل یک مجله به‌عنوان نماینده برای جستجو داشته باشند. در نتیجه در مرحله اول یافتن مجلاتی که در ۵ مجله تاثیرگذار هر زمینه گنجیده و همچنین در زمینه‌های مختلف فعالیت دارند در اولویت انتخاب قرار گرفت. در جدول زیر فهرست مجلات ۹گانه انتخاب شده و حوزه فعالیت آنها مشخص شد. لازم به ذکر است دیگر مشخصات این مجلات در جدول بخش قبل آمده است.

جدول ۳- اطلاعات و حوزه پوشش مجلات منتخب

نام مجله	حوزه‌ی پوشش دهنده	ناشر	کشور	اولین انتشار
Bioinformatics	Computional theory and Mathematics	Elsevier Ltd.	United Kingdom	1985
	Computer Science Application			
Computer Physics Communications	Hardware and Architecture	Elsevier BV	Netherlands	1969
Computers in Human Behavior	Human Computer Interaction	Elsevier Ltd.	United Kingdom	1985
European Journal of Operational Research	Computer Science (miscellaneous)	Elsevier BV	Netherlands	1977
IEEE Communications Magazine	Computer Network and Communication	Institute of Electrical and Electronics Engineers	United States	1979
	Computer Science Application			
IEEE Transactions on Image Processing	Computer Graphics and Computer Aided Design	Institute of Electrical and Electronics Engineers	United States	1992
	Software			
IEEE Transactions on Information Theory	Computer Science Application	Institute of Electrical and Electronics Engineers	United States	1969
	Information Systems			
IEEE Transactions on Pattern Analysis and Machine Intelligence	AI	Institute of Electrical and Electronics	United States	1978
	Computional Theory and Mathematics			

نام مجله	حوزه‌ی پوشش دهنده	ناشر	کشور	اولین انتشار
	Computer Vision and Pattern Recognition	Engineers		
	Software			
Pattern Recognition	AI	Elsevier Ltd.	United Kingdom	1968
	Signal Processing			

با توجه به انتخاب مجموعه فوق و با کد زیر، تعداد ۸۴۵۱ رکورد شامل مشخصات، کلیدواژه، عنوان و چکیده در ۳ سال اخیر دریافت شدند.

SO=(BIOINFORMATICS) OR SO=(Computer Physics Communications) OR SO=(Computers in Human Behavior) OR SO=(European Journal of Operational Research) OR SO=(IEEE Communications Magazine) OR SO=(IEEE Transactions on Image Processing) OR SO=(IEEE Transactions on Information Theory) OR SO=(IEEE Transactions on Pattern Analysis and Machine Intelligence) OR SO=(Pattern Recognition)

بر این اساس ۱۷۴۰۳ کلیدواژه منحصر به فرد از مجموع ۲۶۲۶۸ واژه شناسایی شده و واژگان بالای ۱۰ تکرار در جدول زیر آمده است:

جدول ٤- کلمات پرکاربرد در مجلات دارای ضریب تاثیر بالا

ردیف	واژه	تکرار
1	Deep Learning	171
2	Convolutional Neural Network (CNN)	145
3	Supply Chain Management	70
4	Social Media	66
5	Data Envelopment Analysis	62
15	Game Theory	62
8	Combinatorial Optimization	42
7	Integer Programming	42
9	Optimization	41
10	Feature Extraction	40
11	Health Services	39
12	Machine Learning	38
13	Scheduling	38
14	Image Segmentation	37

ردیف	واژه	تکرار
16	Routing	35
17	Classification	34
18	Sparse Representation	34
19	Transportation	34
20	Task Analysis	33
21	Inventory	31
22	Logistics	31
23	Person Re-Identification	31
24	Energy	29
25	Visual Tracking	29
26	Finance	28
29	Clustering	26
30	Heuristics	26
31	Social Media	26
32	Dynamic Programming	25
33	Object Detection	25
34	Facebook	24
35	Dictionary Learning	22
36	Feature Selection	22
37	Semi-Supervised Learning	22
38	Simulation	22
39	Stochastic Programming	22
40	Training	22
44	Decision Analysis	21
41	Metric Learning	21
42	Neural Networks	21
43	Visualization	21
45	Image Classification	20
46	Image Retrieval	20
47	Robust Optimization	20
48	Super-Resolution	20

۲-٤-۲. انتخاب مجلات با استفاده از نظر تکمیلی خبرگان

با توجه به روش تحقیق مشروح در بخش سوم این فصل، از ۳ خبره در زمینه فناوری اطلاعات برای معرفی مجلات معتبر و مورد نظر ایشان نیز بهره گرفته شد. براین اساس، علاوه بر مجلات انتخاب شده در بخش قبل، مجلات زیر نیز توسط خبرگان معرفی شدند.

- IEEE Intelligent Systems
- IEEE Transactions on Information Theory
- IEEE Transactions on Information Forensics And Security
- ACM Transactions on Computer Systems
- IBM Journal of Research And Development
- Journal of Visual Languages And Computing
- Computational Linguistics
- Information Systems
- IEEE Transactions on Neural Networks And Learning Systems
- International Journal of Human-Computer Studies
- IEEE Transactions on Knowledge And Data Engineering
- Applied Intelligence
- AI Magazine

جستجوی مجلات فوق در بازه ۵ ساله با کد زیر در پایگاه WOS انجام پذیرفت:

so=(ieee intelligent systems) or so=(ieee transactions on information theory) or so=(ieee transactions on information forensics and security) or so=(acm transactions on computer systems) or so=(ibm journal of research and development) or so=(journal of visual languages and computing) or so=(computational linguistics) or so=(information systems) or so=(ieee transactions on neural networks and learning systems) or so=(international journal of human-computer studies) or so=(ieee transactions on knowledge and data engineering) or so=(applied intelligence) or so=(ai magazine)

با استفاده از مجلات معرفی شده توسط خبرگان در طول ۵ سال اخیر ۷۲۴۱ رکورد شامل مشخصات، کلیدواژه، عنوان و چکیده دریافت شدند.

فهرست واژگان بالای ۲۰ تکرار جستجوی فوق در جدول زیر آمده است:

جدول ۵- تکرار کلیدواژگان پرکاربرد در مجلات مورد تایید خبرگان

ردیف	واژه	تکرار
۱	Neural Network(s) (NN(s))	۱۶۵
۲	Machine Learning	۹۸
۳	Classification	۸۴
۴	Feature Selection	۸۰
۵	Deep Learning	۷۹
۶	Clustering	۷۲
۷	Data Mining	۶۸

ردیف	واژه	تکرار
۸	Compressed Sensing	۶۲
۹	Security	۶۰
۱۰	Optimization	۵۹
۱۱	Social Network (s)	۵۹
۱۲	Capacity	۵۸
۱۳	Physical Layer Security	۵۴
۱۴	Privacy	۵۲
۱۵	Network Coding	۴۷
۱۶	Reinforcement Learning (Rl)	۴۷
۱۷	Cloud Computing	۴۵
۱۸	Channel Capacity	۴۲
۱۹	Feature Extraction	۳۶
۲۰	Online Learning	۳۶
۲۱	Synchronization	۳۴
۲۲	Adaptive Control	۳۲
۲۳	Game Theory	۳۲
۲۴	Mutual Information	۳۲
۲۵	Supervised Learning	۳۲
۲۶	Robustness	۳۱
۲۷	Authentication	۲۹
۲۸	Optimal Control	۲۹
۲۹	Stability	۲۹
۳۰	Big Data	۲۸
۳۱	Information Theory	۲۷
۳۲	Unsupervised Learning	۲۶
۳۳	Quantization	۲۴
۳۴	Recommender Systems	۲۴
۳۵	Source Coding	۲۴
۳۶	Crowdsourcing	۲۳
۳۷	Dynamic Programming	۲۳
۳۸	Error Exponent	۲۳
۳۹	Face Recognition	۲۳
۴۰	Side Information	۲۳
۴۱	Feedback	۲۲

ردیف	واژه	تکرار
۴۲	Hypothesis Testing	۲۲
۴۳	Interference Channel	۲۲
۴۴	Query Processing	۲۲
۴۵	Wiretap Channel	۲۲
۴۶	Interference Alignment	۲۱
۴۷	Mimo	۲۱
۴۸	Nonlinear Systems	۲۱
۴۹	Particle Swarm Optimization	۲۱
۵۰	Regularization	۲۱
۵۱	Biometrics	۲۰
۵۲	Convolutional Neural Network	۲۰
۵۳	Entropy	۲۰
۵۴	Image Classification	۲۰
۵۵	Polar Codes	۲۰

کشف این تکرار با توجه به کد نرم‌افزاری توسعه یافته در این طرح پژوهشی در نرم‌افزار R به دست آمده است و گام‌های زیر را انجام می‌دهد:

- مرتب‌سازی کلیدواژگان از نظر " " اضافی که ممکن است ایجاد شده باشد
- کوچک‌سازی همه حروف به‌منظور کشف یکسانی واژگان
- جداسازی کلیدواژه‌های یک مقاله که توسط ";" در فایل خروجی پایگاه از هم تفکیک شده‌اند
- ایجاد سطر متفاوت برای هر کلیدواژه (ساخت ماتریس ۱ در n)
- شمارش تکرار هر کلیدواژه

۲-۵. انتخاب کلیدواژگان منتخب

با توجه به فهرست بالای واژگان پرتکرار در دو جستجوی مجلات پیشرو و مجلات منتخب خبرگان، یافتن سبد واژگان منتخب فناوری اطلاعات با قوانینی (به‌عنوان توصیه به خبرگان) در این گام مورد توجه قرار گرفت. قوانین زیر برای اجماع خبرگان درنظر گرفته شد:

- واژه مختص به فناوری اطلاعات باشد
- واژه شامل ابزار حل مساله نباشد
- واژه‌های علمی عمومی مانند مدل و ... در سبد واژگان قرار نگیرد
- تعداد واژه پر تکرار باشد (در این پژوهش ۱۰ تکرار)
- واژه شامل کاربرد فناوری اطلاعات در زمینه دیگر علمی نباشد

به منظور در اختیار گذاشتن تمامی واژگان یافت شده در دو جستجوی اخیر، پاکسازی واژگان نیز در دستور کار قرار گرفت. استفاده از کلیدواژگان در مجلات مختلف ممکن است یکسان نباشد. بطور مثال واژه IOT و یا Internet of Things ممکن است در کلیدواژگان مقالات استفاده شوند درصورتی که به یک مفهوم اشاره دارند. در جدول زیر کلیدواژگان پرتکرار در اختیار خبرگان قرار گرفته و در لیست زیر کلیدواژگان منتخب آنها انتخاب شدند.

جدول ٦- انتخاب کلیدواژه از دیدگاه خبرگان برای تشکیل سبد واژگان

ردیف	واژه	تایید خبره
۱	Deep Learning	*
۲	Machine Learning	*
۳	Supply Chain Management	
۴	Convolutional Neural Network	
۵	Social Media	*
۶	Data Envelopment Analysis	*
۷	Security	*
۸	Optimization	
۹	Feature Selection	*
۱۰	Privacy	*
۱۱	Classification	
۱۲	Network Coding	*
۱۳	Clustering	
۱۴	Channel Capacity	

ردیف	واژه	تایید خبره
۱۵۱	Image Denoising	
۱۵۲	Information Theoretic Security	*
۱۵۳	Linear Programming	
۱۵۴	Mapreduce	
۱۵۵	Multiple Access Channel	
۱۵۶	Network Security	*
۱۵۷	Physical-Layer Security	
۱۵۸	Private Information Retrieval	*
۱۵۹	Representation Learning	
۱۶۰	Semidefinite Programming	
۱۶۱	Trust	
۱۶۲	Usability	
۱۶۳	Wireless Networks	*
۱۶۴	5g Mobile Communication	*

ردیف	واژه	تایید خبره
۱۵	Combinatorial Optimization	
۱۶	Integer Programming	
۱۷	Compressed Sensing	
۱۸	Feature Extraction	*
۱۹	Health Services	
۲۰	Data Mining	*
۲۱	Neural Network	
۲۲	Scheduling	
۲۳	Image Segmentation	*
۲۴	Online Learning	*
۲۵	Capacity	
۲۶	Game Theory	
۲۷	Routing	
۲۸	Sparse Representation	
۲۹	Synchronization	
۳۰	Transportation	
۳۱	Task Analysis	
۳۲	Adaptive Control	*
۳۳	Mutual Information	
۳۴	Physical Layer Security	*
۳۵	Supervised Learning	*
۳۶	Inventory	
۳۷	Logistics	
۳۸	Person Re-Identification	*
۳۹	Robustness	
۴۰	Authentication	*
۴۱	Energy	
۴۲	Optimal Control	
۴۳	Stability	
۴۴	Visual Tracking	
۴۵	Big Data	*
۴۶	Finance	

ردیف	واژه	تایید خبره
۱۶۵	Adaptive Dynamic Programming (Adp)	
۱۶۶	Capacity Region	
۱۶۷	Channel State Information	
۱۶۸	Computational Complexity	
۱۶۹	Discrete Optimization	
۱۷۰	Diversity	
۱۷۱	Domain Adaptation	
۱۷۲	Hashing	*
۱۷۳	Image Forensics	
۱۷۴	Index Coding	*
۱۷۵	Information Entropy	*
۱۷۶	Maintenance	
۱۷۷	Manifold Learning	
۱۷۸	Maritime Industry	
۱۷۹	Monte Carlo	
۱۸۰	Parallel Computing	*
۱۸۱	Random Forest	
۱۸۲	Self-Esteem	
۱۸۳	Sparse Coding	*
۱۸۴	Www	*
۱۸۵	Attention	
۱۸۶	Benders Decomposition	
۱۸۷	Branch-And-Price	
۱۸۸	Cloud Storage	*
۱۸۹	Data Privacy	*
۱۹۰	Deep Neural Network	
۱۹۱	Differential Evolution	
۱۹۲	Dimension Reduction	
۱۹۳	Estimation	
۱۹۴	Evolutionary Algorithm	
۱۹۵	Forecasting	
۱۹۶	Gaussian Mixture Model	

ردیف	واژه	تایید خبره
۴۷	Information Theory	*
۴۸	Heuristics	
۴۹	Dynamic Programming	
۵۰	Object Detection	*
۵۱	Cloud Computing	*
۵۲	Facebook	
۵۳	Quantization	
۵۴	Recommender Systems	*
۵۵	Social Network	*
۵۶	Source Coding	*
۵۷	Crowdsourcing	*
۵۸	Error Exponent	*
۵۹	Face Recognition	*
۶۰	Side Information	*
۶۱	Dictionary Learning	
۶۲	Feedback	
۶۳	Hypothesis Testing	
۶۴	Interference Channel	*
۶۵	Query Processing	*
۶۶	Reinforcement Learning	
۶۷	Simulation	
۶۸	Stochastic Programming	
۶۹	Training	
۷۰	Wiretap Channel	
۷۱	Cnn	
۷۲	Decision Analysis	*
۷۳	Interference Alignment	
۷۴	Metric Learning	
۷۵	Mimo	
۷۶	Nonlinear Systems	
۷۷	Particle Swarm Optimization	
۷۸	Regularization	
۷۹	Visualization	*

ردیف	واژه	تایید خبره
۱۹۷	Generative Adversarial Networks	
۱۹۸	Information Security	*
۱۹۹	Iterative Decoding	*
۲۰۰	Learning Analytics	
۲۰۱	Mds Codes	*
۲۰۲	Motivation	
۲۰۳	Nonconvex Optimization	
۲۰۴	Org	
۲۰۵	Process Mining	*
۲۰۶	Project Scheduling	
۲۰۷	Reed-Solomon Codes	
۲۰۸	Resource Allocation	
۲۰۹	Second-Order Asymptotics	
۲۱۰	Secrecy	
۲۱۱	Sentiment Analysis	*
۲۱۲	Spectral Clustering	*
۲۱۳	Superposition Coding	*
۲۱۴	Swarm Intelligence	
۲۱۵	System Identification	*
۲۱۶	Visual Analytics	
۲۱۷	Wireless Sensor Networks	*
۲۱۸	Zero-Error Capacity	
۲۱۹	Artificial Neural Networks	
۲۲۰	Boolean Functions	
۲۲۱	Change Detection	
۲۲۲	Computer Security	
۲۲۳	Concept Drift	
۲۲۴	Cooperation	
۲۲۵	Correlation	
۲۲۶	Cuda	
۲۲۷	Data Models	*
۲۲۸	Energy Harvesting	
۲۲۹	Environment And Climate Change	

ردیف	واژه	تایید خبره
۸۰	Biometrics	*
۸۱	Entropy	
۸۲	Image Classification	*
۸۳	Image Retrieval	*
۸۴	Polar Codes	
۸۵	Robust Optimization	
۸۶	Super-Resolution	
۸۷	Broadcast Channel	*
۸۸	Channel Coding	*
۸۹	Dimensionality Reduction	
۹۰	Distributed Storage	
۹۱	Kernel Methods	*
۹۲	Matrix Factorization	
۹۳	Pattern Recognition	*
۹۴	Regenerating Codes	
۹۵	Segmentation	
۹۶	Anomaly Detection	
۹۷	Graph Theory	
۹۸	Image Reconstruction	*
۹۹	Multiple Criteria Analysis	
۱۰۰	Secrecy Capacity	
۱۰۱	Sparsity	
۱۰۲	Support Vector Machine	
۱۰۳	User Experience	*
۱۰۴	Artificial Intelligence	*
۱۰۵	Collaborative Filtering	
۱۰۶	Computational Modeling	*
۱۰۷	Convex Optimization	
۱۰۸	Degrees Of Freedom	
۱۰۹	Genetic Algorithm	
۱۱۰	Gpu	
۱۱۱	Image Restoration	*
۱۱۲	Learning Systems	*

ردیف	واژه	تایید خبره
۲۳۰	Error Correction Codes	*
۲۳۱	Group Testing	
۲۳۲	Homomorphic Encryption	*
۲۳۳	Human-Computer Interaction	*
۲۳۴	Iris Recognition	
۲۳۵	Large Deviations	
۲۳۶	Location Privacy	
۲۳۷	Molecular Dynamics	
۲۳۸	Multi-Objective Optimization	
۲۳۹	Ontology	
۲۴۰	Pose Estimation	
۲۴۱	Production	
۲۴۲	Rate-Distortion Theory	
۲۴۳	Remote Sensing	
۲۴۴	Renyi Entropy	
۲۴۵	Revenue Management	
۲۴۶	Risk Management	
۲۴۷	Salient Object Detection	
۲۴۸	Shannon Theory	
۲۴۹	Smart Grid	
۲۵۰	Subspace Learning	
۲۵۱	Video Games	
۲۵۲	Virtual Reality	
۲۵۳	Wireless Communication	
۲۵۴	Admm	
۲۵۵	Children	
۲۵۶	Competition	
۲۵۷	Compression	
۲۵۸	Convolutional Neural Network (Cnn)	
۲۵۹	Data Streams	
۲۶۰	Decision Processes	
۲۶۱	Decision Support Systems	
۲۶۲	Dispersion	

ردیف	واژه	تایید خبره
۱۱۳	Multiple Objective Programming	
۱۱۴	Multi-Task Learning	*
۱۱۵	Pricing	
۱۱۶	Quantum Entanglement	
۱۱۷	Recurrent Neural Network	
۱۱۸	Regression	
۱۱۹	Saliency Detection	*
۱۲۰	Time Delay	
۱۲۱	Action Recognition	*
۱۲۲	Adolescents	
۱۲۳	Community Operational Research	
۱۲۴	Computer Vision	*
۱۲۵	Convergence	
۱۲۶	Ensemble Learning	*
۱۲۷	Internet Of Things	*
۱۲۸	List Decoding	*
۱۲۹	Location	
۱۳۰	Renyi Divergence	
۱۳۱	Scalability	
۱۳۲	State Estimation	
۱۳۳	Transfer Learning	
۱۳۴	Analytics	
۱۳۵	Efficiency	
۱۳۶	Global Optimization	
۱۳۷	Joint Source-Channel Coding	
۱۳۸	Kernel	
۱۳۹	Linear Codes	
۱۴۰	Marketing	
۱۴۱	Reliability	
۱۴۲	Sampling	
۱۴۳	Semantics	*
۱۴۴	Semisupervised Learning	

ردیف	واژه	تایید خبره
۲۶۳	Extreme Learning Machine (Elm)	
۲۶۴	Gaussian Processes	
۲۶۵	High-Dimensional Data	
۲۶۶	Image Coding	
۲۶۷	Image Fusion	
۲۶۸	Image Recognition	*
۲۶۹	Internet Addiction	
۲۷۰	Intrusion Detection	
۲۷۱	Link Prediction	*
۲۷۲	Locally Repairable Codes	
۲۷۳	Loneliness	
۲۷۴	Low-Density Parity-Check (Ldpc) Codes	
۲۷۵	Metaheuristics	
۲۷۶	Mixed Integer Programming	
۲۷۷	Multi-View Learning	*
۲۷۸	Natural Language Processing	
۲۷۹	Natural Resources	
۲۸۰	Networks	
۲۸۱	Parameter Estimation	
۲۸۲	Personalization	
۲۸۳	Quantum Mechanics	
۲۸۴	Random Coding	*
۲۸۵	Redundancy	
۲۸۶	Secrecy Outage Probability	
۲۸۷	Social Support	
۲۸۸	Spiking Neural Network (Snn)	
۲۸۹	Stability Analysis	
۲۹۰	Stochastic Geometry	
۲۹۱	Strong Converse	
۲۹۲	Subspace Clustering	
۲۹۳	Subspace Codes	
۲۹۴	Sustainability	

ردیف	واژه	تایید خبره
۱۴۵	Uncertainty	
۱۴۶	Column Generation	
۱۴۷	Community Detection	*
۱۴۸	Cryptography	*
۱۴۹	Cyberbullying	*
۱۵۰	Differential Privacy	

ردیف	واژه	تایید خبره
۲۹۵	Text Mining	*
۲۹۶	Time-Varying Delay	
۲۹۷	Topic Modeling	*
۲۹۸	Twitter	
۲۹۹	Upper Bounds	
۳۰۰	Verification	

از مجموع ۳۰۰ واژه با تکرار مناسب، 88 واژه توسط خبرگان برای قرارگیری در سبد انتخاب شدند.

۲-۶. بررسی روند توسعه کلیدواژه‌ها

کلیدواژه‌های یافت شده در دسته‌بندی فناوری اطلاعات در پایگاه WOS در بازه زمانی ۱۰ ساله مورد جستجو قرار می‌گیرند. همانگونه که در فصل ۲-۴ ذکر شد، دسته‌بندی فناوری اطلاعات به‌صورت زیر در این پایگاه شکل گرفته است:

- Computer Science, Artificial Intelligence
- Computer Science, Cybernetics
- Computer Science, Hardware & Architecture
- Computer Science, Information Systems
- Computer Science, Interdisciplinary Applications
- Computer Science, Software Engineering
- Computer Science, Theory & Methods
- Information Science & Library Science
- Medical Informatics
- Telecommunications

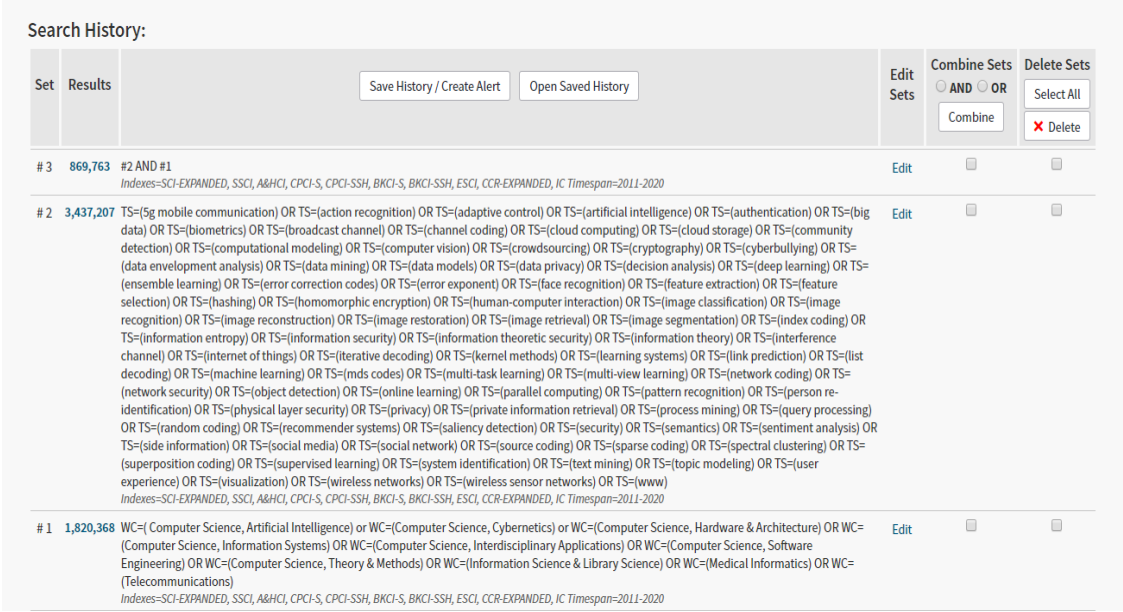
کد جستجوی زیر برای محدود کردن جستجو در زمینه مرتبط با فناوری در ده سال اخیر مورد استفاده قرار گرفته است:

WC=(Computer Science, Artificial Intelligence) or WC=(Computer Science, Cybernetics) or WC=(Computer Science, Hardware & Architecture) OR WC=(Computer Science, Information Systems) OR WC=(Computer Science, Interdisciplinary Applications) OR WC=(Computer Science, Software Engineering) OR WC=(Computer Science, Theory & Methods) OR WC=(Information Science & Library Science) OR WC=(Medical Informatics) OR WC=(Telecommunications)

همچنین کد زیر برای جستجوی سبد واژگان در مجموعه پایگاه و در ۱۰ سال اخیر مورد استفاده قرار گرفته است:

TS=(5g mobile communication) OR TS=(action recognition) OR TS=(adaptive control) OR TS=(artificial intelligence) OR TS=(authentication) OR TS=(big data) OR TS=(biometrics) OR TS=(broadcast channel) OR TS=(channel coding) OR TS=(cloud computing) OR TS=(cloud storage) OR TS=(community detection) OR TS=(computational modeling) OR TS=(computer vision) OR TS=(crowdsourcing) OR TS=(cryptography) OR TS=(cyberbullying) OR TS=(data envelopment analysis) OR TS=(data mining) OR TS=(data models) OR TS=(data privacy) OR TS=(decision analysis) OR TS=(deep learning) OR TS=(ensemble learning) OR TS=(error correction codes) OR TS=(error exponent) OR TS=(face recognition) OR TS=(feature extraction) OR TS=(feature selection) OR TS=(hashing) OR TS=(homomorphic encryption) OR TS=(human-computer interaction) OR TS=(image classification) OR TS=(image recognition) OR TS=(image reconstruction) OR TS=(image restoration) OR TS=(image retrieval) OR TS=(image segmentation) OR TS=(index coding) OR TS=(information entropy) OR TS=(information security) OR TS=(information theoretic security) OR TS=(information theory) OR TS=(interference channel) OR TS=(internet of things) OR TS=(iterative decoding) OR TS=(kernel methods) OR TS=(learning systems) OR TS=(link prediction) OR TS=(list decoding) OR TS=(machine learning) OR TS=(mds codes) OR TS=(multi-task learning) OR TS=(multi-view learning) OR TS=(network coding) OR TS=(network security) OR TS=(object detection) OR TS=(online learning) OR TS=(parallel computing) OR TS=(pattern recognition) OR TS=(person re-identification) OR TS=(physical layer security) OR TS=(privacy) OR TS=(private information retrieval) OR TS=(process mining) OR TS=(query processing) OR TS=(random coding) OR TS=(recommender systems) OR TS=(saliency detection) OR TS=(security) OR TS=(semantics) OR TS=(sentiment analysis) OR TS=(side information) OR TS=(social media) OR TS=(social network) OR TS=(source coding) OR TS=(sparse coding) OR TS=(spectral clustering) OR TS=(superposition coding) OR TS=(supervised learning) OR TS=(system identification) OR TS=(text mining) OR TS=(topic modeling) OR TS=(user experience) OR TS=(visualization) OR TS=(wireless networks) OR TS=(wireless sensor networks) OR TS=(www)

اشتراک دو جستجوی فوق در پایگاه، مجموعه سبدواژگان منتخب در حوزه فناوری اطلاعات را در ده سال اخیر جستجو و ۸۶۹۷۶۳ رکورد بازیابی شدند.



Search History:

Set	Results	Save History / Create Alert / Open Saved History	Edit Sets	Combine Sets (AND / OR) Combine	Delete Sets Select All / Delete
# 3	869,763	#2 AND #1 *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=2011-2020*	Edit		
# 2	3,437,207	TS=(5g mobile communication) OR TS=(action recognition) OR TS=(adaptive control) OR TS=(artificial intelligence) OR TS=(authentication) OR TS=(big data) OR TS=(biometrics) OR TS=(broadcast channel) OR TS=(channel coding) OR TS=(cloud computing) OR TS=(cloud storage) OR TS=(community detection) OR TS=(computational modeling) OR TS=(computer vision) OR TS=(crowdsourcing) OR TS=(cryptography) OR TS=(cyberbullying) OR TS=(data envelopment analysis) OR TS=(data mining) OR TS=(data models) OR TS=(data privacy) OR TS=(decision analysis) OR TS=(deep learning) OR TS=(ensemble learning) OR TS=(error correction codes) OR TS=(error exponent) OR TS=(face recognition) OR TS=(feature extraction) OR TS=(feature selection) OR TS=(hashing) OR TS=(homomorphic encryption) OR TS=(human-computer interaction) OR TS=(image classification) OR TS=(image recognition) OR TS=(image reconstruction) OR TS=(image restoration) OR TS=(image retrieval) OR TS=(image segmentation) OR TS=(index coding) OR TS=(information entropy) OR TS=(information security) OR TS=(information theoretic security) OR TS=(information theory) OR TS=(interference channel) OR TS=(internet of things) OR TS=(iterative decoding) OR TS=(kernel methods) OR TS=(learning systems) OR TS=(link prediction) OR TS=(list decoding) OR TS=(machine learning) OR TS=(mds codes) OR TS=(multi-task learning) OR TS=(multi-view learning) OR TS=(network coding) OR TS=(network security) OR TS=(object detection) OR TS=(online learning) OR TS=(parallel computing) OR TS=(pattern recognition) OR TS=(person re-identification) OR TS=(physical layer security) OR TS=(privacy) OR TS=(private information retrieval) OR TS=(process mining) OR TS=(query processing) OR TS=(random coding) OR TS=(recommender systems) OR TS=(saliency detection) OR TS=(security) OR TS=(semantics) OR TS=(sentiment analysis) OR TS=(side information) OR TS=(social media) OR TS=(social network) OR TS=(source coding) OR TS=(sparse coding) OR TS=(spectral clustering) OR TS=(superposition coding) OR TS=(supervised learning) OR TS=(system identification) OR TS=(text mining) OR TS=(topic modeling) OR TS=(user experience) OR TS=(visualization) OR TS=(wireless networks) OR TS=(wireless sensor networks) OR TS=(www) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=2011-2020*	Edit		
# 1	1,820,368	WC=(Computer Science, Artificial Intelligence) or WC=(Computer Science, Cybernetics) or WC=(Computer Science, Hardware & Architecture) OR WC=(Computer Science, Information Systems) OR WC=(Computer Science, Interdisciplinary Applications) OR WC=(Computer Science, Software Engineering) OR WC=(Computer Science, Theory & Methods) OR WC=(Information Science & Library Science) OR WC=(Medical Informatics) OR WC=(Telecommunications) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=2011-2020*	Edit		

شکل ۹- نمای تلفیق دو جستجو در پایگاه

با توجه به اینکه برخی واژگان مانند امنیت می‌توانند در حوزه‌های دیگر علمی نیز یافت شوند، شمار واژگان یافت شده در زمینه علمی مرتبط با فناوری اطلاعات نزدیک به دو برابر شمار انتشارات در حوزه فناوری اطلاعات (جستجوی اول پایین عکس) بوده است.

نمونه کلیدواژگان یافت شده در مقالات یافت شده در جدول زیر آمده است:

جدول ۷- نمونه کلیدواژگان مقالات یافت شده در جستجوی نهایی

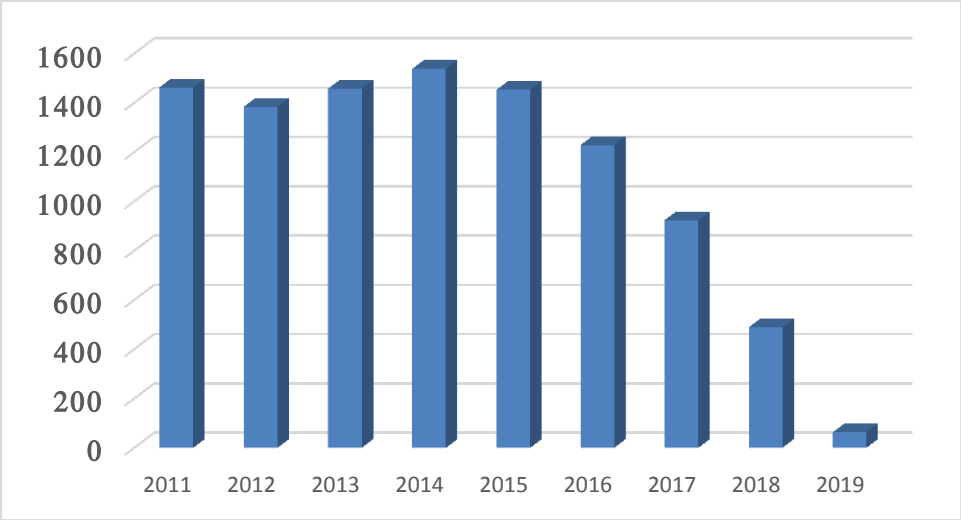
ردیف	کلیدواژه مقالات
۱	Business analytics; Practice of OR; Data science; Topic models; Computational literature review
۲	Business analytics; Data science; Business ethics canvas; Markkula; GDPR
۳	Text analytics; Multi-method approach; Technical features; Augmented reality; Post-adoption use intention
۴	Analytics; Ensemble learning; Predictive analysis; Hospital readmission; Natural language processing
۵	Machine learning; Physical activity status prediction; Multi-objective genetic programming; Hidden Markov model
۶	Data mining; Remanufactured products; Machine learning; Regression trees
۷	Decision support systems; Big data; Analytic infrastructure; Competence set; Deduction graph
۸	Student retention; Prediction; Elastic net; Bayesian Belief Network (BBN); Imbalance data

ردیف	کلیدواژه مقالات
۹	Analytics; Purchase prediction; Sales forecast; Non-contractual setting; Machine learning
۱۰	Analytics; Multi-class patients; Multi-priority patients; Dynamic scheduling; Advance scheduling
۱۱	Analytics; Online returns; Predictive model; Variable selection; LASSO; Machine learning
۱۲	Analytics; Deep learning; Deep neural networks; Managerial implications; Research agenda
۱۳	Business analytics; Technology affordances; Effective use; Value creation; Organizational transformation
۱۴	Analytics; Dynamic capabilities; Agility; Operations research
۱۵	Analytics; Innovation; Big Data; Data-driven culture; Absorptive capacity
۱۶	Analytics; Artificial intelligence; Data mining; Decision support systems; telecommunications; Validation of OR Computations
۱۷	Game theory; Crowdsourcing; Contest; All-pay auction
۱۸	Scheduling; Aircraft maintenance; Dynamic programming; Forward induction
۱۹	Inventory; Order expediting; Lead time flexibility; Service level
۲۰	Inventory; Approximate dynamic programming; Random yield; Tracking; Value of information
۲۱	energy; Demand-side management; Bilevel programming; Trilevel programming
۲۲	Group decisions and negotiations; Minimum cost consensus; Cost metric; Consensus Reaching Process; AFRYCA 3.0
۲۳	Discrete time optimization; Seasonal fisheries; Bioeconomics; Periodicity
۲۴	Control; S-shaped utility; Trading constraint; Value-at-Risk constraint; Defined contribution pension plan

۲-۷. دسته‌بندی حوزه‌ها با بهره‌گیری از فرکانس همزمانی استفاده از کلیدواژگان

با توجه به جستجوی بیش از ۸۰۰ هزار مدرک علمی در این پژوهش، تصمیم بر آن شد که بر اساس ارتباط پژوهش‌ها با حوزه جستجو ۱۰ هزار سند مرتبط انتخاب شوند. این انتخاب می‌تواند بر اساس سال انتشار، تعداد کلمات موجود در عنوان در جستجو، میزان ارجاع و شباهت باشد. معیار شباهت بر اساس تعداد کلمات مورد جستجو در عنوان و کلیدواژه با وزن بیشتر و در چکیده و کلیدواژه‌های تکمیلی با وزن کمتر است (Web of Science 2020).

میزان مدارک انتخابی به میزان ۱۰۰۰۰ مدرک (با توجه به پیشنهادیه این طرح و تایید خبرگان) و در طول زمان ۱۰ ساله به ترتیب زیر قرار گرفته است:

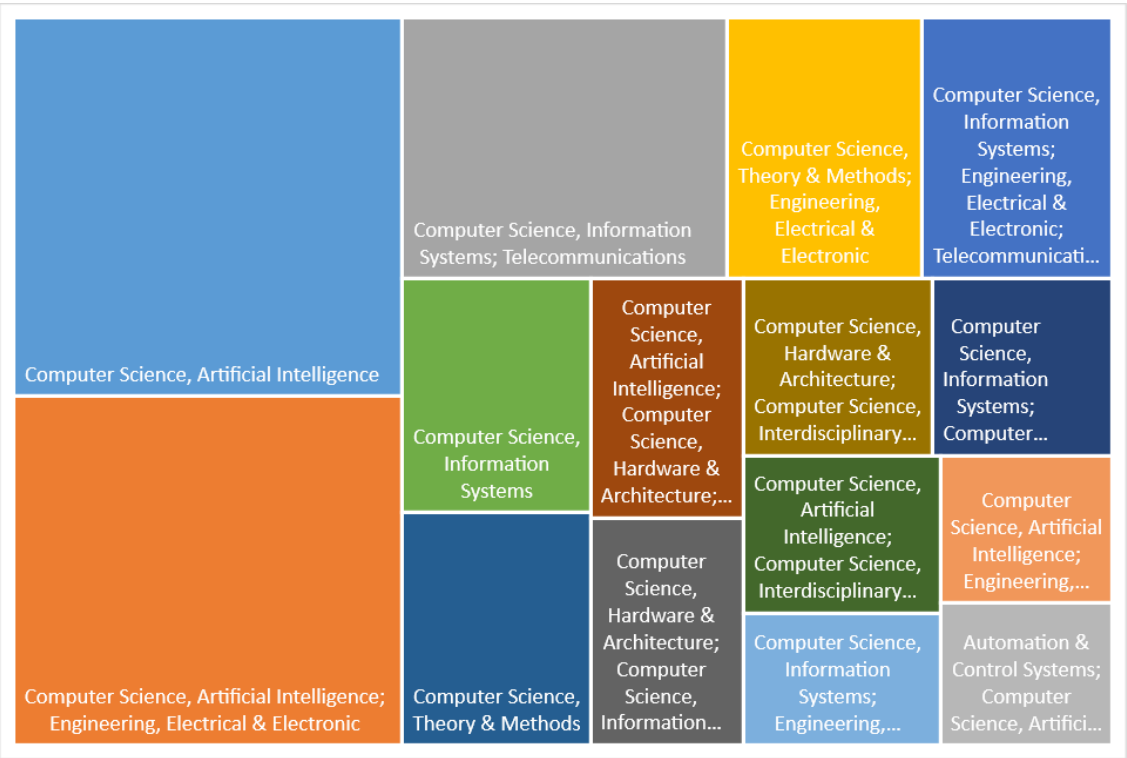


شکل ۱۰- میزان استخراج منبع بر حسب سال

کمتر بودن منابع در سال‌های اخیر به‌دلیل عدم تکمیل استنادات کافی به این مطالعات بوده است. زیرا معیار اولویت جستجوی و انتخاب، در این طرح، نزدیکی به موضع مورد جستجو است. همچنین عدم نمایه شدن تمام مطالعات در سال‌های اخیر و زمان‌بر بودن آن نیز ممکن است در این زمینه موثر باشد.

۲-۷-۱. بررسی چگونگی کاربرد کلیدواژگان

بر این اساس پانزده حوزه ترکیبی پرکاربرد در زمینه فناوری اطلاعات به شرح جدول زیر قرار گرفته است:



شکل ۱۱- حوزه‌های ترکیبی پرکاربرد

سه حوزه ترکیبی که بیش از ۲۵ درصد مطالعات در زمینه فناوری اطلاعات را به خود اختصاص داده‌اند شامل حوزه‌های ترکیبی زیر هستند:

- Computer Science, Artificial Intelligence
- Computer Science, Artificial Intelligence; Engineering, Electrical & Electronic
- Computer Science, Information Systems; Telecommunications

نکته قابل توجه حضور همزمان علوم کامپیوتر و هوش مصنوعی در صدر این مطالعات است. همچنین در ۱۵ حوزه برتر نمایش داده شده که بیش از نیمی از مطالعات را پوشش می‌دهند. زمینه علمی علوم کامپیوتر موجود است.

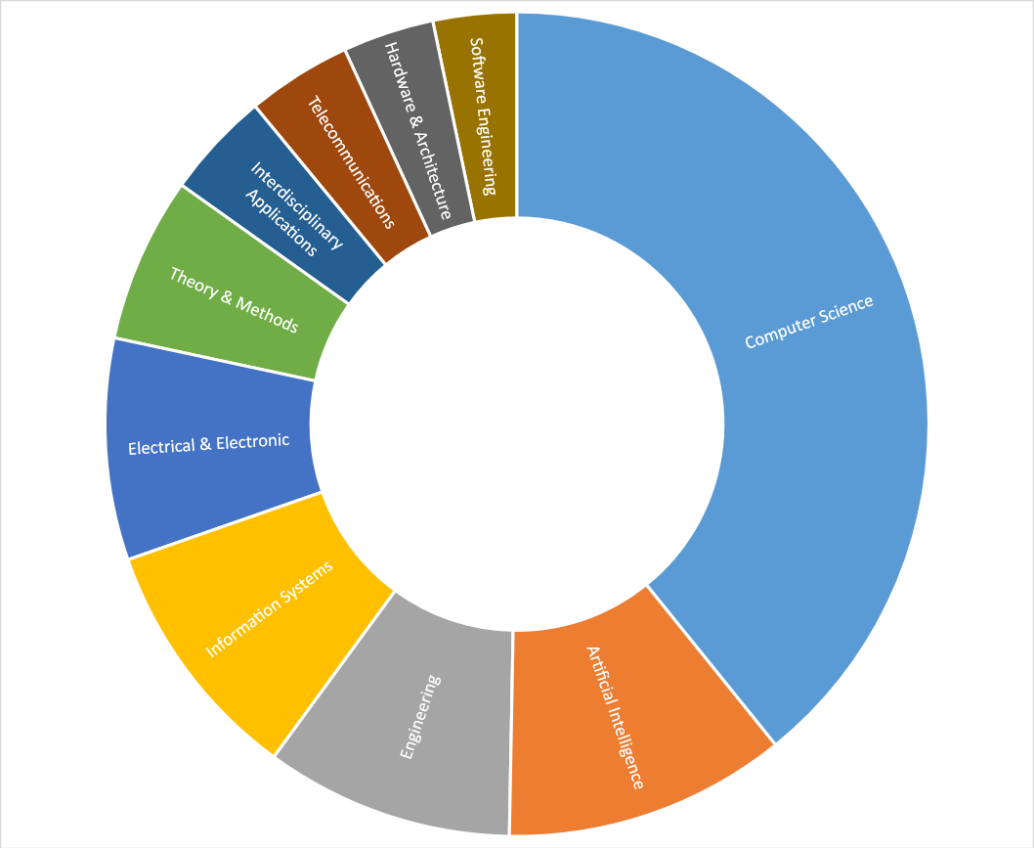
بر این اساس ۱۰ حوزه تخصصی برتر در جدول زیر آمده است. باقی حوزه‌های یافت شده کمتر از ۱ درصد مطالعات استخراجی را دربر داشته‌اند.

جدول ۸- حوزه‌های پر در مطالعات مرتبط با فناوری اطلاعات

ردیف	حوزه	درصد در ۱۰ حوزه برتر	درصد در کل منابع استخراجی
۱	Computer Science	39.2	35.6
۲	Artificial Intelligence	11.1	10.1
۳	Engineering	9.8	8.9
۴	Information Systems	9.6	8.7
۵	Electrical & Electronic	8.7	7.9
۶	Theory & Methods	6.5	5.9
۷	Interdisciplinary Applications	4.2	3.8
۸	Telecommunications	4.1	3.7
۹	Hardware & Architecture	3.6	3.3
۱۰	Software Engineering	3.3	3

لازم به ذکر است دو حوزه مهندسی (در حوزه‌های عمومی و کاربردی) و الکترونیک نیز جزو دسته‌بندی‌های WC در جدول پیوست بوده‌اند که با جستجوی مطالعات مرتبط با فناوری اطلاعات به لیست استخراجی اضافه شدند. با توجه به انتخاب لیست کوتاه توسط خبرگان، عدم درنظر گرفتن زمینه‌های فوق در جستجوی اولیه خللی در نتایج بدست آمده ایجاد نکرده است و دو دسته فوق جایگاه خود را در میان نتایج یافته‌اند.

در شکل زیر نسبت حوزه‌های ذکر شده در جدول فوق مصور شده است:



شکل ۱۲- درصد تمرکز بر حوزه‌ها در مطالعات فناوری اطلاعات

در گام بعدی ازین تحلیل، به نظر رسید دسته‌بندی مطالعات فناوری اطلاعات در دسته فوق می‌تواند به این سوال منجر شود که هر دسته فوق در نمونه مورد بررسی بیشتر از چه کلیدواژگانی بهره می‌جوید. بدین منظور دسته‌بندی‌های ذکر شده در هر مقاله در کنار کلیدواژه‌های مقاله قرار گرفته و بار دیگر تحلیل و بررسی احتمال حضور کلیدواژه‌های هم‌خانواده (برای هر دسته‌بندی) انجام شد. در جدول زیر کلیدواژه‌های مورد استفاده (بالاتر از ۵ درصد) برای هر دسته‌بندی پایگاه WOS استخراج شده است:

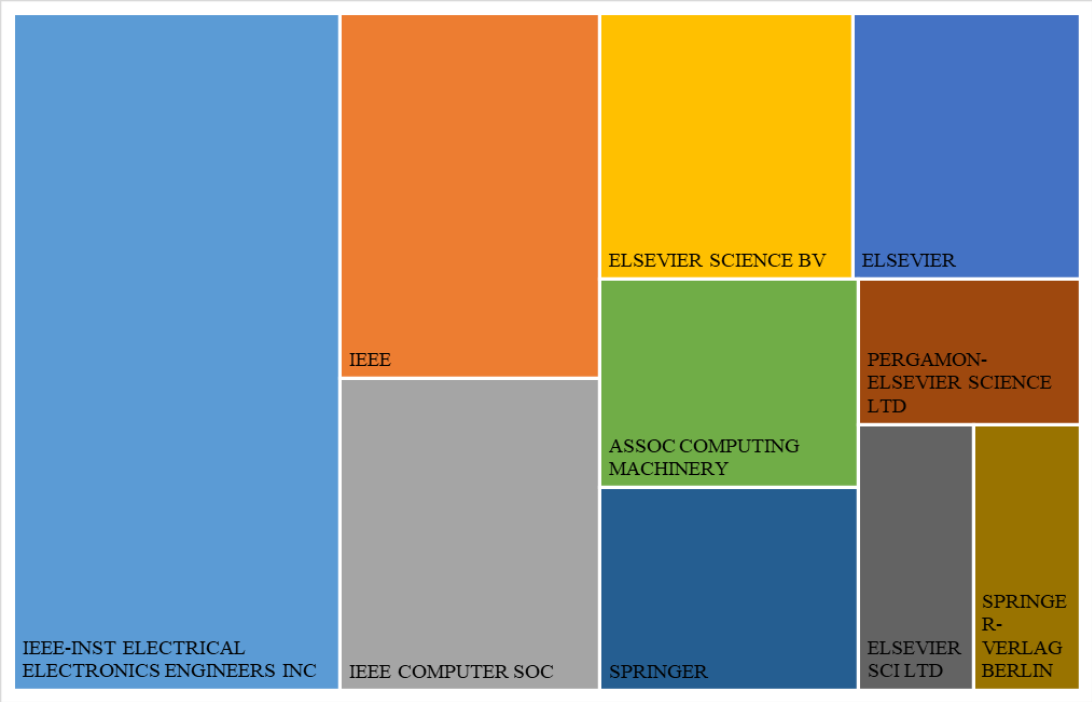
جدول ۹- کلیدواژه‌های مورد استفاده هر دسته پایگاه WOS

ردیف	دسته‌بندی فناوری اطلاعات	کلیدواژه‌های پر تکرار
۱	Computer Science	• Adaptive Dynamic Programming (ADP) • Neural Network (NN) • Optimal Control • Reinforcement Learning (RL) • Adaptive Control • Adaptive Neural Control

ردیف	دسته‌بندی فناوری اطلاعات	کلیدواژه‌های پر تکرار
۲	Theory & Methods	• Cloud Computing • Catoptics • Cloud Storage • Fully Homomorphic Encryption • Big Data
۳	Artificial Intelligence	• Feature Selection • Classification • Feature Extraction • Sparse Representation • Image Classification • Extreme Learning Machine
۴	Automation & Control Systems	• Industrial Informatics • Industrial Internet Of Things (IIOT) • Catchemistry • Catmultidisciplinary • Neural Networks (NNS) • Enterprise Systems • Extreme Learning Machine (ELM) • Neuro-Dynamic Programming • Soft Sensor • Industrial Wireless Sensor Networks (IWSNS)
۵	Hardware & Architecture	• Adaptive Neural Control • Cloud Computing • Neural Network (Nn) • Optimal Control • Reinforcement Learning (Rl) • Adaptive Dynamic Programming • Approximate Dynamic Programming • Attack Graph • Internet of Things
۶	Information Systems	• Security • Experimentation • Internet of Things (IOT) • Privacy • Wireless Sensor Networks • Algorithms • Interference Alignment • Social Media
۷	Electrical & Electronic	• Face Recognition • Physical-Layer Security • Biometrics • Interference Alignment • Network Coding • Physical Layer Security • Wiretap Channel

ردیف	دسته‌بندی فناوری اطلاعات	کلیدواژه‌های پر تکرار
۸	Interdisciplinary Applications	• Electronic Health Records • Cloud Forensics • Digital Forensics • Fuzzy Logic • Image Segmentation • Industrial Informatics • Industrial Internet of Things (IIOT) • Interactive Learning Environments • Media In Education • Medical Image Analysis • Pedagogical Issues
۹	Software Engineering	• Cloud Computing • Defect Prediction • Information Visualization • Software Defect Prediction • Visual Analytics • Attack Graph • Feature Detection • Multimedia
۱۰	Telecommunications	• Wireless Sensor Networks • Internet of Things (IOT) • Security • Authentication • Edge Computing

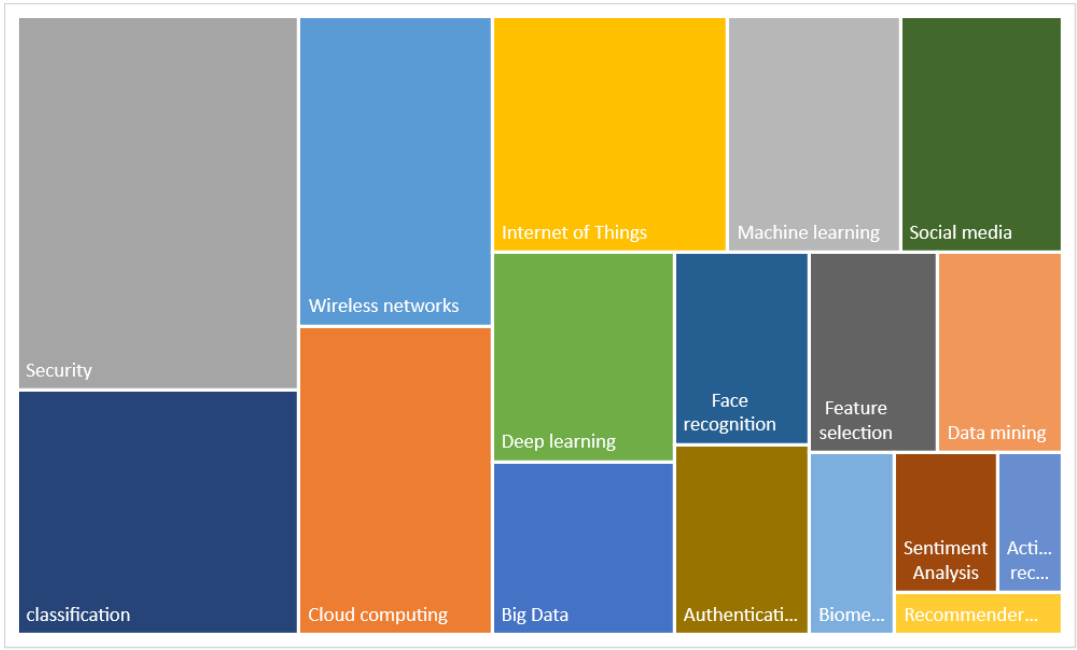
از بعد انتشارات در زمینه فناوری اطلاعات در بازه ذکر شده به این نکته باید توجه داشت که از بیش از ۱۲۰ منبع انتشار، نزدیک به نیمی از انتشارات در مجموعه‌های Elsevier و IEEE بوده است. ده دسته‌بندی انتشارات برتر که بیش از ۸۰ درصد فعالیت‌ها را منتشر نموده‌اند، در شکل زیر نمایش داده شده‌اند:



شکل ۱۳- انتشارات فعال در زمینه فناوری اطلاعات

۲-۷-۲. بررسی روند کاربرد کلیدواژگان و تمرکز بر حوزه‌های مختلف

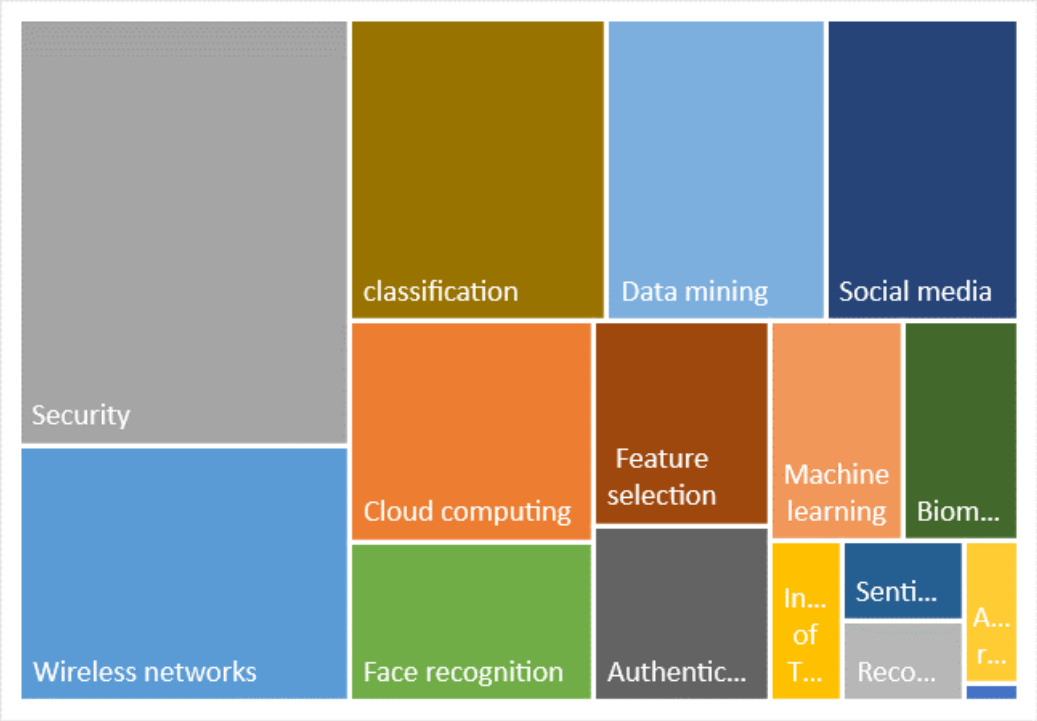
با توجه به فرکانس کلیدواژه‌های مقالات در ۱۰هزار مجله مرتبط با جستجو، برای ترسیم ۱۵ کلیدواژه برتر به به نمودار زیر می‌رسیم:



شکل ۱٤- کلیدواژه‌های پرکاربرد حوزه فناوری اطلاعات در ده سال

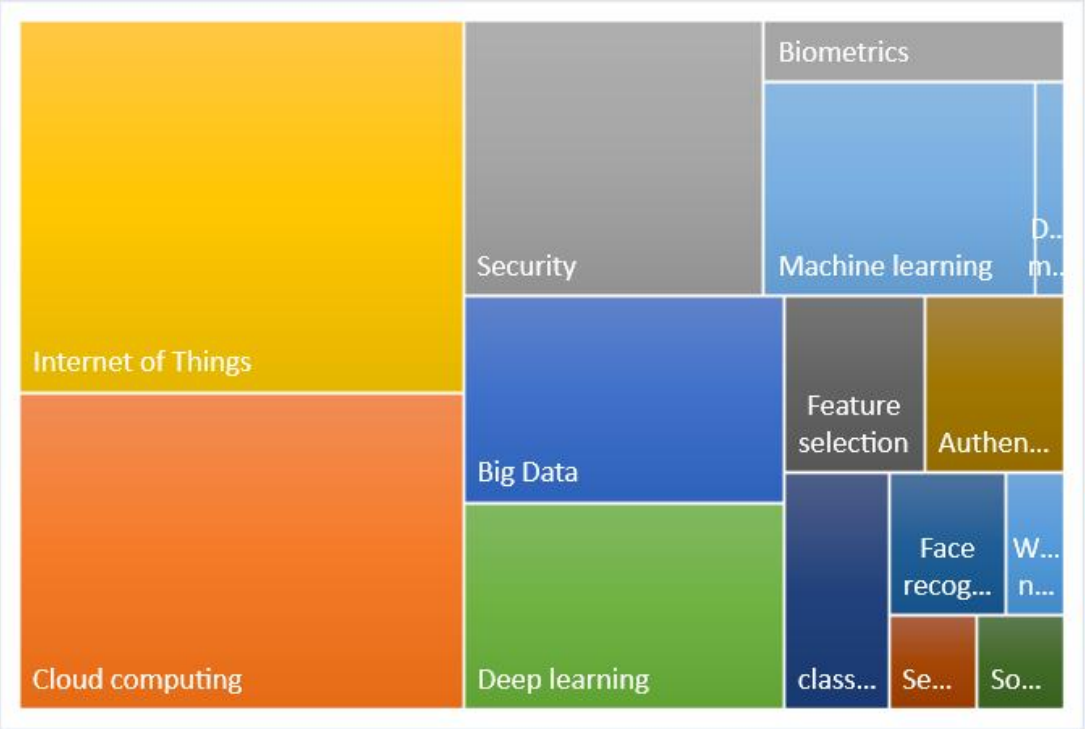
با توجه به نمودار فوق می‌توان مشاهده نمود که مباحث امنیت، طبقه‌بندی و پردازش ابری جایگاه ویژه‌ای را در میان کلیدواژگان پرکاربرد در زمینه فناوری اطلاعات دارند. این دسته‌بندی با توجه به

پیشنهاده طرح در بازه ۱۰ ساله مورد بررسی قرار گرفته است. حال اینکه کلیدواژه‌های پرکاربرد در زمینه فناوری اطلاعات با توجه به پویایی زیاد این زیمنه علمی ممکن است در طول زمان دچار تغییرات گردد. بدین منظور اگر بخواهیم نگاهی به تعریف کلیدواژه‌ها در ابتدا و انتهای بازه مورد بررسی ۲۰۱۱ و ۲۰۱۹ بیاندازیم به نمودارهای زیر می‌رسیم:



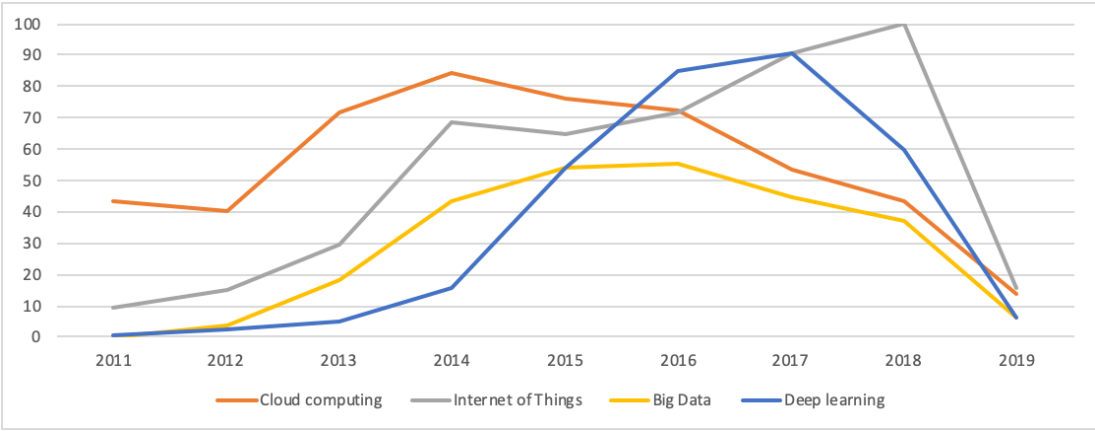
شکل ۱۵- کاربرد و تکرار کلیدواژه‌ها در سال ۲۰۱۱

همانگونه که در شکل فوق دیده می‌شود، ۴ حوزه امنیت، شبکه‌های بی‌سیم، طبقه‌بندی و داده‌های عظیم بزرگترین حوزه‌های تمرکز کلیدواژه‌های فناوری اطلاعات در سال ۲۰۱۱ هستند. حال اگر بخواهیم همین گزارش را در سال ۲۰۱۹ به تحلیل بکشیم داریم:



شکل ١٦ -کاربرد و تکرار کلیدواژه‌ها در سال ۲۰۱۹

همانگونه که در شکل فوق ملاحظه می‌شود حوزه‌های نوظهوری همانند اینترنت اشیا و محاسبات ابری و داده‌های عظیم در کنار حوزه‌های کاری مانند یادگیری ماشین و یادگیری عمیق تعریف حوزه‌های فناوری اطلاعات را در بازه زمانی جدید تشکیل داده‌اند. با توجه به دو شکل فوق رشد ۴ حوزه‌ی جدید اینترنت اشیا، محاسبات ابری و داده‌های بزرگ و یادگیری عمیق را به شکل (در مقیاس ۱۰۰) زیر خواهیم داشت:



شکل ١٧- روند رشد ٤ حوزه نوظهور کشف شده در پژوهش

همانگونه که مشاهده می‌شود رشد این زمینه علمی بویژه در حوزه‌های اینترنت اشیا و یادگیری عمیق بسیار مشهود است و زمینه‌های علمی مانند اینترنت اشیا همچنان در حال ادامه است. تعداد کمتر مشاهدات علمی در سال ۲۰۱۹ به علت کامل نشدن تمام انتشارات (یا آنلاین نشدن آنها) است.

جدول ۱۰- مقایسه کلیدواژه‌های پرکاربرد در دو بازه زمانی

کلیدواژه	درصد در بین کلیدواژگان پرکاربرد در سال ۲۰۱۱	درصد در بین کلیدواژگان پرکاربرد در سال ۲۰۱۹	رشد
Internet of Things	1.71	22.99	21.28
Cloud Computing	7.85	19.54	11.69
Big Data	0.00	9.20	9.20
Deep Learning	0.14	9.20	9.05
Machine Learning	4.28	10.34	6.07
Sentiment Analysis	1.43	1.15	-0.28
Authentication	4.56	3.45	-1.12
Action Recognition	1.14	0.00	-1.14
Recommender Systems	1.43	0.00	-1.43
Feature Selection	5.28	3.45	-1.83
Face Recognition	5.71	2.30	-3.41
Biometrics	3.71	0.00	-3.71
Social Media	8.56	1.15	-7.41
Classification	11.41	3.45	-7.96
Data Mining	9.70	1.15	-8.55
Security	20.68	11.49	-9.19
Wireless Networks	12.41	1.15	-11.26

در این میان اینترنت اشیا با بیشترین رشد و شبکه‌های بی‌سیم با بیشترین نزول در میان برترین کلیدواژگان مجموع اخیر قرار گرفته‌اند. همچنین می‌توان نتیجه گرفت ۵ حوزه زیر نقش و جایگاه بالاتری در آینده فناوری اطلاعات خواهند داشت:

- Internet of Things
- Cloud Computing
- Big Data
- Deep Learning
- Machine Learning

حال می‌توانیم کلیدواژه‌های هم‌خانواده با کلیدواژه‌های فوق را یافت نماییم. در حقیقت با مدل‌سازی موضوعی کلیدواژه‌های برتر یا پرتکرار می‌توان احتمال همراهی دیگر کلیدواژگان را با این کلیدواژه‌ها بدست آورد. در جدول زیر احتمال حضور کلیدواژه دیگر در کلیدواژه‌های پرتکرار

نمایش داده شده است. لازم به ذکر است میزان خویشاوندی برابر با احتمال حضور کلیدواژه در کنار دیگر کلیدواژگان در هر رکورد است:

جدول ۱۱- خویشاوندی کلیدواژگان

ردیف	کلیدواژه	هم‌خانواده	احتمال حضور
1	Cloud Computing	Virtualization	0.20
2		Big Data	0.18
3		Data Sharing	0.13
4		Fog Computing	0.12
5		Data Center	0.11
6		Data Security	0.11
7		Ranked Search	0.11
8		Searchable Encryption	0.11
9		Cloud Manufacturing	0.10
10		Privacy	0.10
11	Big Data	Cloud Computing	0.18
12		Data Analytics	0.17
13		Mapreduce	0.17
14		Hadoop	0.16
15		Analytics	0.15
16		Web And Internet Services	0.13
17		Web Technologies	0.13
18		Internet Of Things	0.11
19		Industry 4.0	0.10
20	Deep Learning	Convolutional Neural Networks	0.28
21		Convolutional Networks	0.15
22		Convolutional Neural Network	0.15
23		Convolution Neural Network	0.10
24		Lstm	0.10
25		Model Combination	0.10
26		Transfer Learning	0.10
69	Machine Learning	Artificial Intelligence	0.11
70		Data Mining	0.11
71		Intelligent Sensors	0.10
72	Social Media	Sentiment Analysis	0.17
73		Twitter	0.15
74		Web 2.0	0.14
75		Collaborative Tagging	0.13
76		Competitive Analytics	0.13
77		Czech Language	0.13
78		Political Participation	0.13
79		Word-Of-Mouth	0.13
80		Information Diffusion	0.11
81		Online Communities	0.11
82		User-Generated Content	0.11
83		Text Mining	0.10
84	Authentication	Key Agreement	0.25
85		Smart Card	0.25
86		Password	0.24
87		Proverif	0.21
88		Anonymity	0.17
89		Security	0.17
90		Biometrics	0.15
91		Elliptic Curve Cryptosystem	0.15
92		Key Establishment	0.15
93		Impersonation Attack	0.14
94		Anonymous	0.13

ردیف	کلیدواژه	هم‌خانواده	احتمال حضور
27	Internet Of Things	Smart Cities	0.14
28		Web Of Things	0.13
29		Connected Objects	0.12
30		Electronic Devices	0.12
31		Iot Technologies	0.12
32		Iot Weaknesses	0.12
33		Privacy Issues	0.12
34		Risks Issues	0.12
35		Web And Internet Services	0.12
36		Web Technologies	0.12
37		Big Data	0.11
38		Fog Computing	0.11
39		Security	0.11
40		Smart Gateway	0.11
41		Blockchain	0.10
42		Dtls	0.10
43		Iot Marketplace	0.10
44		Machine-To-Machine Communications	0.10
45	Security	Privacy	0.31
46		Authentication	0.17
47		Confidentiality	0.11
48		Internet Of Things	0.11
49		Attacks	0.10
50	Classification	Regression	0.16
51		Ensembles	0.13
52	Networks	Capacity Scaling	0.20
53		Through-Wall Surveillance	0.20

ردیف	کلیدواژه	هم‌خانواده	احتمال حضور
95		Bloom Filter	0.12
96		Elliptic Curve Cryptography	0.12
97		Ethereum	0.12
98		Global Mobility Networks	0.12
99		Graphical Passwords	0.12
100		Identity-Based Signature	0.12
101		Non-Interactive Witness-Indistinguishable	0.12
102		Non-Interactive Zero-Knowledge Proof	0.12
103		Roaming	0.12
104		Vehicular Ad-Hoc Networks	0.12
105		Avispa	0.11
106		Privacy	0.11
107		Ban Logic	0.10
108		Unlinkability	0.10
109	Feature Selection	Mutual Information	0.16
110		Binary Particle Swarm Optimization	0.11
111		Filters	0.11
112		Rough Set Theory	0.11
113		Sparse Learning	0.11
114		Instance Selection	0.10
115	Face Recognition	Feret	0.16
116		Cas-Peal-R1	0.14
117		Dog Filter	0.14
118		Face Images	0.14
119		Feature Descriptors	0.14
120		Gabor Descriptor	0.14
121		Global-Local Hog Feature Fusion	0.14

ردیف	کلیدواژه	هم‌خانواده	احتمال حضور
54		Caching	0.16
55		Network Information Theory	0.16
56		11 Standards	0.14
57		Ant Colony	0.14
58		Anypath Routing	0.14
59		Broadcast-Relay Channels	0.14
60		Broadcast Network	0.14
61		Colluding Eavesdroppers	0.14
62		Computer Intrusion Detection	0.14
63		Congestion Games	0.14
64		Data And Communication Security	0.14
65		Data Center Placement	0.14
66		Data Delivery	0.14
67		Degrees-Of-Freedom	0.14
68		Delayed Effects	0.14

ردیف	کلیدواژه	هم‌خانواده	احتمال حضور
122		Global Classifier	0.14
123		Histogram-Of-Oriented Gradient Descriptor	0.14
124		Hog Descriptor	0.14
125		Large-Scale Face Database	0.14
126		Lbp Descriptor	0.14
127		Local Classifier	0.14
128		Nearest Neighbour Classifier	0.14
129		Weighted Sum Rule	0.14
130		Feature Extraction	0.12
131		Forensic Sketch	0.12
132		Heterogeneous Face Recognition	0.12
133		Sparse Representation	0.12
134		Discriminant Analysis	0.11
135		Local Descriptors	0.11
136		Heterogeneous Face Matching	0.10

۲-۷-۳. دسته‌بندی حوزه‌های علمی بر اساس مدل‌سازی موضوعی کلیدواژه‌ها

در بخش «بررسی چگونگی کاربرد کلیدواژگان»، بر اساس دسته‌بندی‌های انجام شده پایگاه، کلیدواژه‌های مرتبط با هر زمینه فناوری اطلاعات مشخص شد. با توجه به اینکه این حوزه‌های مفهومی توسط خود پایگاه اختصاص می‌یابد، میزان دقت و یا درستی این دسته‌بندی مشخص نیست. در ادامه این طرح پژوهشی، با در نظر نگرفتن تمامی دسته‌بندی‌های موضوعی پایگاه و فقط درنظر گرفتن روابط بین کلیدواژه‌ها (رابطه وجود در یک مقاله)، به بررسی دسته‌بندی بر اساس کمترین احتمال پراکندگی با روش مدل‌سازی موضوعی می‌پردازیم. مدل‌سازی موضوعی از روش‌های کارای تحلیل متن است که هدف آن گروه‌بندی مجموعه‌ای از متون مختلف به‌صورت زیرمجموعه‌های معنادار که موضوع نامیده می‌شوند است. الگوریتم‌های روش‌های مدل‌سازی موضوعی با حداقل دخالت انسان این فرایند را انجام می‌دهند و همین نکته باعث جذابیت و استفاده روزافزون آن در تحلیل حجم عظیمی از متون شده است. در این روش بجای استفاده از عناوین از قبل تعیین شده، تنها با تعیین تعداد

موضوعات، زیرمجموعه‌های موضوعی استخراج شده و میزان ارتباط هر سند (شامل مقالات منتشر شده در مجلات پایگاه) با هر زیرمجموعه نیز تعیین می‌شود (John W and Petko 2013).

روش‌های استخراج موضوع بر مبنای این فرض عمل می‌کنند که کلمات موجود در یک متن از نظر معنایی به یکدیگر وابسته هستند و معنایی که در مستندات یک موضوع مسنجم وجود دارد از مجموعه کلمات این مستندات به دست آمده است. به عبارتی صرف‌نظر از معانی یا مکان قرارگیری متن، هم رخدادی کلمات مورد تحلیل قرار گرفته و اسناد به‌صورت مجموعه‌ای از کلمات در نظر گرفته می‌شود. بنابر این و مطابق با ادبیات در مدل‌سازی موضوعی مفروضات زیر در نظر گرفته می‌شود (Blei, Andrew and Jordan 2003):

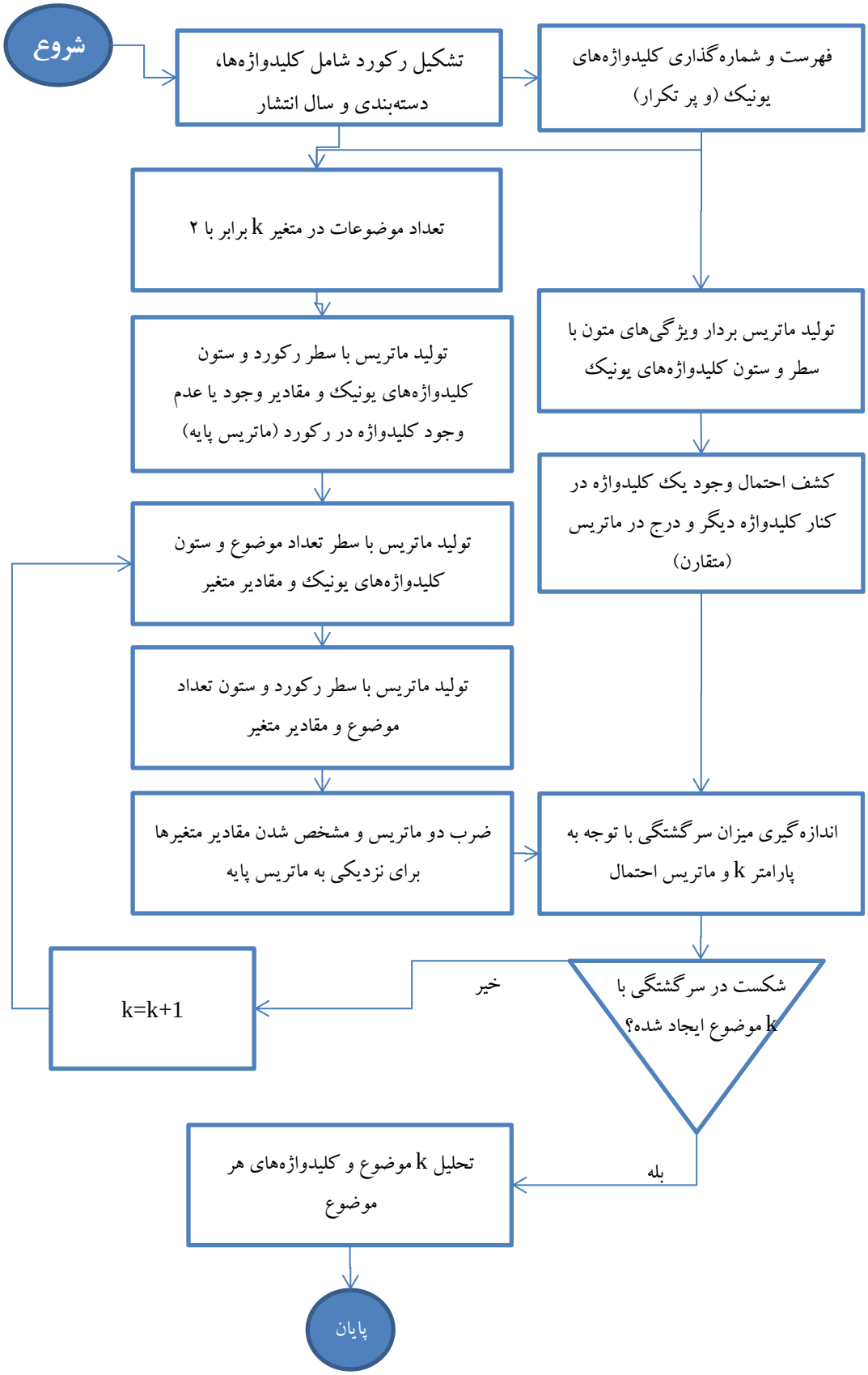
۱. ترتیب حضور مستندات در پیکره اهمیتی ندارد؛
۲. ترتیب قرار گرفتن کلمات در هر مستند اهمیتی ندارد؛
۳. موضوعات توزیع چندجمله‌ای از کلمات موجود در یک واژه‌نامه هستند؛
٤. کلمات یک مستند از تعدادی موضوع پنهان نشأت گرفته‌اند.

پیدا کردن موضوعات یک سند حتی برای اسنادی که توسط خبرگان نمایه‌سازی شده و موضوعی به آن‌ها اختصاص یافته است، موردتوجه است. هر سند از تعدادی واژه تشکیل شده است و میزان ارتباط یک سند با یک موضوع به میزان تکرار کلمات مرتبط با آن موضوع در سند وابسته است. با عبارتی اگر میزان اهمیت یا وزن یک کلمه در یک سند را w نشان دهیم و یک سند شامل m واژه باشد که هر یک وزن و اهمیت خاص خود را برای آن سند دارند و با فرض اینکه پیکره متنی موردنظر شامل n مستند است. آنگاه LDA می‌تواند تعداد k موضوع را شناسایی کند به‌نحوی‌که (D. M. Blei 2012):

$$\begin{array}{c} \\ \text{Term}_1 \\ \text{Term}_2 \\ \vdots \\ \text{Term}_m \end{array} \begin{array}{c} \begin{array}{cccc} \text{Doc}_1 & \text{Doc}_2 & \dots & \text{Doc}_n \end{array} \\ \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,n} \\ w_{2,1} & w_{2,2} & \dots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m,1} & w_{m,2} & \dots & w_{m,n} \end{bmatrix} \end{array} \xrightarrow{\text{LDA(BoW(Term,Doc), k)}}$$

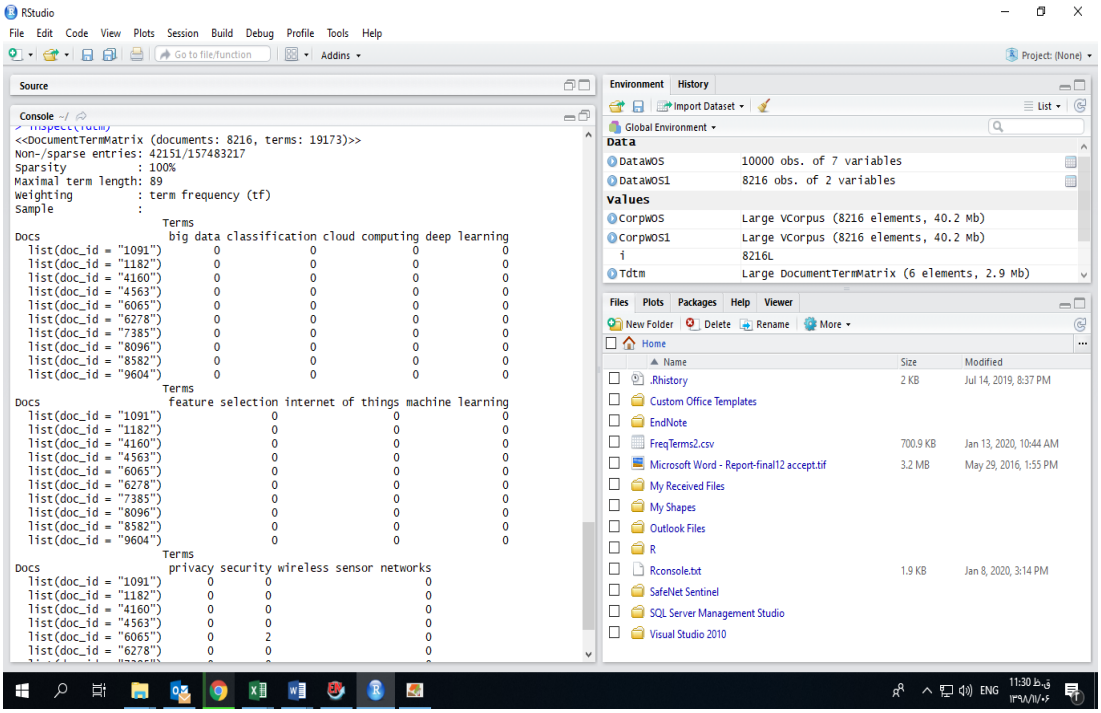
$$\begin{array}{c} \\ \text{Term}_1 \\ \text{Term}_2 \\ \vdots \\ \text{Term}_m \end{array} \begin{array}{c} \begin{array}{cccc} \text{Topic}_1 & \text{Topic}_2 & \dots & \text{Topic}_k \end{array} \\ \begin{bmatrix} \beta_{1,1} & \beta_{1,2} & \dots & \beta_{1,k} \\ \beta_{2,1} & \beta_{2,2} & \dots & \beta_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{m,1} & \beta_{m,2} & \dots & \beta_{m,k} \end{bmatrix} \end{array} * \begin{array}{c} \begin{array}{cccc} \text{Doc}_1 & \text{Doc}_2 & \dots & \text{Doc}_n \end{array} \\ \begin{array}{c} \text{Topic}_1 \\ \text{Topic}_2 \\ \vdots \\ \text{Topic}_k \end{array} \begin{bmatrix} \theta_{1,1} & \theta_{1,2} & \dots & \theta_{1,n} \\ \theta_{2,1} & \theta_{2,2} & \dots & \theta_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{k,1} & \theta_{k,2} & \dots & \theta_{k,n} \end{bmatrix} \end{array}$$

در زیر الگوریتم اجرای روش به تصویر کشیده شده است:



شکل ۱۸- فلوچارت اجرای الگوریتم LDA

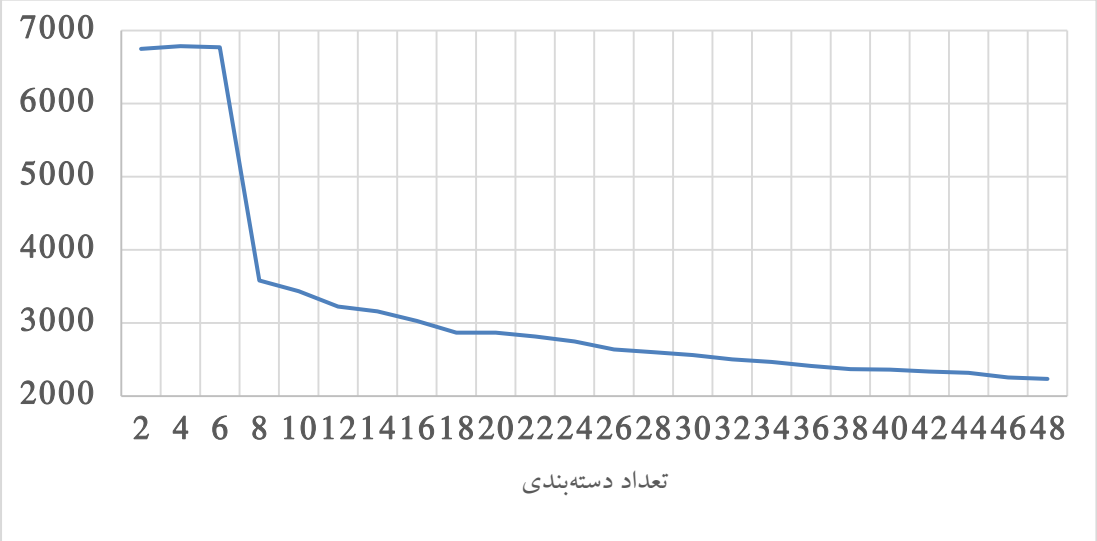
کد این روش نیز در نرم‌افزار R توسعه داده شده است:



شکل ۱۹- نمای خروجی مدل‌سازی موضوعی

همانگونه که در شکل قابل نمایش است، مدلسازی با تمامی مقادیر نمونه مورد بررسی در این تحقیق انجام شده است. بر اساس ادبیات، میزان درست یا مناسب دسته‌ها می‌تواند از ملاک‌های زیادی پیروی کند. یکی از این معیارها میزان سرگشتگی (Perplexity) جواب در تعداد دسته‌های انتخابی است. پارامتر ورودی مهم دیگر LDA، مقدار k است که تعداد موضوعات مدنظر را نشان می‌دهد. روش‌های مختلفی برای تعیین تعداد موضوعات وجود دارد (Hall, Jurafsky and Manning 2008). مزیت انتخاب تعداد موضوعات زیاد، این است که تمامی حوزه‌های موضوعی پژوهش پوشش داده خواهد شد، از طرفی مزیت تعداد حوزه‌های محدود این است که خبرگان و سیاست گذاران درک بهتری از تعداد موضوعات محدود خواهند داشت و تفسیر و تحلیل موضوعات راحت تر خواهد بود (De Battisti, Ferrara and Salini 2015). ازاین‌رو انتخاب معیاری برای تعیین تعداد موضوعات مناسب بسیار ضروری است. برای این منظور از معیار سرگشتگی استفاده می‌شود. سرگشتگی یک معیار اندازه‌گیری در مدل‌های آماری است که برای بررسی میزان مناسب بودن یک توزیع احتمالی یا پیش‌بینی مدل احتمالی یک نمونه مورد استفاده قرار می‌گیرد و می‌توان از آن برای مقایسه مدل‌های احتمالی استفاده کرد. برای ارزیابی مدل‌سازی موضوعی از سرگشتگی اسناد مشاهده نشده استفاده می‌شود که تابعی نزولی از لگاریتم احتمال وقوع اسناد مشاهده نشده است و هر چه مقدار آن کمتر باشد نشان می‌دهد که مدل بهتر است و قابلیت تعمیم‌پذیری بالاتری دارد

(D. M. Blei 2012) و مقدار بالاتر سرگشتگی نشان‌دهنده این است که عبارت‌ها در زمان یادگیری مدل به موضوعات مناسبی تخصیص نیافته‌اند (De Battisti, Ferrara and Salini 2015). بدین‌منظور کمترین میزان به‌صورت معمول برای انتخاب اندازه دسته مناسب برگزیده می‌شود. در شکل زیر میزان شاخص برای دسته‌های ۲ تا ۵۰ نشان می‌دهد.



شکل ۲۰- میزان انحراف سرگشتگی با توجه به اندازه دسته‌بندی

همانگونه که از شکل فوق بر می‌آید، تعداد دسته‌های مورد انتظار از ۸ به بعد با افزایش تعداد به ندرت باعث کاهش در شاخص می‌شود. این کاهش تدریجی به علت تنوع بالا و حجم زیاد کلیدواژه‌های مورد استفاده است. زیرا با تعداد دسته‌بندی بالا می‌توان حجم زیادی از آنها را در یک محیط پراکنده در یک گروه قرار داد. بر این اساس، تعداد گروه ۸ که نقطه شکست نمودار است، به عنوان تعداد دسته‌های مدلسازی موضوعی انتخاب می‌شود. بر این اساس ۸ دسته ارتباطات کلیدواژه‌های فناوری اطلاعات برای تشکیل ۸ موضوع در جدول زیر (نمایش ده کلیدواژه پر احتمال برای هر موضوع) ترسیم می‌شوند. با توجه به جدول زیر به‌نظر می‌رسد حوزه شناسایی شده اول، حوزه علوم توسعه یافته با اینترنت اشیا، موضوع دوم مرتبط با مباحث یادگیری ماشین، موضوع سوم حوزه داده‌های عظیم، موضوع چهارم حوزه محاسبات ابری، موضوع پنجم حوزه تشخیص الگو و دید ماشین، موضوع ششم طراحی شبکه، موضوع هفتم شبکه‌های اجتماعی و کاربرد فناوری و نهایتا موضوع هشتم مرتبط با توسعه نرم‌افزار در علوم دیگر باشد. لازم به ذکر است به‌منظور مقایسه بهتر و جلوگیری از تاثیر مقالات ناشی از فعالیت‌های علمی پژوهشگران داخلی، مقالات مربوط به این پژوهش‌های داخلی (نویسنده اول یا نویسنده مسئول از ایران)، از نمونه مطالعات استخراج شده حذف شدند.

جدول ۱۲- دسته‌بندی موضوعات با روش مدلسازی موضوعی

شماره موضوع/ کلیدواژه	موضوع ۱	موضوع ۲	موضوع ۳	موضوع ۴	موضوع ۵	موضوع ۶	موضوع ۷	موضوع ۸
نام پیشنهادی	یادگیری و انتخاب ویژگی	اینترنت اشیا	شبکه	رمزنگاری و امنیت	کاربردهای فناوری اطلاعات و اینترنت اشیا	پردازش تصویر	کلان‌داده	یادگیری ماشین
۱	cloud computing	wireless sensor networks	security	classification	machine learning	feature selection	cloud computing	machine learning
۲	face recognition	deep learning	wireless sensor networks	security	cloud computing	pattern recognition	big data	security
۳	online learning	machine learning	deep learning	face recognition	data mining	biometrics	privacy	deep learning
۴	data mining	big data	internet of things	wireless sensor networks	feature selection	computer vision	internet of things	privacy
۵	internet of things	face recognition	cloud computing	privacy	feature extraction	authentication	deep learning	feature selection
۶	text mining	feature selection	machine learning	internet of things	authentication	security	security	wireless sensor networks
۷	classification	privacy	authentication	cloud computing	sentiment analysis	deep learning	social media	cloud computing
۸	privacy	clustering	multi-task learning	internet of things (iot)	social media	internet of things	multi-task learning	image classification
۹	feature extraction	neural networks	neural networks	pattern recognition	internet of things (iot)	support vector machine	sparse representation	internet of things (iot)
۱۰	Wireless Sensor Network	Recommender Systems	Adaptive Control	Cryptography	Extreme Learning Machine	Measurement	Anonymity	Face Recognition

۸-۲. جمع‌بندی و نتیجه‌گیری

در این فصل به بررسی و مطالعه حوزه‌های (موضوعات) فناوری اطلاعات در مطالعات بین‌المللی پرداخته شد. در این بررسی از روش‌های مختلفی برای دستیابی به نتایج استفاده شد. در بخش اول و در انتخاب مجلات پیشرو در زمینه فناوری اطلاعات، با وجود تاکید در پیشنهاده بر استفاده از مجلات پر استناد و پیشرو (اچ ایندکس بالا) برای تشکیل سبد واژگان کلیدی اولیه، با این‌حال از خبرگان نیز در مورد مجلات پرسیده شده و ۱۱ مورد مجله دیگر به فهرست جستجوی اولیه افزون شد. این گام باعث گردید سبد واژگان اولیه معتبر گردد. همچنین در فاز دوم علی‌رغم وجود دسته‌بندی موضوعی در پایگاه مورد مطالعه (WOS) و دسته‌بندی کلیدواژه‌های پرکاربرد در این دسته‌ها در این تحقیق، روش مدلسازی موضوعی نیز مورد توجه قرار گرفت و محققین فارغ از نتایج بدست آمده در توزیع کلیدواژه‌ها در دسته‌بندی مذکور، مجدد دسته‌بندی را بر اساس ادبیات ذکر شده در بخش قبل به انجام رساندند. با توجه به افت ضریب سرگشتگی در عدد ۸، هشت دسته‌بندی برای موضوعات ترکیبی کلیدواژه‌های مرسوم انتخاب شده و این دسته‌بندی در بخش قبل به نمایش درآمد. مقایسه دسته‌بندی فوق با دسته‌بندی پایگاه در جدول زیر آمده است:

جدول ۱۳- مقایسه دسته‌بندی پایگاه WOS و دسته‌بندی انجام شده

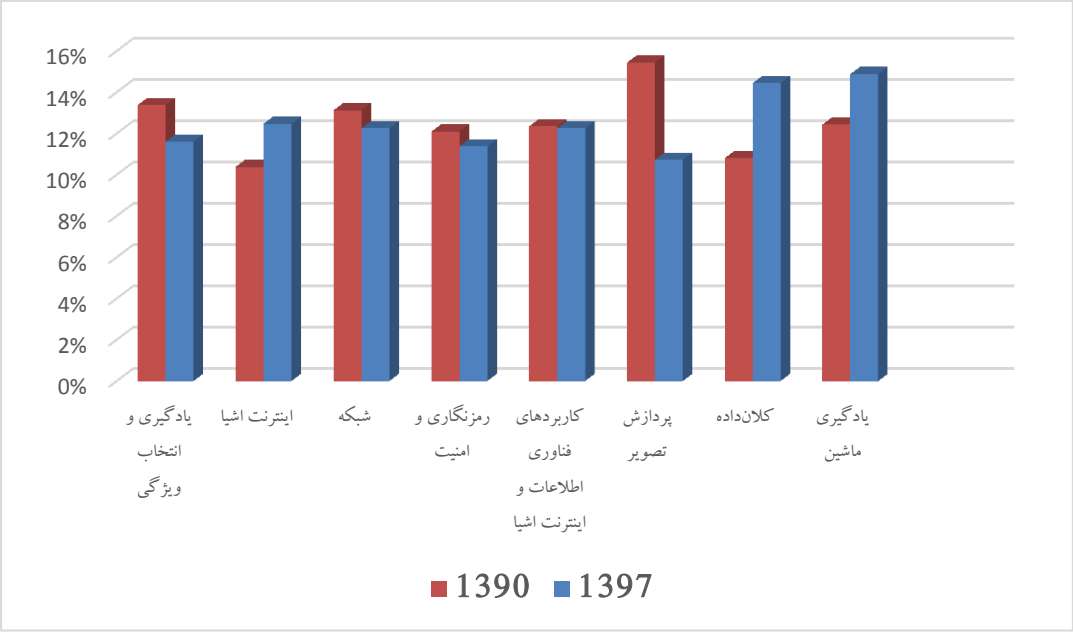
ردیف	دسته‌بندی فناوری اطلاعات پایگاه	کلیدواژه‌های پر تکرار	دسته‌بندی استخراج شده
۱	Computer Science	• Adaptive Dynamic Programming (ADP) • Neural Network (NN) • Optimal Control • Reinforcement Learning (RL) • Adaptive Control	موضوع ۸
۲	Theory & Methods	• Cloud Computing • Catoptics • Cloud Storage • Fully Homomorphic Encryption • Big Data	موضوع ۷
۳	Artificial Intelligence	• Feature Selection • Classification • Feature Extraction • Sparse Representation • Image Classification • Extreme Learning Machine	موضوع ۸

ردیف	دسته‌بندی فناوری اطلاعات پایگاه	کلیدواژه‌های پر تکرار	دسته‌بندی استخراج شده
۴	Automation & Control Systems	• Industrial Informatics • Industrial Internet Of Things (IIOT) • Catchemistry • Catmultidisciplinary • Neural Networks (NNS) • Enterprise Systems • Extreme Learning Machine (ELM) • Neuro-Dynamic Programming • Soft Sensor • Industrial Wireless Sensor Networks (IWSNS)	موضوع ۱
۵	Hardware & Architecture	• Adaptive Neural Control • Cloud Computing • Neural Network (Nn) • Optimal Control • Reinforcement Learning (Rl) • Adaptive Dynamic Programming • Approximate Dynamic Programming • Attack Graph • Internet of Things	موضوع ۲
۶	Information Systems	• Security • Experimentation • Internet of Things (IOT) • Privacy • Wireless Sensor Networks • Algorithms • Interference Alignment • Social Media	موضوع ۴
۷	Electrical & Electronic	• Face Recognition • Physical-Layer Security • Biometrics • Interference Alignment • Network Coding • Physical Layer Security • Wiretap Channel	موضوع ۶

ردیف	دسته‌بندی فناوری اطلاعات پایگاه	کلیدواژه‌های پر تکرار	دسته‌بندی استخراج شده
۸	Interdisciplinary Applications	• Electronic Health Records • Cloud Forensics • Digital Forensics • Fuzzy Logic • Image Segmentation • Industrial Informatics • Industrial Internet of Things (IIOT) • Interactive Learning Environments • Media In Education • Medical Image Analysis • Pedagogical Issues	موضوع ۵
۹	Software Engineering	• Cloud Computing • Defect Prediction • Information Visualization • Software Defect Prediction • Visual Analytics • Attack Graph • Feature Detection • Multimedia	موضوع ۱
۱۰	Telecommunications	• Wireless Sensor Networks • Internet of Things (IOT) • Security • Authentication • Edge Computing	موضوع ۳

با توجه به جدول فوق می‌توان یافت که در زمینه فناوری اطلاعات موضوعات در حال ترکیب و میان در دسته‌های مختلف فناوری اطلاعات در حال تغییر است.

با توجه به سرعت تغییرات در زمینه فناوری اطلاعات می‌بایست اطمینان یافت که تمرکز کلیدواژگان در طی مدت زمانی ده ساله چگونه در مجموعه دسته‌ها تغییر می‌کند. بدین منظور در شکل زیر تعداد مدارک بازیابی شده در هر دسته در ابتدا و انتهای بازه زمانی تحقیق مورد مطالعه قرار گرفته است (معادل تقویم شمسی تاریخ درج شده است):



شکل ۲۱- تفاوت در تعداد مقالات دسته‌بندی‌ها در ابتدا و انتهای بازه پژوهش جاری

همانگونه که در شکل مشخص شده است تفاوت محسوسی در تمرکز مطالعات در دسته‌بندی‌ها ایجاد نشده است. هرچند تمرکز برخی مطالعات پایه پردازش تصویر به سمت کاربردی شدن و موضوعات یادگیری ماشین و کلان داده سوق پیدا نموده است. این مساله اما درباره کلیدواژگان هر دسته در حال تغییر است. بطور مثال در دسته سوم (شبکه) در ابتدای بازه کلیدواژگان زیر نقش پررنگتری داشته اند:

- Algorithms
- Algebraic coding
- Clustering
- Location Privacy
- Wireless sensor networks

ولی در انتهای بازه جستجو تعدد کلیدواژه‌هایی مانند کلیدواژگان زیر در حوزه مربوطه در حال افزایش بوده است:

- Hybrid machine learning
- multi-task learning
- Cloud storage
- Spam filtering

این مساله باعث شده است که در این پژوهشی تنها به مقایسه روند دسته‌های موضوعی بسنده نکنیم و بنابراین مقایسه کلیدواژه‌های پرتکرار در هر دسته را با توجه به روند زمانی در فصل بعد (بخش‌های ۳-۴-۱ و ۳-۴-۲) مورد مقایسه قرار خواهیم داد.

فصل سوم

مدل‌سازی پژوهش‌های ملی در حوزه فناوری اطلاعات و مقایسه با روندهای بین‌المللی

۳-۱. مقدمه

در دهه‌های اخیر، شاهد افزایش چشم‌گیر تعداد پژوهش‌هایی هستیم که در زمینه‌های مختلف علوم صورت گرفته است، اما رشد کمی پژوهش‌ها، لزوماً با افزایش کیفیت آن‌ها همراه نبوده است و در بسیاری از موارد دستاوردهای این پژوهش‌ها منجر به رفع مشکلات جامعه نشده و به‌ندرت مشاهده شده است که نتایج و یافته‌های پژوهشی برای بهبود اوضاع سیاسی، اجتماعی، اقتصادی و فرهنگی جامعه بکار گرفته شود. در این میان شاید رشته‌های مهندسی و پژوهش‌های کاربردی این گروه وضعیت بهتری را نسبت به دیگر گروه‌های آموزشی داشته باشند (قنادینژاد , حیدری و چینی پرداز ۱۳۹۷).

فاصله یا شکاف میان پژوهش‌های دانشگاهی، به‌ویژه پایان‌نامه‌ها با نیازهای جامعه یکی از چالش‌های اساسی نظام آموزش عالی در ایران است. گسترش کمی دانشگاه‌ها و افزایش پذیرش دانشجویان تحصیلات تکمیلی باعث شده است تا ازیک‌طرف شاهد انتقادها و واکنش اندیشمندان و خبرگان آموزش و پژوهش باشیم و از طرف دیگر کارفرمایان صنایع و اقتصاد از توانمندی اندک دانش‌آموختگان در حل مسائل واقعی کشور گله‌مند باشند (فتاحی ۱۳۹۰). همچنین عرصه پژوهش شاهد دوباره‌کاری‌های بسیاری است که دلیل عمده آن،عدم توانایی استادان و دانشجویان در یافتن مسئله پژوهش در حوزه‌های مختلف موضوعی است و در بسیاری از موارد جامعه دانشگاهی بجای «مسئله‌یابی» به دنبال «مسئله‌سازی» شده است.

.بر اساس تعریف رابینسون[1]، کاستی پژوهشی، محدودیت اطلاعاتی در حوزه یا موضوعی است که مانع از نتیجه‌گیری در مورد یک سؤال پژوهشی می‌شود و نیاز پژوهشی محدودیت اطلاعاتی است که تصمیم‌گیری سیاست‌گذار را محدود می‌کند (Robinson, Saldanha and Mckoy 2011).

[1] Robinson, Karen A

به عبارتی بسیاری از پژوهش‌هایی که با عنوان تعیین یا شناسایی نیازهای پژوهشی مطرح شده‌اند و در فرایند شناسایی از تحلیل الگوهای جستجوی اطلاعات کاربران پایگاه‌های اطلاعاتی یا کاربران کتابخانه‌ها استفاده کرده‌اند ولی به نظر می‌رسد که نیازهای پژوهشی فراتر از خلأ پژوهشی است و حوزه‌های دانشی وجود دارد که ممکن است جزو نیازهای پژوهشی باشند اما بخشی از خلأ پژوهشی یا اطلاعاتی نباشند زیرا کمتر کسی به دنبال جستجوی این حوزه‌ها بوده است و در اصل تقاضایی برای این حوزه‌ها وجود نداشته و ازاین‌رو در خلأ پژوهشی جای نگرفته‌اند. همچنین بسیاری از نیازهای پژوهشی ممکن است به دلیل پیرو بودن پژوهشی رخ دهد و باعث ایجاد یک فصله زمانی چند ساله میان پژوهشهای داغ دنیا و در سطح ملی گردد.

سیاست‌گذاران پژوهش پس از اطلاع از خلأ پژوهشی و نیز پژوهش‌هایی که تقاضای قابل‌توجهی در جامعه پژوهشی برای آن‌ها وجود ندارد و با بهره‌گیری از دانش و اطلاعات دیگری که در اختیار دارند و نیز در نظر گرفتن سیاست‌های کلان پژوهشی، نیازهای پژوهشی را شناسایی و تعیین می‌کنند. در این مرحله سیاست‌گذار ممکن است سیاست‌ها و برنامه‌هایی را اعمال کند که تقاضای پژوهشی خاصی در جامعه پژوهشی ایجاد شود و این نیاز به پژوهشگران منتقل گردد. درنهایت سیاست‌گذاران و مدیران بر اساس معیارهای مختلف و با توجه به منابع موجود اولویت‌های پژوهشی را در حوزه‌های مختلف تعیین می‌کنند.

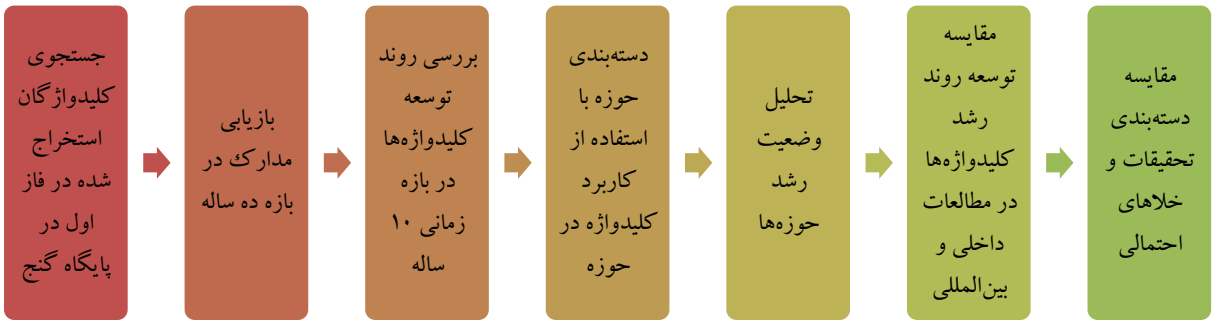
این فصل به سیاست‌گذاران کمک خواهد نمود که در زمینه فناوری اطلاعات بتوانند فاصله زمانی میان ایجاد یک پژوهش بین‌المللی و ملی را رصد نموده و برای رفع آن سرمایه‌گذاری لازم را انجام دهند. همچنین عدم پرداختن یک زمینه علمی به موضوعی خاص را مورد مطالعه قرار می‌دهیم و از خبرگان خواهیم خواست نظر خود را در مورد چرایی عدم پرداختن پژوهش‌های ملی به موضوعی خاص بین نمایند.

در این فصل در ابتدا کلیدواژگان استخراج شده از فصل قبل مجددا در پایگاه داده گنج مورد جستجو قرار می‌گیرد. بدین ترتیب مختلف بازیابی و استخراج می‌شوند. بدون احتساب دسته بندی فصل قبل، مدارک مجددا از طریق روش مدلسازی موضوعی دسته بندی شده و جایگاه هر دسته بصورت عمومی با دسته بندی فصل قبل مقایسه می‌شود. سپس هر مدرک کشف شده نیز با تابع برازش در دسته بندی پایگاه بین‌المللی جای می‌گیرد تا مشخص شود کدام مدرک در چه دسته ملی و چه دسته بین‌المللی قرار گرفته است. با بهترین تخصیص می‌توان دسته‌بندی ایجاد شده را بین دو پایگاه را معادلسازی کرده و میزان رشد و نمو مطالعات در طول زمان را استخراج کرد.

این مطالعه همچنین مقایسه رشد و نمو مطالعات را در طی زمان بین دو پایگاه را از طریق مقایسه اختلاف میزان تمرکز هر دسته (بخش ۳-۵-۱) و هر کلیدواژه (بخش ۳-۴-۲) را نمایان خواهد نمود.

۳-۲. روش پژوهش

روش پژوهش جاری در فاز دوم این طرح پژوهشی (فصل جاری) به مراحل زیر انجام خواهد پذیرفت:



شکل ۲۲- مراحل پژوهش جاری در فاز دوم

در جدول زیر مراحل مختلف شکل فوق تشریح می‌شوند:

جدول ١٤- شرح روش پژوهش

مرحله	روش
استحراج و بررسی مطالعات	جستجوی کلیدواژگان استخراج شده در فاز اول در پایگاه گنج
	بازیابی مدارک در بازه ده ساله
	بررسی روند توسعه کلیدواژه‌ها در بازه زمانی ۱۰ ساله
دسته‌بندی حوزه‌ها و تحلیل روند	دسته‌بندی حوزه با استفاده از کاربرد کلیدواژه در حوزه
	تحلیل وضعیت رشد حوزه‌ها
مقایسه با مطالعات بین‌المللی	مقایسه توسعه روند رشد کلیدواژه‌ها در مطالعات داخلی و بین‌المللی
	مقایسه دسته‌بندی تحقیقات و خلاهای احتمالی

مدل پژوهش جاری در فاز دوم در شکل زیر نمایش داده شده است. لازم به ذکر است که مراحل داده در بخش اول پژوهش برای داده‌های پایگاه‌داده‌های بین‌المللی انجام پذیرفت (شکل ۶) و در این فاز از پژوهش مجددا برای پایگاه داده گنج تکرار شد. اما تحلیل‌های این فصل شامل تحلیل‌های مقایسه‌ای بین روندهای ملی و بین‌المللی نیز است.

داده

با جستجو در پایگاه گنج با استفاده از کلیدواژه‌های منتخب در فاز ۱

پالایش داده

پاکسازی داده، حذف تکرار و

انتقال داده

تبدیل داده به فرمت قابل تبدیل برای پردازش در نرم‌افزار، تهیه

داده‌کاوی (مدل‌سازی موضوعی)

- کشف ارزش هر واژه (تعداد تکرار)
- دسته‌بندی به ازای واژه‌های پر تکرار و خویشاوندی
- ارزیابی کیفیت دسته‌بندی
- تهیه دسته‌های مناسب و احتمال حضور هر کلیدواژه در دسته

تحلیل

- بررسی روند کلیدواژه‌ها در زمان
- دسته‌بندی موضوعی فناوری اطلاعات
- بررسی روند رشد و افول کلیدواژگان
- مقایسه روند در دو پایگاه WOS و گنج

شکل ۲۳- مدل پژوهش جاری در فاز دوم

۳-۳. انتخاب کلیدواژگان قابل جستجو

همانگونه که در فصل قبل ذکر شد، سبدی از واژگان تایید شده توسط خبرگان با عنوان کلیدواژگان معتبر فناوری اطلاعات انتخاب شدند. انتخاب یا عدم انتخاب کلیدواژه‌های پیشنهادی یافت شده در جدول ۶ ذکر شد. به‌دلیل نیاز به یکپارچگی جستجو و قابل مقایسه بودن خروجی بدست آمده از پایگاه گنج و پایگاه‌های بین‌المللی، می‌بایست از همین فهرست در جستجوی پایگاه گنج استفاده نماییم. لازم به ذکر است این فهرست تایید شده شامل ۸۰ کلیدواژه است که در جدول زیر آمده است. نحوه جستجو به نحوی شکل گرفته است که اگر هر تحقیقی فقط شامل یکی از کلیدواژه‌های زیر باشد نیز جزو مجموعه جواب‌های این مطالعه قرار می‌گیرد. بطور مثال ممکن است تحقیقی در زمینه مهندسی صنایع بوده و فقط از یکی از کلیدواژه‌های زیر مانند Social Media استفاده کرده باشد. این تحقیق نیز جزو مجموعه جواب‌های فرض شده در نظر گرفته می‌شود. هرچند به دلیل فرکانس کمتر احتمالی دیگر کلیدواژه‌های این مطالعه، ارزش موضوعی زیادی پیدا نخواهد کرد. این رویکرد را به‌این دلیل انتخاب نموده‌ایم که از خطای جاافتادگی زمینه‌ی علمی خاصی توسط خبرگان چشم‌پوشی نشود. قابل ذکر است با وجود انتخاب این رویکرد، کلیدواژه قابل بازیابی به‌صورت پرفرکانس یا ترند که در فهرست زیر موجود نباشد یافت نشد.

جدول ۱۵- کلیدواژه منتخب برای تشکیل سبد واژگان جستجو در پایگاه گنج

ردیف	واژه	ردیف	واژه
۱	Deep Learning	۴۵	Learning Systems
۲	Machine Learning	۴۶	Multi-Task Learning
۳	Social Media	۴۷	Saliency Detection
۴	Data Envelopment Analysis	۴۸	Action Recognition
۵	Security	۴۹	Computer Vision
۶	Feature Selection	۵۰	Ensemble Learning
۷	Privacy	۵۱	Internet Of Things
۸	Network Coding	۵۲	List Decoding
۹	Feature Extraction	۵۳	Semantics
۱۰	Data Mining	۵۴	Community Detection
۱۱	Image Segmentation	۵۵	Cryptography

ردیف	واژه
۱۲	Online Learning
۱۳	Adaptive Control
۱۴	Physical Layer Security
۱۵	Supervised Learning
۱۶	Person Re-Identification
۱۷	Authentication
۱۸	Big Data
۱۹	Information Theory
۲۰	Object Detection
۲۱	Cloud Computing
۲۲	Recommender Systems
۲۳	Social Network
۲۴	Source Coding
۲۵	Crowdsourcing
۲۶	Error Exponent
۲۷	Face Recognition
۲۸	Side Information
۲۹	Interference Channel
۳۰	Query Processing
۳۱	Decision Analysis
۳۲	Visualization
۳۳	Biometrics
۳۴	Image Classification
۳۵	Image Retrieval
۳۶	Broadcast Channel
۳۷	Channel Coding
۳۸	Kernel Methods

ردیف	واژه
۵۶	Cyberbullying
۵۷	Information Theoretic Security
۵۸	Network Security
۵۹	Private Information Retrieval
۶۰	Wireless Networks
۶۱	۵g Mobile Communication
۶۲	Hashing
۶۳	Index Coding
۶۴	Information Entropy
۶۵	Parallel Computing
۶۶	Sparse Coding
۶۷	Www
۶۸	Cloud Storage
۶۹	Data Privacy
۷۰	Information Security
۷۱	Iterative Decoding
۷۲	Mds Codes
۷۳	Process Mining
۷۴	Sentiment Analysis
۷۵	Spectral Clustering
۷۶	Superposition Coding
۷۷	System Identification
۷۸	Wireless Sensor Networks
۷۹	Data Models
۸۰	Error Correction Codes
۸۱	Homomorphic Encryption
۸۲	Human-Computer Interaction

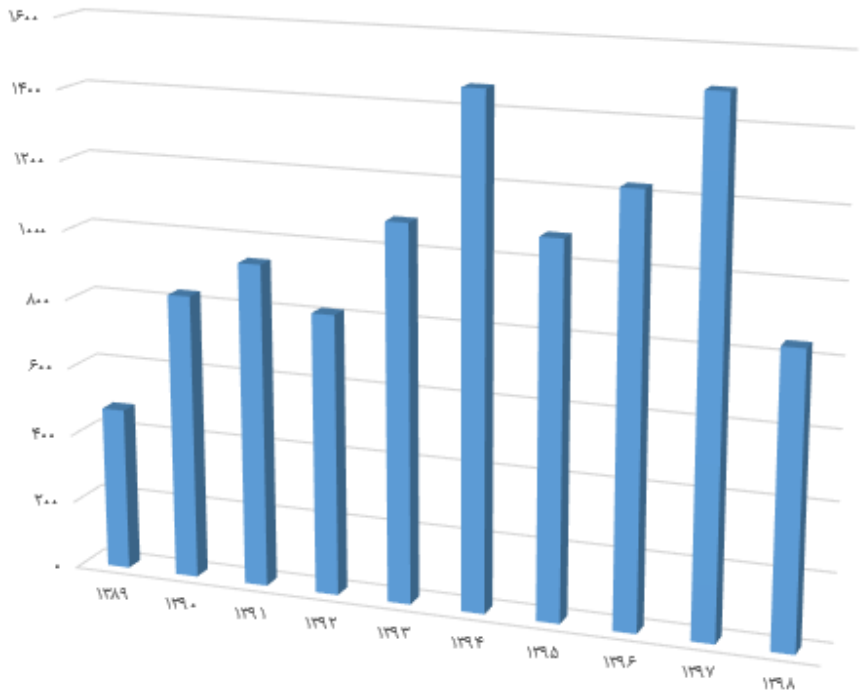
ردیف	واژه
۳۹	Pattern Recognition
۴۰	Image Reconstruction
۴۱	User Experience
۴۲	Artificial Intelligence
۴۳	Computational Modeling
۴۴	Image Restoration

ردیف	واژه
۸۳	Image Recognition
۸۴	Link Prediction
۸۵	Multi-View Learning
۸۶	Random Coding
۸۷	Text Mining
۸۸	Topic Modeling

۳-٤. بررسی روند توسعه کلیدواژه‌ها

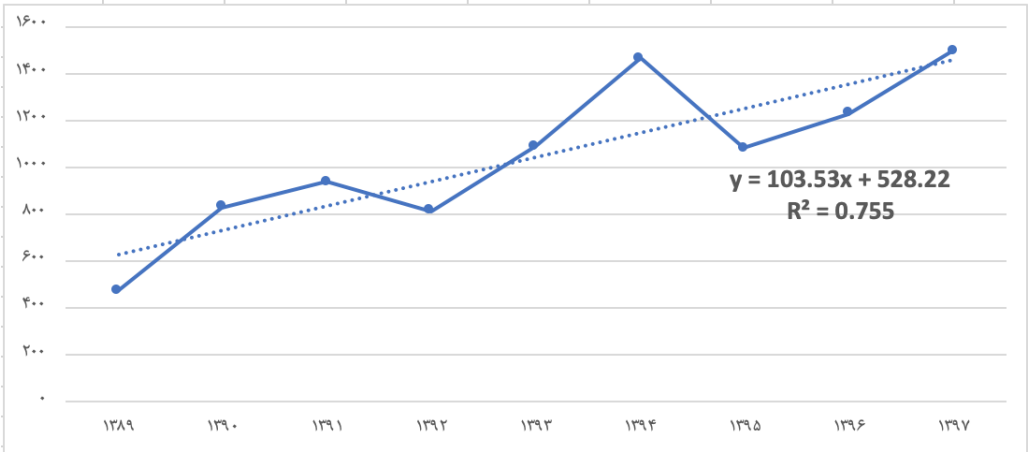
پایگاه اطلاعات علمی ایران (گنج)، دستاورد کاربرد فناوری اطلاعات در خدمات مدیریت اطلاعات علم و فناوری در پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) است. پیشینه این بخش از خدمات ایرانداک به دهه چهل بازمی‌گردد که کار گردآوری، سازمان‌دهی، و اشاعه اسناد علمی و فناورانه در مرکز اسناد علمی آغاز شد. هر چند اکنون بیشترین داده‌های روزآمد گنج از پایان‌نامه‌ها و رساله‌هاست*، ولی قلمرو مدیریت اطلاعات علم و فناوری بسیار گسترده بوده و طرح‌های پژوهشی، پایان‌نامه‌ها، رساله‌ها، پیشنهاده‌ها، پروانه‌های اختراع، گزارش‌های دولتی، همایش‌های علمی، نشریه‌های علمی، مقاله‌های علمی، ناشران علمی، و روزنامه‌ها و همچنین کتاب‌شناسی‌ها، واژه‌نامه‌ها، و فهرست‌های مستند نام‌ها را در بر داشته است. کار پایگاه اطلاعات علمی ایران، نخستین بار در سال ۱۳۷۰ در رایانه‌های «مین‌فریم» و محیط نرم‌افزار «سی‌دی‌اس‌آی‌اس‌آی‌اس» شروع شد. در سال ۱۳۷۳ این پایگاه با فناوری «دایال‌آپ» در شبکه‌ای با نام «صبا» از دور در دسترس کاربران قرار گرفت. سال ۱۳۷۸ بود که اینترنت به ایرانداک وارد شد و دسترسی به گنج نیز در وب فراهم شد. از آن پس با پیشرفت فناوری اطلاعات، این پایگاه نیز قابلیت‌های تازه‌ای یافته است و در گذر زمان، نسخه‌های گوناگونی از آن با فناوری‌های روز به‌کار رفته‌اند. این پایگاه هم‌اکنون مرجع بسیاری از پژوهشگران ایران و جهان است و روزانه بیش از ده هزار کاربر، ده‌ها هزار جست‌وجو در آن انجام می‌دهند. این سامانه بنیان سامانه‌های دیگری در ایرانداک همچون سامانه همانندجو، سامانه پیشینه پژوهش، و برخی از داشبوردهای رصدخانه پژوهش و فناوری نیز هست. پایگاه ایرانداک که به نام «گنج» در اختیار کاربران قرار دارد و مهم‌ترین کاربران آن را پژوهشگران و دانشجویان تحصیلات تکمیلی تشکیل می‌دهند منابع گوناگون علمی از رشته‌های علمی مختلف را گردآوری، سازمان‌دهی و اشاعه می‌کند.

تمرکز اصلی این پایگاه بر سازمان‌دهی پایان‌نامه‌ها و رساله‌های تحصیلات تکمیلی است و بزرگ‌ترین و تنها مرجع رسمی است که وظیفه ثبت، سازمان‌دهی و اشاعه پایان‌نامه‌ها و رساله‌ها را به عهده دارد.
این پایگاه در حال حاضر بیش از ۱ میلیون رکورد علمی را شامل می‌شود که در حال حاضر بیش از ۵۰۰ هزار رکورد آن مربوط به پایان‌نامه‌ها و رساله‌های تحصیلات تکمیلی داخل کشور و بیش از ۱۵ هزار رکورد آن مربوط به پایان‌نامه‌ها و رساله‌های ایرانیان خارج از کشور است. قدیمی‌ترین پایان‌نامه فارسی موجود در پایگاه ایرانداک در رشته دامپزشکی دانشگاه تهران و مربوط به سال ۱۳۲۳ و قدیمی‌ترین پایان‌نامه لاتین مربوط به سال ۱۹۲۰ است. به‌طور متوسط روزانه بیش از ۱۳۰ هزار جستجو در این پایگاه انجام می‌شود.
ازآنجاکه دانشجویان تحصیلات تکمیلی کشور موظف به ثبت پایان‌نامه و رساله خود در این پایگاه هستند، در حال حاضر نزدیک به ۷۰ درصد از کل پایان‌نامه و رساله‌های کشور در این سامانه موجود است. از طرفی بخش عمده مقالات تولیدشده در دانشگاه‌ها و مراکز آموزشی و پژوهشی از پایان‌نامه و رساله مستخرج می‌شود ازاین‌رو این پایگاه تا حد قابل قبولی می‌تواند نماینده پژوهش‌های دانشگاهی باشد. از طرف دیگر دانشجویان برای اطمینان از تکراری نبودن فعالیت پژوهشی خود و نیز استفاده از منابع علمی موجود در این پایگاه، جستجوهای فراوانی را در این پایگاه انجام می‌دهند که موجب شده است تا روزانه نزدیک به ۱۳۰ هزار جستجو در این پایگاه ثبت شود، به همین دلیل این پایگاه می‌تواند نشانگر خوبی برای تقاضاهای پژوهشی باشد که از طرف پژوهشگران بیان می‌شود.
وجود اعضای هیئت‌علمی متخصص در حوزه‌های مدیریت دانش، کتابداری و اطلاع‌رسانی و فناوری اطلاعات و نیز تعامل سازنده مدیران این پژوهشگاه نیز از نقاط مثبت و قابل‌توجهی است که باعث شده است این پایگاه به‌عنوان مورد مطالعه انتخاب شود.
کلیدواژه‌های یافت شده در دسته‌بندی فناوری اطلاعات، در پایگاه گنج منجر به بازیابی بیش از ۱۰ هزار مدرک شدند که از بعد تناسب با تعداد واژگان انتخابی از پایگاه بین‌المللی تناسب دارند. لازم به ذکر استمیزان مدارک انتخابی به میزان ۱۰۲۵۵ مدرک (با توجه به پیشنهادیه این طرح و تایید خبرگان) و در طول زمان ۱۰ ساله به ترتیب زیر قرار گرفته است:



شکل ۲۴- میزان استخراج منبع بر حسب سال

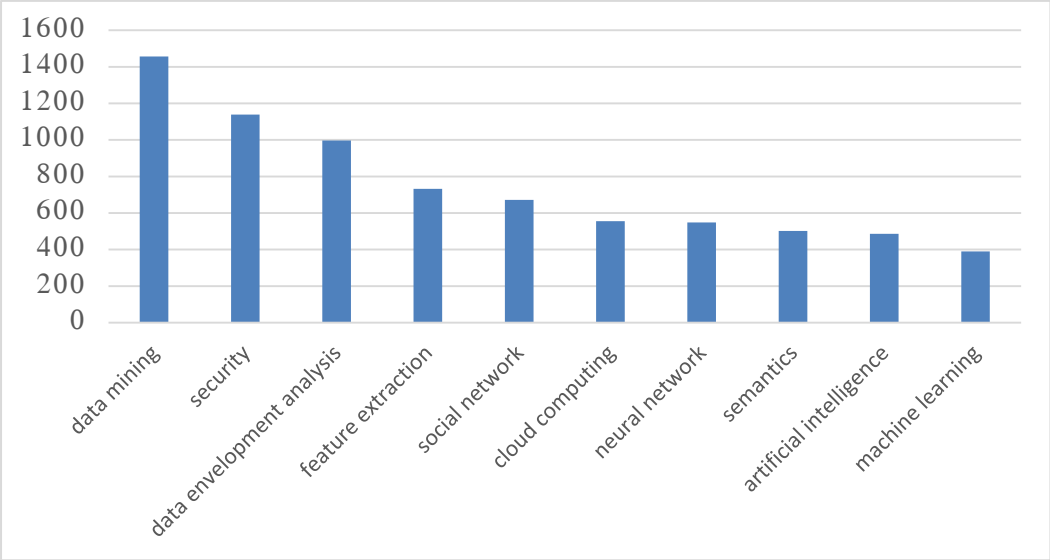
کمتر بودن منابع در سال ۹۸ به‌دلیل این مساله است که بسیاری از دانشجویان ثبت رساله و پایان‌نامه خود را اندکی پس از دفاع و بعد از انجام پابلیکیشن‌های لازم انجام می‌دهند. با حذف سال ۹۸ از پیش‌بینی، همانگونه که در نمودار زیر قابل مشاده است، با اطمینان نسبتا خوبی روند مطالعات در این زمینه صعودی است.



شکل ۲۵- رشد داده‌های استخراج شده از پایگاه گنج در سال

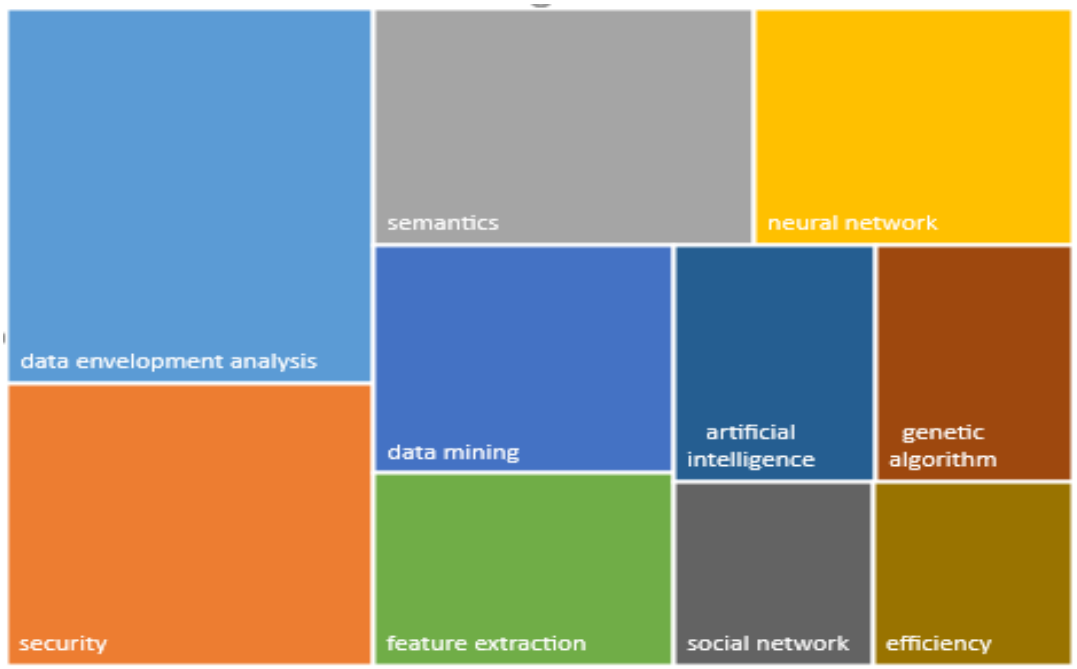
۳-٤-۱. بررسی روند کاربرد کلیدواژگان و تمرکز بر حوزه‌های مختلف

با توجه به فرکانس کلیدواژه‌های مقالات در بیش از ۱۰هزار رکورد مرتبط با جستجو در پایگاه گنج، برای ترسیم ۱۵ کلیدواژه برتر به به نمودار زیر می‌رسیم:



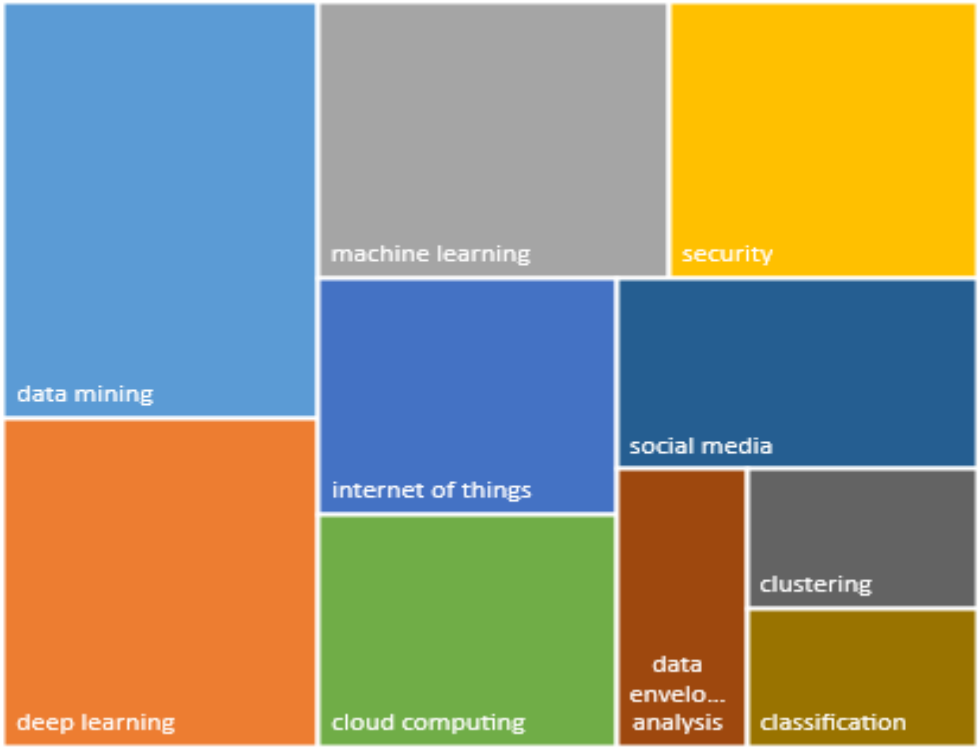
شکل ۲۶- کلیدواژه‌های پرکاربرد حوزه فناوری اطلاعات در ده سال

با توجه به نمودار فوق می‌توان مشاهده نمود که مباحث داده‌کاوی و امنیت جایگاه ویژه‌ای را در میان کلیدواژگان پرکاربرد در زمینه فناوری اطلاعات در کشور دارند. این دسته‌بندی با توجه به پیشنهاده طرح در بازه ۱۰ ساله مورد بررسی قرار گرفته است. حال اینکه همانگونه که برای مطالعات بین‌المللی ذکر شد، کلیدواژه‌های پرکاربرد در زمینه فناوری اطلاعات با توجه به پویایی زیاد این زمینه علمی ممکن است در طول زمان دچار تغییرات گردد. بدین منظور اگر بخواهیم نگاهی به تعریف کلیدواژه‌ها در ابتدا و انتهای بازه مورد بررسی یعنی سال ۱۳۸۹ و ۱۳۹۸ بیاندازیم، به نمودارهای زیر می‌رسیم:



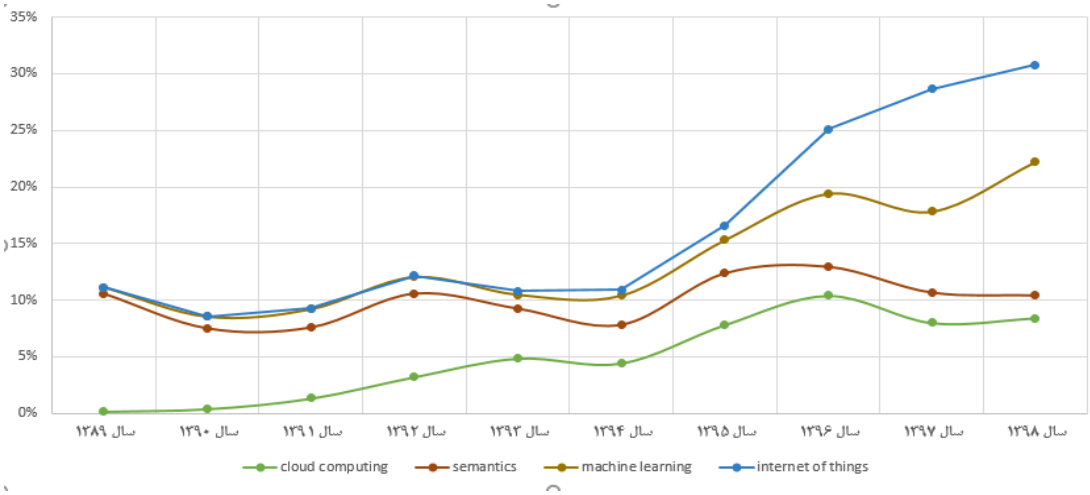
شکل ۲۷- کاربرد و تکرار کلیدواژه‌های پایگاه گنج در سال ۱۳۸۹

همانگونه که در شکل فوق دیده می‌شود، ۳ حوزه تحلیل پوششی داده‌ها، امنیت و تحلیل معنایی، بزرگترین حوزه‌های تمرکز کلیدواژه‌های فناوری اطلاعات در سال ۱۳۸۹ هستند. حال اگر بخواهیم همین گزارش را در سال ۱۳۹۸ به تحلیل بکشیم داریم:



شکل ۲۸ -کاربرد و تکرار کلیدواژه‌ها در سال ۲۰۱۹

قابل مشاهده است که حوزه‌های داده‌کاوی، یادگیری عمیق و یادگیری ماشین جایگاه ویژه‌ای در این سال پیدا کرده‌اند. همچنین همانگونه که در شکل فوق ملاحظه می‌شود حوزه‌های نوظهوری همانند اینترنت اشیا و محاسبات ابری و داده‌های عظیم در کنار حوزه‌های کاری مانند یادگیری ماشین و یادگیری عمیق تعریف حوزه‌های فناوری اطلاعات را در بازه زمانی جدید تشکیل داده‌اند که این ۴ کلیدواژه در ۱۵ کلیدواژه برتر ۱۰ سال قبل (۱۳۸۹) جود نداشته‌اند. با توجه به دو شکل فوق رشد ۴ حوزه‌ی جدید اینترنت اشیا، محاسبات ابری و داده‌های بزرگ و یادگیری عمیق را به شکل (در مقیاس ۱۰۰) زیر خواهیم داشت:



شکل ۲۹- روند رشد ٤ حوزه نوظهور کشف شده در پژوهش

همانگونه که مشاهده می‌شود رشد این زمینه علمی بویژه در حوزه‌های اینترنت اشیا و یادگیری عمیق بسیار مشهود است و زمینه‌های علمی مانند اینترنت اشیا همچنان در حال ادامه است. در جدول زیر می‌توان رشد و افول هر کلیدواژه برتر فناوری اطلاعات را در ابتدا و انتهای بازه ۱۰ ساله مشاهده نمود:

جدول ١٦- مقایسه کلیدواژه‌های پرکاربرد در دو بازه زمانی

کلیدواژه	درصد در بین کلیدواژگان پرکاربرد در سال ۲۰۱۱	درصد در بین کلیدواژگان پرکاربرد در سال ۲۰۱۹	رشد
Data Mining	8%	16%	8%
Security	12%	10%	-2%
Data Envelopment Analysis	16%	4%	-12%
Feature Extraction	7%	3%	-3%
Social Network	4%	4%	-1%
Cloud Computing	0%	8%	8%
Neural Network	9%	2%	-6%
Semantics	10%	2%	-8%
Artificial Intelligence	6%	2%	-3%
Machine Learning	1%	12%	11%
Image Processing	4%	1%	-3%
Privacy	4%	4%	0%
Clustering	1%	4%	2%
Genetic Algorithm	5%	3%	-3%
Efficiency	4%	1%	-3%
Social Media	0%	8%	8%
Support Vector Machine	3%	1%	-2%
Cryptography	4%	1%	-3%
Internet Of Things	0%	9%	9%
Classification	1%	4%	2%

در این میان یادگیری ماشین و اینترنت اشیا با بیشترین رشد و تحلیل پوششی داده‌ها و تحلیل معنایی با بیشترین نزول در میان برترین کلیدواژگان مجموع اخیر قرار گرفته‌اند. همچنین می‌توان نتیجه گرفت ۵ حوزه زیر نقش و جایگاه بالاتری در آینده فناوری اطلاعات کشور خواهند داشت:

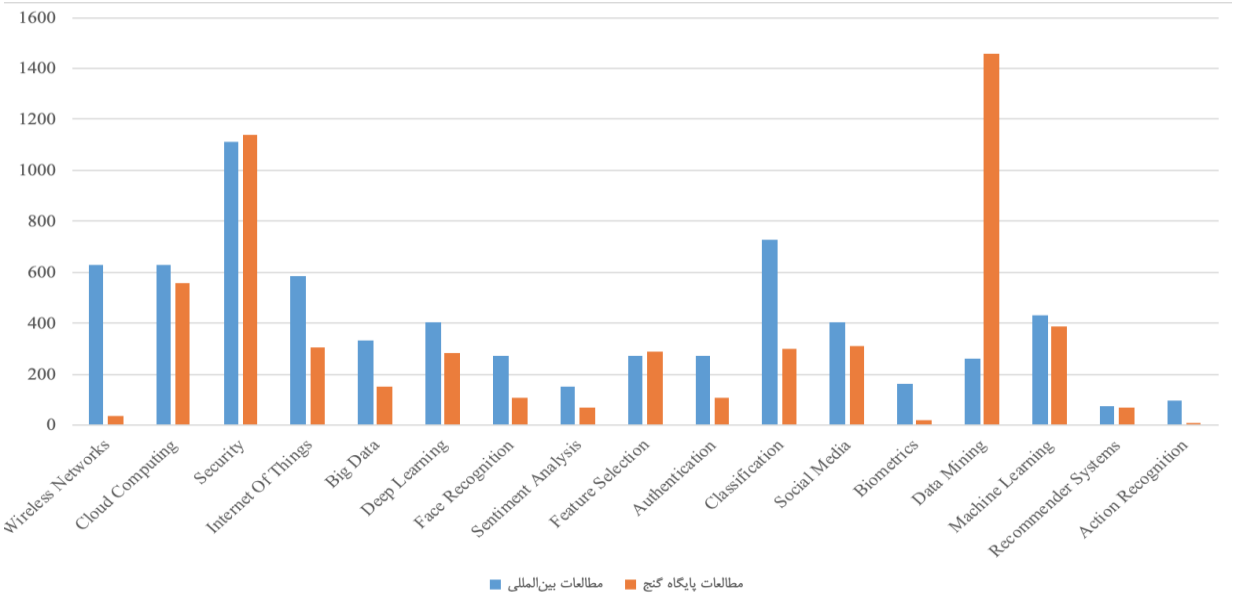
- Internet of Things
- Cloud Computing
- Social Media
- Data Mining
- Machine Learning

اگر بخواهیم نگاهی به کلیدواژگان توسعه یافته در سطح بین‌المللی داشته باشیم (با توجه به فصل قبل) عدم پررنگ بودن نقش Big Data به اندازه دیگر کلیدواژگان در سطح ملی نسبت به سطح بین‌المللی است. در بخش بعدی از خبرگان خواهیم خواست که دلایل و چالش‌های این فاصله را تشریح نمایند.

۳-٤-۲. مقایسه روند کلیدواژگان در مطالعات ملی و بین‌المللی

برای مقایسه روند ظهور و رشد کلیدواژگان در دو نمونه داخلی و بین‌المللی و با توجه به ثبت داده سال برای داده‌ها به‌صورت میلادی و شمسی، بازه مشترک ۹ ماهه اول سال را به‌عنوان معادل قرار می‌دهیم. این معادل‌سازی ممکن است خطایی را ایجاد نماید ولی به‌دلیل عدم ثبت سال میلادی بر داده‌های پایگاه گنج، این تصمیم لازم به نظر می‌رسد. بدین منظور داده‌های سال ۲۰۱۱ میلادی را با سال ۱۳۹۰ و به همین ترتیب داده‌های سال ۲۰۱۹ را با سال ۱۳۹۸ مقایسه خواهیم نمود.

به علت نزدیکی میزان نمونه تحقیقات مورد مطالعه (تمامی پایگاه گنج و به همان میزان نمونه متناسب سال‌های مختلف در پایگاه‌های بین‌المللی) مقایسات این بخش با اعداد مطلق صورت می‌پذیرد. بر این اساس در نگاهی کلی تعداد کلیدواژه‌های مستخرج از هر دوپایگاه را در مدت زمان یکسان، داریم:

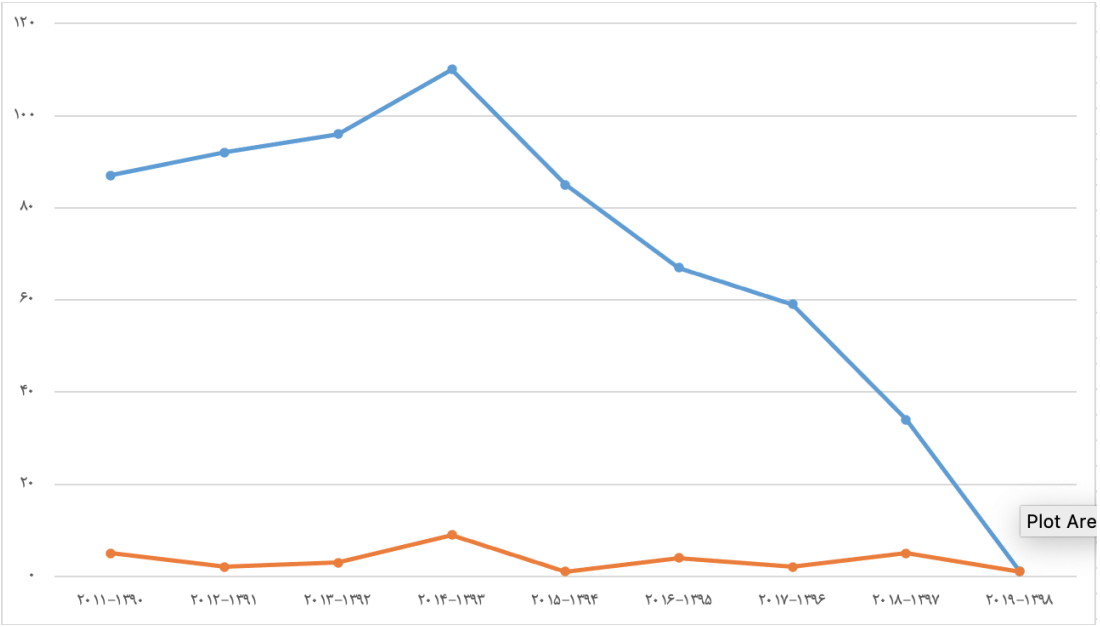


شکل ۳۰- مقایسه تعداد کلیدواژه‌های پرکاربرد

با توجه به نمودار فوق می‌توان دریافت در سه کلیدواژه Wireless Network(s)، Biometrics و Action Recognition اختلاف زیادی میان تعداد (یا نسبت مطالعات) داخلی و بین‌المللی مشاهده می‌شود.

با توجه به نمونه ارایه شده به خبرگان، تمامی این سه زمینه تحقیقاتی در فناوری اطلاعات زمینه‌هایی بودند که برای توسعه پژوهش در آنها، نیازمندی‌های سخت‌افزاری ویژه مورد نیاز بود. همچنین خبرگان اشاره داشتند که نمونه تحقیقات ملی انجام شده در این زمینه (بر خلاف سطح بین‌المللی) بیشتر روی داده‌های نمونه و فرضی و نه داده‌های تجربی و عملی فعالیت نموده‌اند. همچنین دو کلیدواژه داده‌کاوی و امنیت، با توجه به نمونه مورد بررسی توسط خبرگان، به‌علت کاربرد این دو کلیدواژه در زمینه‌های علمی دیگر در فهرست بالا قرار گرفته و آمار مذکور از دید ایشان مختص به فناوری اطلاعات نیست.

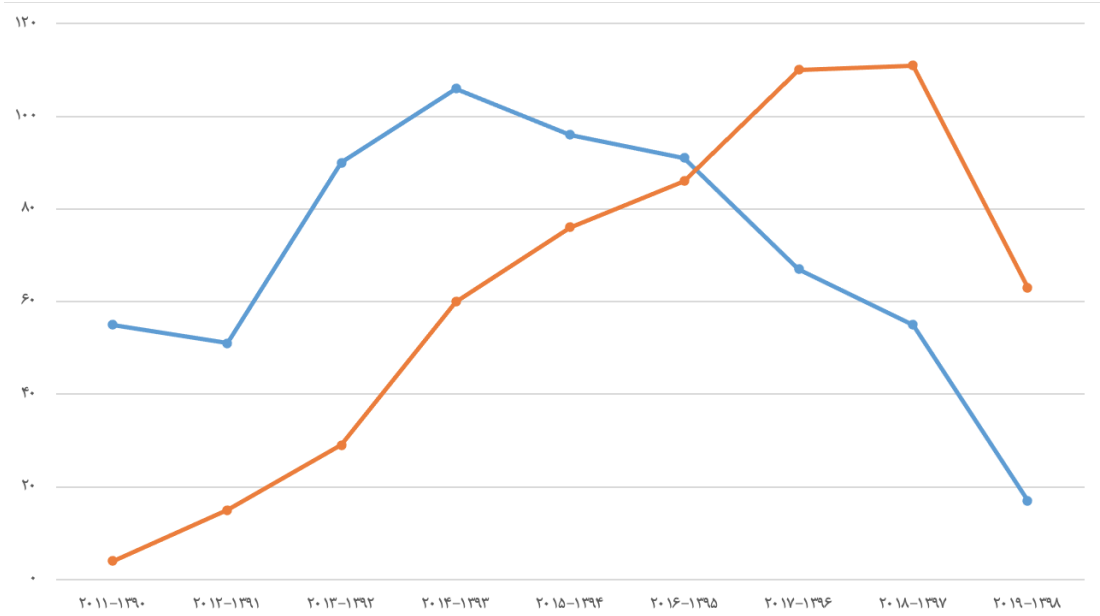
بر این اساس، مقایسه برای هر کلیدواژه برتر یا درحال رشد (در هر دو نمونه) انجام پذیرفت. در شکل زیر روند مطالعات در کلیدواژه Wireless Network (s) مورد بررسی قرار می‌گیرد:



شکل ۳۱- مقایسه روند مطالعات در کلیدواژه Wireless Network (s)

همانگونه که پیشتر ذکر شد، عدم بهره‌مندی از سخت‌افزارهای مورد نیاز در مطالعات یکی از دلایل اختلاف میزان تمرکز مطالعات در سطح ملی و بین‌المللی در زمینه فوق یافت شد.

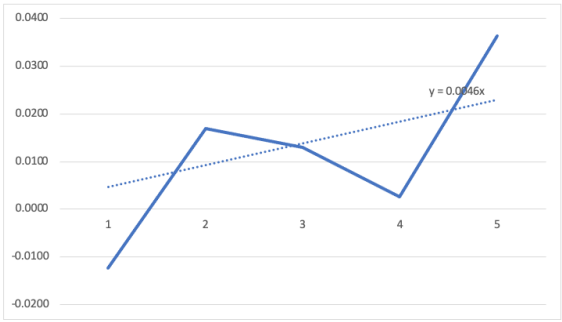
در زمینه مطالعات رایانش ابری نمودار زیر قابل مشاهده است:



شکل ۳۲- مقایسه روند مطالعات در کلیدواژه Cloud Computing

همانطور که از نمودار بر می‌آید مطالعات بین‌المللی در سال ۲۰۱۴ بیشترین رشد را داشته که این پیک در مطالعات ملی به سال ۱۳۹۷ باز می‌گردد. اختلاف ۴ ساله میان ترند ملی و بین‌المللی در این زمینه ناامیدکننده است.

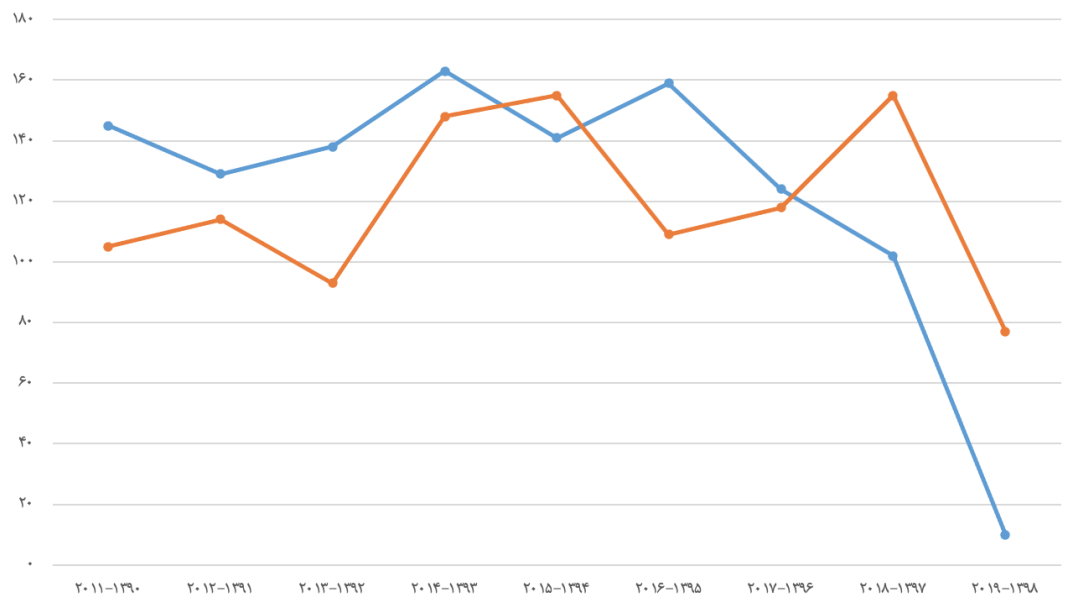
سال	پایگاه گنج	٪	WOS	٪	فرض برابری با اختلاف ٤ سال
۱۳۹۰	۴	۰٪	۵۵	۵٪	-۰.۰۱۲۳
۱۳۹۱	۱۵	۱٪	۵۱	۹٪	۰.۰۱۶۹
۱۳۹۲	۲۹	۲٪	۹۰	۱۰٪	۰.۰۱۳۰
۱۳۹۳	۶۰	۵٪	۱۰۶	۹٪	۰.۰۰۲۶
۱۳۹۴	۷۶	۶٪	۹۶	۹٪	۰.۰۳۶۴
۱۳۹۵	۸۶	۷٪	۹۱	۶٪	
۱۳۹۶	۱۱۰	۹٪	۶۷	۵٪	
۱۳۹۷	۱۱۱	۹٪	۵۵	۲٪	
۱۳۹۸	۶۳	۵٪	۱۷		



t	m	n	S	n	معیار پذیرش ۹۰ درصد	Result
۱.۴۰۶۶۱۲۵۹۱	۰.۰۱۱۳	۰	۰.۰۱۷۹۸۶۳۸۶	۵	۱.۹۴۳	Approved

شکل ۳۳- مقایسه تست شکاف زمانی بین مطالعات در کلیدواژه Cloud Computing

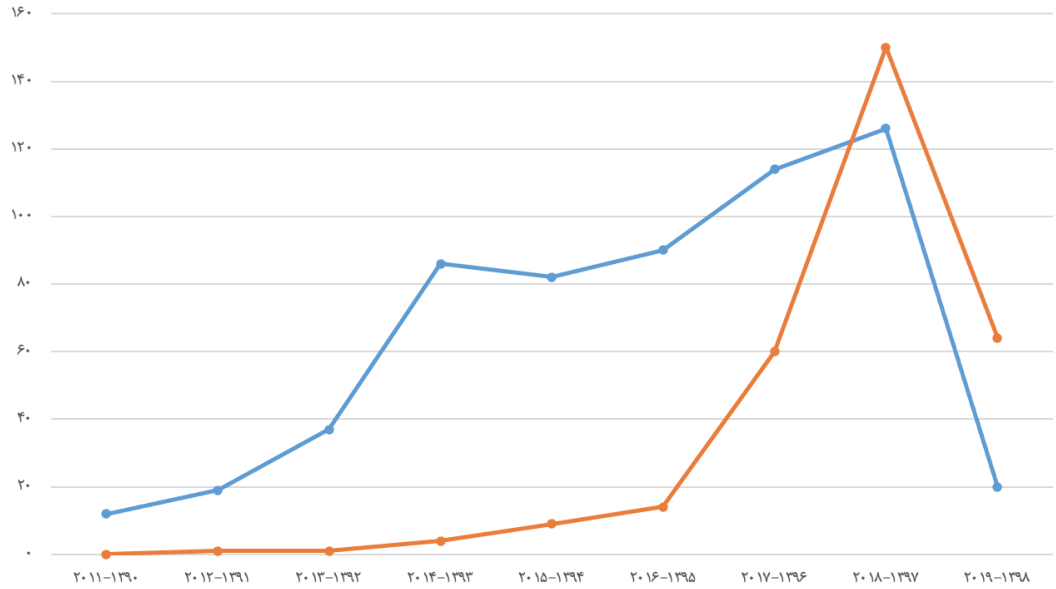
روند پرداختن به موضوع امنیت تقریبا در بازه زمانی مورد بررسی یکسان و در سطح ملی و بین‌المللی نیز متفاوت نیست. البته در بخش بعدی مشخص خواهد شد که بخشی ازین کلیدواژه در حوزه فناوری اطلاعات نبوده است (در پایگاه گنج).



شکل ۳۴- مقایسه روند مطالعات در کلیدواژه Security

بنابراین افزایش این تعداد در سال‌های اخیر در پژوهش‌های داخلی به‌دلیل استفاده ازین کلیدواژه در علوم دیگر است و خطای استخراج داده در این طرح پژوهشی است.

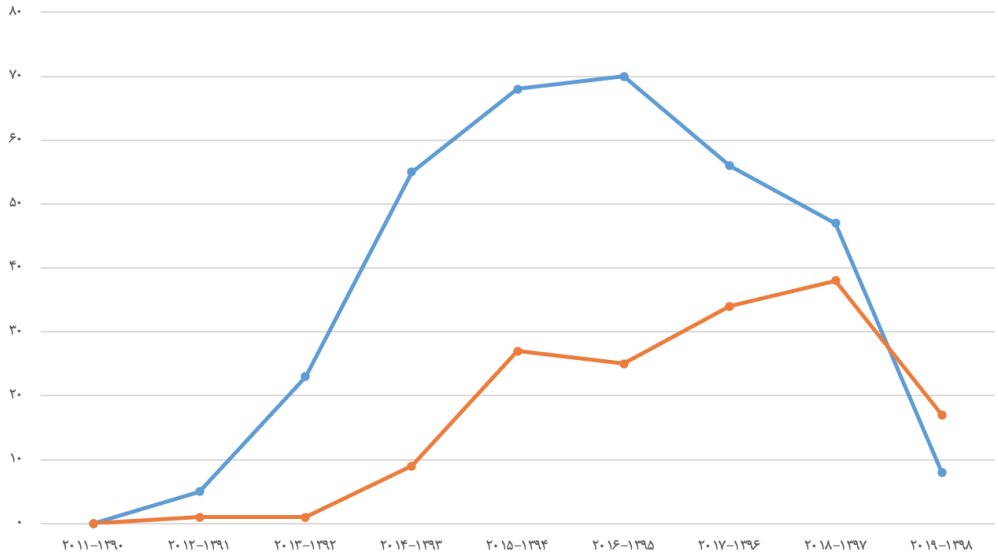
شروع و اوجگیری مطالعات در زمینه اینترنت اشیا در سطح بین‌المللی در سال ۲۰۱۱ و ۲۰۱۴ به وقوع پیوسته است و این شروع و اوج در مطالعات ملی به ۱۳۹۳ و ۱۳۹۶ باز می‌گردد.



شکل ۳۵- مقایسه روند مطالعات در کلیدواژه Internet of Things (IOT)

همانگونه که مشاهدات فوق نشان می‌دهد بازه اختلاف ۳ تا ۴ سال میان شروع و اوجگیری مطالعات در این زمینه علمی میان سطح داخلی و بین‌المللی مشهود است. البته تست آماری این فرضیه با اختلاف کمی پذیرفته نشد (به دلیل شیب متفاوت رشد).

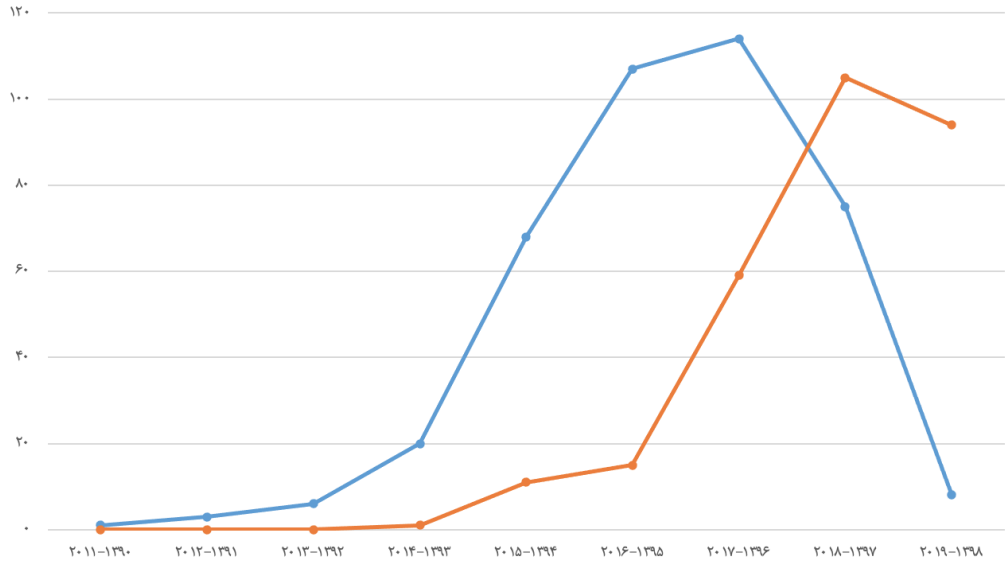
کلیدواژه داده‌های عظیم از هر دو عارضه رنج می‌برد. فاصله زمانی میان داغ شدن موضوع علمی و همچنین پرداختن کمتر به این موضوع در شکل زیر قابل مشاهده است.



شکل ۳۶- مقایسه روند مطالعات در کلیدواژه Big Data

از دیدگاه خبرگان این کاهش به‌علت نیازمندی این حوزه‌ی علمی به داده‌های واقعی و انحصاری و همچنین رایانه‌های پیشرفته برای پیش‌بینی و مدل‌سازی پرسرعت باز می‌گردد.

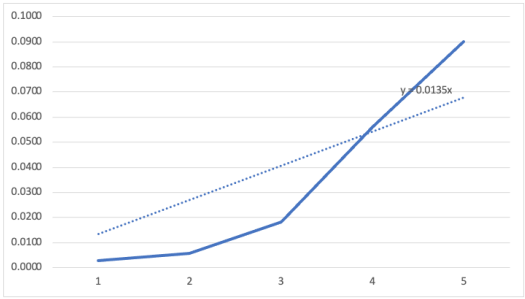
در زمینه کلیدواژگان Deep Learning، Face Recognition، Sentiment Analysis ، Feature Selection، Authentication و Classification نیز دو الگوی فوق قابل بازیابی است. گذر مطالعات بین‌المللی در این مباحث از سمت Classification در ده سال پیش شروع شده و به سمت Deep Learning در سال‌های اخیر رشد داشته است.



شکل ۳۷- مقایسه روند مطالعات در کلیدواژه‌ی Deep Learning

با وجود پیش‌بینی نسبی زمینه فناورانه Deep Learning می‌توان مشاهده نمود که با الگوی کشف شده، در سال جاری یا سال بعد بیشترین رشد مطالعات داخلی کشور در این زمینه خواهد بود.

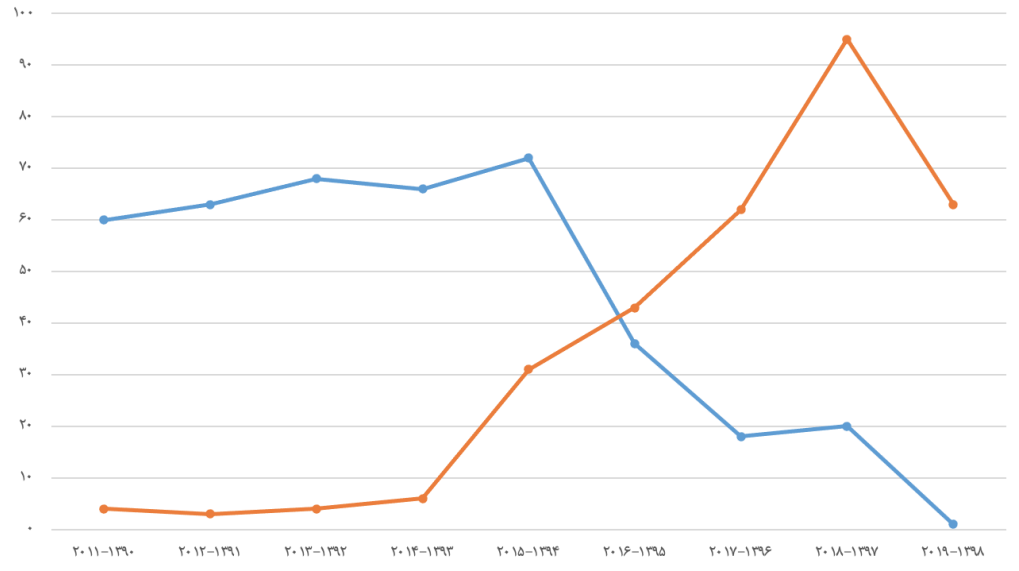
سال	پایگاه گنج	٪	WOS	٪	فرض برابری با اختلاف ۳ سال
۱۳۹۰	0	۰٪	1	۰٪	۰.۰۰۲۹
۱۳۹۱	0	۰٪	3	۱٪	۰.۰۰۵۷
۱۳۹۲	0	۰٪	6	۲٪	۰.۰۱۸۳
۱۳۹۳	1	۰٪	20	۶٪	۰.۰۵۶۱
۱۳۹۴	11	۱٪	68	۱۰٪	۰.۰۹۰۲
۱۳۹۵	15	۱٪	107	۱۱٪	۰.۰۶۱۵
۱۳۹۶	59	۵٪	114	۷٪	-۰.۰۱۲۶
۱۳۹۷	105	۸٪	75	۱٪	-۰.۰۶۷۸
۱۳۹۸	94	۸٪	8		



t	m	n	S	n	معیار پذیرش ۹۰ درصد	Result
۱.۰۹۹۶۱۲۸۵۷	۰.۰۱۹۳	۰	۰.۰۴۹۶۰۰۷۸۲	۸	۱.۹۴۳	Approved

شکل ۳۸- مقایسه تست شکاف زمانی بین مطالعات در کلیدواژه Deep Learning

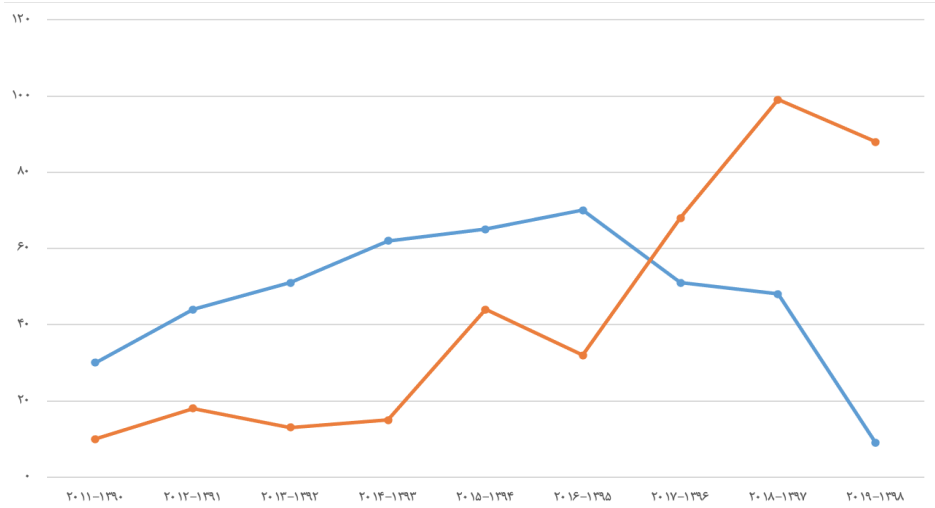
بازه زمانی اختلاف میان اوجگیری مطالعات در زمینه شبکه‌های اجتماعی به ۵ سال باز می‌گردد. این اختلاف بیشترین شکاف زمانی برای مطالعات جاری است که در شکل زیر قابل مشاهده است. البته به دلیل کمی تعداد داده‌های لازم، امکان تست این فرضیه وجود ندارد.



شکل ۳۹- مقایسه روند مطالعات در کلیدواژه Social Media

از دیدگاه خبرگان، درهم تنیختن موضوعات فرهنگی و فناوری در کلیدواژگان مرتبط با شبکه‌های اجتماعی باعث کندی جایگرینی مباحث آن در دنیای علمی داخل کشور شده است.

این اختلاف زمانی در موضوع یادگیری ماشین در همان ۳ سال فاصله باقی مانده است.



شکل ٤٠- مقایسه روند مطالعات در کلیدواژه‌های Machine Learning

۳-۵. دسته‌بندی حوزه‌های علمی بر اساس مدل‌سازی موضوعی کلیدواژه‌ها

همانگونه که پیشتر اشاره شد یکی از شاخه‌های مهم داده‌کاوی، متن‌کاوی است که هدف آن استخراج دانش از داده‌های متنی غیر ساختاریافته یا نیمه ساختاریافته است. از متن‌کاوی با عناوین دیگری چون داده‌کاوی متنی (TDM) و کشف دانش از پایگاه‌های متنی نیز نام برده شده است. کاربردهای عمده متن-کاوی در شناخت متون مشابه، خلاصه‌سازی متون، ترجمه ماشینی متون، انتخاب کلیدواژه یا برچسب‌زنی متون، رده‌بندی اطلاعات متنی و ... است که در فرایند آن از علوم مختلف، مخصوصاً پردازش زبان طبیعی استفاده می‌شود (Choudhary, et al. 2009).

هدف از متن‌کاوی استنتاج دانش ضمنی[1] نهان در متن و ارائه آن به شکل دانش صریح[2] است. برای دستیابی به این هدف معمولاً چهار فعالیت مهم در فرایند متن‌کاوی انجام می‌شود که عبارت‌اند از: بازیابی اطلاعات[3]، استخراج اطلاعات[4]، کشف دانش[5] و فرضیه‌سازی[6]. سیستم بازیابی اطلاعات درصدد است تا متن‌های متناسب با حوزه موردنظر را به دست آورد درحالی‌که استخراج اطلاعات به دنبال یافتن اطلاعاتی با ساختار از پیش تعیین‌شده (مانند استخراج روابط) از این متون است. سیستم کشف دانش موجب دست‌یابی به دانش جدید از اطلاعات استخراج‌شده می‌گردد و درنهایت فرضیه‌سازی به تفسیر دانش کشف‌شده در حوزه مربوطه می‌پردازد. در مرحله بازیابی اطلاعات با استفاده از تکنیک‌های مختلف ازجمله به‌کارگیری اصطلاح‌نامه[7]، بسط جستجو[1]، به‌کارگیری شبکه

[1] Implicit Knowledge
[2] Explicit Knowledge
[3] Information Retrieval
[4] Information Extraction
[5] Knowledge Discovery
[6] Hypothesis Generation
[7] Thesaurus

واژگان[۲] تخصصی حوزه مدنظر و ... تلاش می‌شود تا پایگاهی از متون مرتبط ایجاد شود و در مراحل بعدی از این پایگاه به‌عنوان منبع اطلاعاتی متن‌کاوی استفاده شود. رویکرد موضوعی و محتوایی نیز بر ماهیت و ساختار دانش موجود در منابع موردبررسی دلالت دارد. به عبارتی این رویکرد بر مبنای شاخه‌ها و زیرشاخه‌های علمی شناخته‌شده در حوزه موردنظر بنا شده است. این شاخه‌ها می‌توانند از منابع موردبررسی استخراج شده باشند و یا از شاخه‌های موضوعی رشته موردنظر در دیگر منابع (مانند دانشگاه‌ها یا پایگاه‌های علمی) به دست آمده باشند.

هرچند که استفاده از روش‌های ماشینی و تکیه بر تحلیل حجم انبوه اطلاعات، اخیراً در فرایندهای تصمیم‌گیری و برنامه‌ریزی موردنظر قرار گرفته‌اند اما کاربست این روش‌ها در منابع زبان فارسی بسیار اندک بوده و الگوریتم‌ها و ابزارهای توسعه داده‌شده در این روش‌ها نیز بیشتر مبتنی بر متون عمومی مانند روزنامه و وب‌سایت‌های خبری و ... است و کمتر به متون علمی و پژوهشی پرداخته شده است. در بخش به مدل‌سازی موضوعی کلیدواژگان یافت شده در پایگاه گنج می‌پردازیم.

پیشتر اشاره شد که مدل‌سازی موضوعی از روش‌های کارای تحلیل متن است که هدف آن گروه‌بندی مجموعه‌ای از متون مختلف به‌صورت زیرمجموعه‌های معنادار که موضوع[۳] نامیده می‌شوند است. الگوریتم‌های روش‌های مدل‌سازی موضوعی با حداقل دخالت انسان این فرایند را انجام می‌دهند و همین نکته باعث جذابیت و استفاده روزافزون آن در تحلیل حجم عظیمی از متون شده است. در این روش بجای استفاده از عناوین از قبل تعیین شده، تنها با تعیین تعداد موضوعات، زیرمجموعه‌های موضوعی استخراج شده و میزان ارتباط هر سند با هر زیرمجموعه نیز تعیین می‌شود.

روش‌های استخراج موضوع بر مبنای این فرض عمل می‌کنند که کلمات موجود در یک متن از نظر معنایی به یکدیگر وابسته هستند و معنایی که در مستندات یک موضوع مسنجم وجود دارد از مجموعه کلمات این مستندات به دست آمده است. به عبارتی صرف‌نظر از معانی یا مکان قرارگیری متن، هم‌رخدادی کلمات مورد تحلیل قرار گرفته و اسناد به‌صورت مجموعه‌ای از کلمات در نظر گرفته می‌شود. در مدل‌سازی موضوعی مفروضات زیر در نظر گرفته می‌شود (Wallach 2006):

۱. ترتیب حضور مستندات در پیکره اهمیتی ندارد؛
۲. ترتیب قرار گرفتن کلمات در هر مستند اهمیتی ندارد؛
۳. موضوعات توزیع چندجمله‌ای از کلمات موجود در یک واژه‌نامه هستند؛
۴. کلمات یک مستند از تعدادی موضوع پنهان نشأت گرفته‌اند.

---

[1] Query Expansion
[2] WordNet
[3] Topic

دو فرض اول که ترتیب مستندات در یک پیکره و ترتیب کلمات در یک مستند را نادیده می‌گیرند منجر به خلق مفهومی به نام سبد واژگان[1] می‌شود که نماینده‌ای از یک مستند خواهد بود و اصلی‌ترین ورودی الگوریتم‌های مدل‌سازی موضوعی است (Wallach 2006). فرض سوم و چهارم تبیین‌کننده مفهوم موضوع است که توزیعی از کلمات موجود در یک واژه‌نامه است. درصورتی‌که نیاز به تعیین نام یا برچسب برای یک موضوع وجود داشته باشد، معمولاً از چند کلمه اولی که بیشترین ارتباط را به آن موضوع دارند استفاده می‌شود. در این فصل نیز همانند فصل گذشته از زبان برنامه‌نویسی R استفاده شد. مهم‌ترین ویژگی‌هایی که موجب انتخاب این زبان برنامه‌نویسی در این پژوهش شد، عبارت‌اند از:

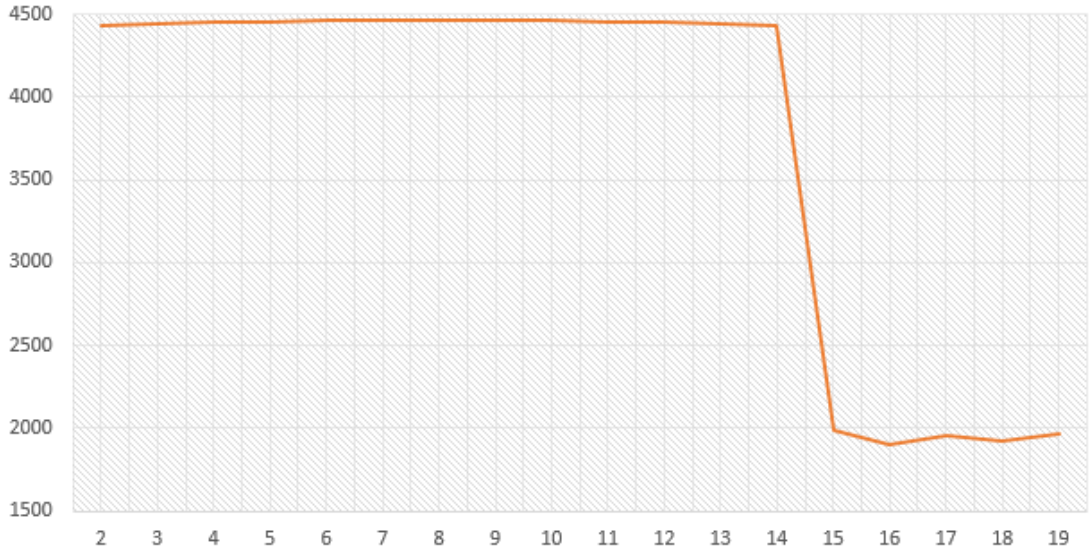
- یک زبان علمی است و قابلیت‌ها و ابزار فراوانی در مدیریت داده‌ها، توابع آماری و ریاضیاتی، توابع داده‌کاوی، متن‌کاوی و پردازش زبان طبیعی را داراست و سطح بالایی از همخوانی بین مفاهیم علمی و آکادمیک در حوزه‌های پردازش و تحلیل داده و داده‌کاوی با ابزارها و توابع توسعه داده‌شده برای این محیط وجود دارد، به‌نحوی‌که بسته‌های نرم‌افزاری، توابع، ابزارها و نحوه کار آن‌ها عمدتاً در مقالات علمی و به زبان آکادمیک تشریح شده و در پایگاه این زبان در دسترس است.
- به دلیل متن‌باز بودن این امکان را فراهم می‌کند تا اجزای داخلی الگوریتم‌ها،مدل‌ها و توابع را مشاهده کرده و امکان تغییر آن‌ها را فراهم می‌کند.
- امکان بسط، توسعه و بومی‌سازی بسته‌ها و توابع در این زبان فراهم است.
- به دلیل ساختار ماتریسی که برای متغیرها در نظر می‌گیرد تطابق زیادی با نگاشت‌های مفهومی در فرایندهای متن‌کاوی و داده‌کاوی دارد و انجام عملیات کشف دانش در آن ملموس است.
- به دلیل متن‌باز بودن گستره زیادی از توابع، ابزارها و راه‌حل‌های موردنیاز در حوزه‌های آماری، پردازش متن و مدل‌سازی را می‌توان در فضای وب جستجو کرده و از آن‌ها استفاده نمود.
- امکان انجام پردازش‌های موازی و توزیع‌شده برای کار با حجم انبوه داده‌ها و پردازش‌های سنگین را دارد که این قابلیت برای داده‌ها و پردازش‌های مربوط به این پژوهش بسیار حیاتی است.
- دارای راهنمای داخلی بسیار خوبی است و در فضای وب گروه‌های علمی زیادی پاسخگوی پرسش‌ها و نیازهای احتمالی خواهند بود.
- رایگان بوده و قابلیت نصب روی سیستم‌عامل‌های گوناگون را دارد و امکان تبادل داده و کد را روی نسخه‌های مختلف داراست.

همانگونه که در فصل قبل نیز مطرح شد پارامتر ورودی مهم LDA، مقدار k است که تعداد موضوعات مدنظر را نشان می‌دهد و ازاین‌رو انتخاب معیاری برای تعیین تعداد موضوعات مناسب بسیار ضروری است. برای این منظور از معیار سرگشتگی (Perplexity) استفاده می‌شود.

[1] Bag of Words

درصورتی‌که برای این آستانه، مقدار پایینی انتخاب شود، هر مستند توسط موضوعات زیادی پوشش داده خواهد شد و نتیجه این خواهد شد که اکثر موضوعات توسط اکثر مستندات تشریح خواهند شد و اگرچه عمده مستندات موجود در پیکره دست‌کم در یک موضوع قرار خواهند گرفت اما موضوع و مستند با ارتباطی بسیار ضعیف به یکدیگر متصل خواهند شد. در عوض درصورتی‌که این آستانه خیلی سخت‌گیرانه انتخاب شود، برخی از موضوعات با تعداد بسیار کمی از مستندات تشریح خواهند شد و بسیاری از مستندات نیز هیچ موضوعی را تشریح نخواهند کرد. برای انتخاب مقدار مناسب k، مدل با استفاده از مقادیر مختلف k اجرا شد و مقدار سرگشتگی برای هر اجرا محاسبه شد. نتایج این انتخاب در زیر آمده است.

این ضریب با توجه به نوع داده‌ها متفاوت بوده و با توجه به محتوای آنها (مثلا فناوری اطلاعات) قابل تعمیم نیست. بنا براین برای داده‌های برامده از پایگاه گنج، مجددا باید میزان سرگستگی را محاسبه نمود. بر این اساس نمودار زیر از ۲ تا ۱۹ دسته برای مجموعه داده‌ها مدلسازی شد.



شکل ٤١- میزان انحراف سرگشتگی با توجه به اندازه دسته‌بندی

همانگونه که از شکل فوق مشاهده می‌شود، تعداد دسته‌های مورد انتظار از عدد ۱۵ کاهش زیادی یافته و قبل از آن دارای سرگشتگی نسبتا بالایی است.

بر این اساس ۱۵ دسته ارتباطات کلیدواژه‌های فناوری اطلاعات برای تشکیل ۱۵ موضوع در جدول زیر (نمایش ده کلیدواژه پر احتمال برای هر موضوع) ترسیم می‌شوند. تعداد بالاتر این دسته‌بندی نسبت به دسته‌بندی‌های بین‌المللی نشان‌دهنده بهره‌گیری کمتر تحقیقات ملی از زمینه‌های پژوهشی چندگانه یا ترکیبی است.

جدول ۱۷- کلیدواژه‌های دسته‌بندی موضوعات با روش مدلسازی موضوعی

دسته	نام گذاری	کلیدواژگان
موضوع ۱	هوش مصنوعی و سیستم‌های خبره	artificial intelligence fuzzy logic expert system reliability recommender systems quality of services electronic banking information security fuzzy system multi-agent system innovation data bases tracking management evaluation
موضوع ۲	شبکه‌های مجازی	social media social network information technology internet electronic commerce trust marketing recommender system information security information and communication technology knowledge management advertising business mobile telephone information retrieval
موضوع ۳	یادگیری و انتخاب ویژگی	feature extraction pattern recognition feature selection learning system identification algorithm face detection neural networks detection biometrics (access control) hidden markov model

دسته	نام گذاری	کلیدواژگان
		image recognition information processing emotion speech recognition
موضوع ۴	اینترنت اشیا	internet of things deep learning wireless sensor networks face recognition information theory entropy routing convolutional neural network anomaly detection problem solving reinforcement learning process mining traffic network coding information
موضوع ۵	پردازش تصویر	image processing feature extraction image segmentation image reconstruction image retrieval edge detection image classification authentication computer vision image recognition wavelet transform machine vision remote sensing magnetic resonance imaging digital image
موضوع ۶	کاربرد فناوری در بانکداری و بیمه	data envelopment analysis efficiency performance evaluation performance assessment ranking productivity decision making unit tehran stock exchange

دسته	نام گذاری	کلیدواژگان
		decision making analytic hierarchy process performance linear programming performance testing multiple criteria decision making banking
موضوع ۷	زبان‌شناسی و قرآن شناسی	semantics persian language quran natural language processing linguistics content analysis koran english language technology semiotics social networks translation discourse analysis poetry information system
موضوع ۸	کاربرد شبکه‌های مجازی	social network social capital facebook internet tehran (city) virtual space iran cyberspace architectural design link prediction trust network analysis tourism satisfaction urban space
موضوع ۹	یادگیری ماشین	machine learning text mining particle swarm optimization system identification signal processing k-nearest neighbor algorithm

دسته	نام گذاری	کلیدواژگان
		ontology game theory metaheuristic algorithm wavelet transform electroencephalography classifier fuzzy clustering knowledge brain-computer interface
موضوع ۱۰	داده‌کاوی	data mining clustering classification decision tree customer relationship management forecasting prediction k-means algorithm ensemble learning business intelligence classification (of information) time series clustering algorithm cluster analysis decision support system
موضوع ۱۱	امنیت حقوقی	privacy visualization quran legal system learning iran graph education anonymity human rights criminal law civil liability jurisprudence cyberspace student
موضوع ۱۲	کاربرد فناوری در پزشکی	neural network support vector machine diagnosis ant colony optimization

دسته	نام گذاری	کلیدواژگان
		artificial neural network data processing breast cancer imperialist competitive algorithm data analysis credit risk patient diagnosis principal component analysis malware geographic information system multilayer perceptron neural network
موضوع ۱۳	امنیت ملی و منطقه‌ای	security iran foreign policy middle east international relations persian gulf threat iran islamic republic crime national security crime prevention strategy terrorism energy america
موضوع ۱۴	پردازش ابری	cloud computing genetic algorithm big data optimization cryptography resource allocation energy consumption evolutionary algorithm scheduling virtual machine load balancing steganography community detection parallel computing resource management
موضوع ۱۵	رمزنگاری و امنیت شبکه	cryptography adaptive control

دسته	نام گذاری	کلیدواژگان
		information security network security security of data intrusion detection system wireless sensor network sentiment analysis computer network design access control intrusion detection opinion mining confidentiality simulation

با توجه به جدول فوق به‌نظر می‌رسد حوزه‌های شناسایی شده هوش مصنوعی و سیستم‌های خبره، شبکه‌های مجازی‌یادگیری و انتخاب ویژگی‌اینترنت اشیا، پردازش تصویر، کاربرد فناوری در بانکداری و بیمه، زبان‌شناسی و قرآن شناسی، کاربرد شبکه‌های مجازی، یادگیری ماشین، داده‌کاوی، امنیت حقوقی، کاربرد فناوری در پزشکی، امنیت ملی و منطقه‌ای، پردازش ابری و نهایتا رمزنگاری و امنیت شبکه است.

موضوع ششم به دلیل استفاده از کلیدواژه فناوری اطلاعات در پیش‌بینی بورس و اقتصاد بانکی، این مدارک بازیابی و دسته‌بندی شده‌اند. همچنین موضوع هفتم به‌علت وجود کلیدواژه Semantics در علوم قرآنی تولید شده و موضوع هشتم به مباحث اجتماعی شبکه‌های مجازی اشاره دارد. همچنین موضوع یازدهم نیز به دلیل وجود کلیدواژه Secuirity در تحقیقات ملی و سیاسی، تشکیل شده و موضوع ۱۳ به دلیل وجود واژه secuitity و وجود آن در مطالعات امنیت ملی ایجاد شده است.

در صورت جدا کردن دسته‌های فوق از دسته‌های فناوری اطلاعات، به ۱۰ دسته‌بندی موضوعی فناوری اطلاعات در تحقیقات ملی می‌رسیم.

۳-۵-۱. مقایسه دسته‌بندی‌ها

در این بخش موضوعات برآمده از روش مدلسازی موضوعی پایگاه WOS را با موضوعات مدلسازی شده پایگاه گنج مقایسه می‌کنیم. بیشتر بودن تعداد دسته‌های برآمده از پایگاه گنج می‌تواند به‌دلیل این باشد که برخی مطالعات در سطح ملی از دایره کمتری از کلیدواژگان وام می‌گیرد و در اصطلاح موضوعات ترکیبی، کمتر مورد توجه قرار می‌گیرد. با توجه به امکان دسته‌بندی مقالات با توجه به کلیدواژه‌های آنها در هر دسته، در این قسمت مقالات پایگاه گنج را در دسته‌های ۱۵ گانه تقسیم

می‌کنیم. همچنین این مقالات را با استفاده از کلیدواژگانشان در دسته‌های ۸ گانه پایگاه WOS نیز افراز خواهیم نمود. بدین شکل تطبیق دسته‌بندی‌های دو پایگاه و نواقص احتمالی نمایان خواهد شد. با توجه به بیشتر بودن تعداد گروه‌های پایگاه گنج نسبت به دسته‌بندی پایگاه WOS، ابتدا مقالات مختلف تخصیص داده شده به دسته‌های مختلف در پایگاه گنج را با توجه به تابع نزدیکی به دسته‌بندی پایگاه WOS افراز خواهیم کرد. در جدول زیر نمونه نزدیکی مجموعه کلیدواژه‌های چند فعالیت را با دسته‌بندی تولید شده در پایگاه WOS نشان می‌دهد. در این نمونه، اعداد بیشینه با رنگ سبز مشخص شده‌اند.

جدول ۱۸- نمونه محاسبه تابع نزدیکی به هر دسته موضوعی

ID	WOS CAT 1	WOS CAT 2	WOS CAT 3	WOS CAT 4	WOS CAT 5	WOS CAT 6	WOS CAT 7	WOS CAT 8
1	0.124285	0.126701	0.124489	0.124293	0.124446	0.125438	0.126011	0.124337
7302	0.124746	0.126186	0.124889	0.123868	0.124387	0.125148	0.126554	0.124222
7307	0.122666	0.124268	0.123655	0.121999	0.123688	0.132099	0.12921	0.122415
9991	0.124635	0.123748	0.12375	0.125046	0.128527	0.125523	0.123908	0.124862
9998	0.124335	0.124609	0.123716	0.125286	0.124686	0.128324	0.124789	0.124253
10000	0.124501	0.123948	0.124048	0.124915	0.123413	0.12715	0.125576	0.126449

با توجه به تخصیص بیشینه اعداد بدست آمده از نزدیکی کلیدواژگان، جایگاه بیشینه هر فعالیت پژوهشی در همه گروهای دسته‌بندی گنج و دسته‌بندی WOS از تابع نزدیکی مشخص می‌شود. در جدول زیر نمونه تخصیص هر مقاله و فعالیت پژوهشی به دسته‌های زیر آمده است:

جدول ۱۹- نمونه اختصاص تابع نزدیکی به هر دسته موضوعی دو پایگاه

ID	نام	برچسب‌های تخصیصی محقق	تاریخ	در دسته بندی گنج	در دسته بندی WOS
833	استفاده از یک رویکرد یادگیری مبتنی بر شبکه‌های عصبی و منطق فازی جهت کشف الگوهای شخصیتی از داده‌های فردی	incident graph; neural network; spectra; died ignorant; data mining	1394	12	5
2957	تشخیص حالت چهره در دنباله تصاویر به روش یادگیری مبتنی بر مدل میدان تصادفی شرطی	conditional random field model; facial expression recognition;facial action unit;image processing;machine learning	1397	4	5
1564	بهبود الگوی تشخیص حالات چهره با هدف افزایش مقاومت نسبت به تغییرات شرایط محیطی و نویز	image processing; pre-processing; feature extraction; facial expression recognition	1396	14	5
1505	استخراج ویژگی‌های بهبود یافته بافت، شکل و رنگ جهت بازیابی تصاویر مبتنی بر محتوا	video or computer games; feature extraction; fernald method; mixed services; ecorangelands; particle swarm optimization	1396	2	3
1550	تشخیص حرکات غیرطبیعی انسان با استفاده از بینایی ماشین	background subtraction;human movement analysis;intelligent video surveillance;feature extraction	1397	14	1
5187	پیاده‌سازی الگوریتم‌های پردازش سیگنال رادار در FPGA	acceleraor; high level synthesis; OpenCL; FPGA; parallel computing; GPU	1397	1	6
2427	بهبود دسترس پذیری داده‌ها در ابر تلفن همراه	accessibility; mobile computing; cloud computing;security	1398	13	8

از شمارش مجموع دسته‌بندی‌های فوق از مدارک تخصیص داده شده با توجه به نزدیکی به هر دسته، جدول زیر قایل توسعه است:

جدول ۲۰- تعداد اختصاص هر پژوهش به هر دسته موضوعی پایگاه WOS

8	7	6	5	4	3	2	1	دسته‌بندی پایگاه گنج / WOS
56	66	89	103	69	72	60	61	1
70	112	71	174	100	89	89	141	2
66	56	112	154	86	68	56	66	3
64	73	68	97	142	178	126	73	4
133	81	134	131	107	90	116	64	5
73	45	59	100	90	49	72	83	9
140	35	42	342	73	76	40	109	10
95	28	65	200	50	63	42	60	12
58	301	52	66	76	129	35	65	14
81	90	69	130	92	143	85	83	15

لازم به ذکر است با توجه به فیلترهای گفته شده در بخش قبل، دسته‌های شماره ۶ تا ۸ و ۱۱ و ۱۳ حذف شده‌اند. تعداد کل فعالیت‌های تطبیق داده‌شده و فیلتر شده میزان ۷۳۴۹ عدد خواهد بود (با توجه به حذف فوق). با توجه به بیشتر بودن دو دسته‌ای دسته‌بندی پایگاه گنج نسبت به پایگاه WOS، بیشینه دو عدد در جدول نرمال شده تخصیص داده می‌شود (خانه‌های سبز جدول زیر) و باقی جدول بصورت تخصیص (۸ سطر و ۸ ستون) مدلسازی می‌شود (خانه‌های زرد رنگ). جدول زیر تطابق موضوعات پایگاه داده WOS و پایگاه گنج را نشان می‌دهد:

جدول ۲۱- تطابق موضوعات پایگاه داده WOS و پایگاه گنج

8	7	6	5	4	3	2	1	دسته‌بندی پایگاه گنج / WOS
0.76	0.90	1.21	1.40	0.94	0.98	0.82	0.83	1
0.95	1.52	0.97	2.37	1.36	1.21	1.21	1.92	2
0.90	0.76	1.52	2.10	1.17	0.93	0.76	0.90	3
0.87	0.99	0.93	1.32	1.93	2.42	1.71	0.99	4
1.81	1.10	1.82	1.78	1.46	1.22	1.58	0.87	5
0.99	0.61	0.80	1.36	1.22	0.67	0.98	1.13	9
1.91	0.48	0.57	4.65	0.99	1.03	0.54	1.48	10
1.29	0.38	0.88	2.72	0.68	0.86	0.57	0.82	12
0.79	4.10	0.71	0.90	1.03	1.76	0.48	0.88	14
1.10	1.22	0.94	1.77	1.25	1.95	1.16	1.13	15

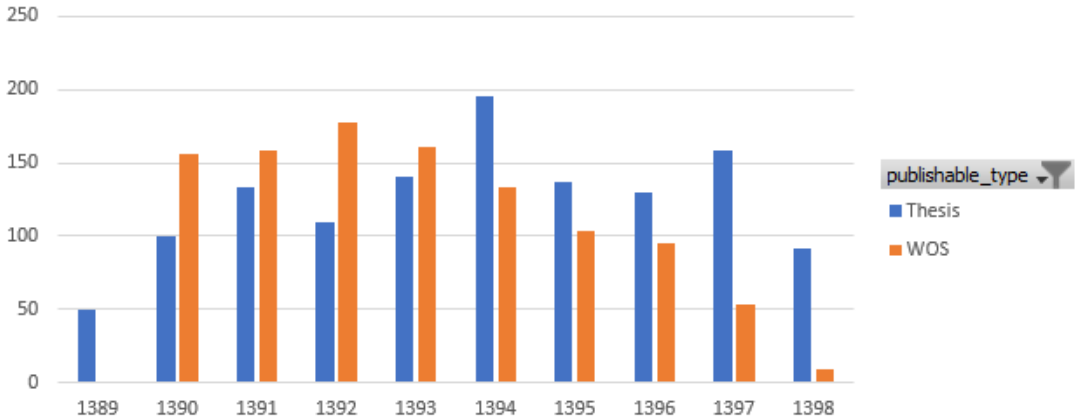
تخصیص فوق نشان می‌دهد که دسته ۱ پایگاه WOS با دسته ۹ پایگاه گنج نزدیکی داشته و به‌ترتیب دسته ۲ و ۳ با ۴، دسته ۴ با ۱۵، دسته ۵ با ۲ و ۱۰ و ۱۲، دسته ۶ با ۵، دسته ۷ با ۱ و ۱۴ و نهایتا دسته ۸ با دسته ۳ پایگاه گنج نزدیکی دارد.

همانطور که پیشتر هم اشاره شد، دسته‌هایی از موضوعات کشف شده مرتبط با فناوری اطلاعات نیستند. این مساله به‌این دلیل پیش می‌آید که در دسته بندی و جستجوی کلیدواژه‌ها در پایگاه WOS، از تقسیم بندی فناوری اطلاعات، کامپیوتر و علوم مرتبط برای استخراج داده‌ها استفاده نمودیم (در فصل قبل مفصل تشریح شد) اما در استخراج داده‌ها از پایگاه گنج، چنین دسته‌بندی موجود نبوده و این اتفاق باعث شده است که کلمه‌ای مانند امنیت که در زمینه‌های علمی زیادی کاربرد دارد، بتواند یک مدل موضوعی (مانند موضوع ۱۱) را اضافه نماید. بدین دلیل در ادامه تحلیل، از موضوعاتی که توسط خبرگان در جدول فوق با هیچیک از موضوعات بین‌المللی تطبیق ندارند (موضوعات ۶، ۷، ۸، ۱۱ و ۱۳) از فهرست ادامه تحلیل کلیدواژگان حذف می‌شوند.

با توجه به تقسیم‌بندی فوق، هریک از مقالات پایگاه گنج می‌توانند در یک دسته‌بندی از پایگاه WOS قرار گیرند. براین اساس در ادامه، فرکانس پرداختن فعالیت‌های مختلف با موضوعات دسته‌بندی شده در این دو پایگاه مورد مقایسه قرار می‌گیرد.

موضوع اول پایگاه WOS مرتبط با موضوع نهم پایگاه گنج و با نامگذاری خبرگان یادگیری یادگیری و انتخاب ویژگی باز می‌گردد.

اگر مطالعات این دو حوزه در در طول زمان مورد توجه قرار دهیم خواهیم داشت:

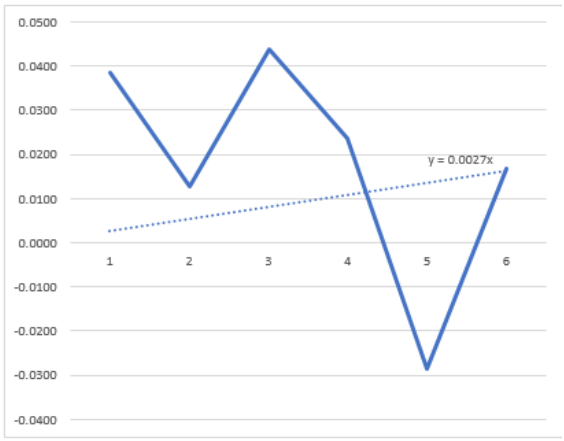


شکل ٤٢- مقایسه تعداد مطالعات در هر دو دسته‌بندی در موضوع اول پایگاه WOS

که نشاندهنده فاصله زمانی ۲ ساله بین موضوعات مطرح شده در سطح ملی و بین‌المللی بوده است.

این فرضیه را با تست آماری T با فرض برابر بودن درصد مطالعات با فاصله زمانی ۲ سال نیز می‌توان اثبات نمود:

سال	پایگاه گنج	٪	WOS	٪	فرض برابری با اختلاف ۲ سال
۱۳۸۹	۵۰	۴٪	۰	۰٪	
۱۳۹۰	۱۰۰	۸٪	۱۵۶	۱۵٪	۰.۰۶۰۷
۱۳۹۱	۱۳۳	۱۱٪	۱۵۸	۱۵٪	۰.۰۳۸۵
۱۳۹۲	۱۱۰	۹٪	۱۷۸	۱۷٪	۰.۰۱۲۷
۱۳۹۳	۱۴۰	۱۱٪	۱۶۱	۱۵٪	۰.۰۴۳۸
۱۳۹۴	۱۹۶	۱۶٪	۱۳۴	۱۳٪	۰.۰۲۳۷
۱۳۹۵	۱۳۷	۱۱٪	۱۰۳	۱۰٪	-۰.۰۲۸۴
۱۳۹۶	۱۳۰	۱۰٪	۹۵	۹٪	۰.۰۱۶۹
۱۳۹۷	۱۵۸	۱۳٪	۵۳	۵٪	
۱۳۹۸	۹۲	۷٪	۹	۱٪	

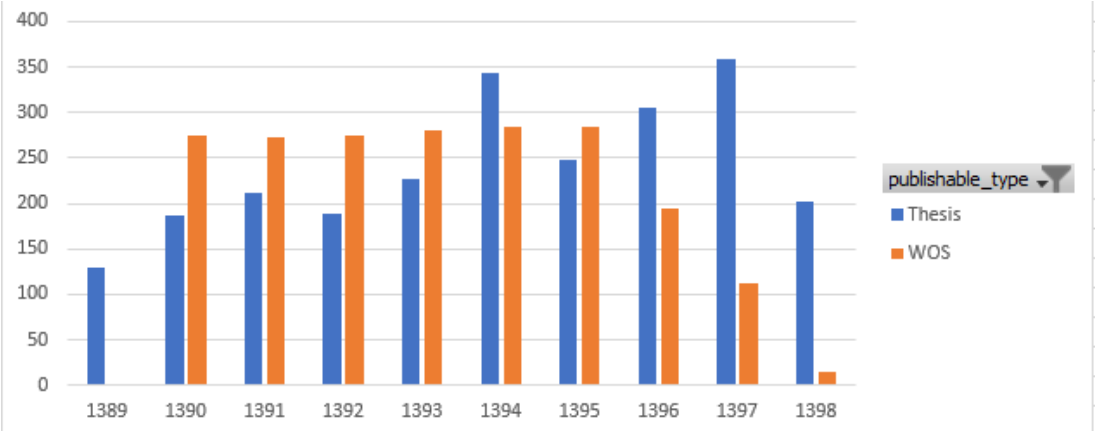


t	m	n	S	n	معیار پذیرش	Result
۱.۷۰۱۹۸۱۱	۰.۰۱۷۹	۰	۰.۰۲۵۷۱۲۶	۶	۱.۸۶	Approved

شکل ۴۳- مقایسه تست شکاف زمانی بین مطالعات در موضوع اول پایگاه WOS

موضوع دوم و سوم پایگاه WOS مشترکا مرتبط با موضوع چهارم پایگاه گنج و با نامگذاری خبرگان اینترنت اشیا باز می‌گردد.

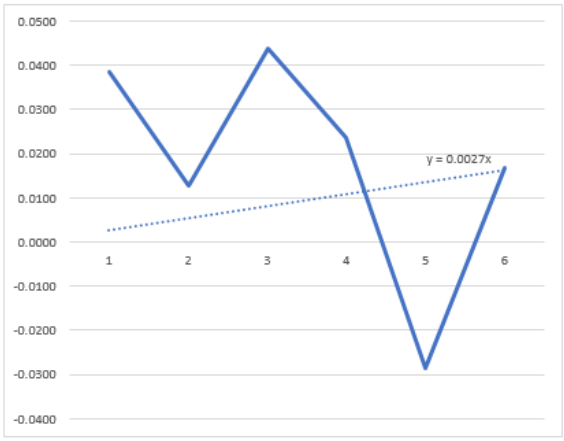
اگر مطالعات این دو حوزه در در طول زمان مورد توجه قرار دهیم خواهیم داشت:



شکل ٤٤- مقایسه تعداد مطالعات در هر دو دسته‌بندی در موضوع دوم و سوم پایگاه WOS

علاوه بر بیش از ۳ سال فاصله زمانی بین این مطالعات، می‌توان میزان بیشتر پرداختن به این موضوعات را در مطالعات بین‌المللی یافت.

سال	پایگاه گنج	٪	WOS	٪	فرض برابری با اختلاف ۲ سال
۱۳۸۹	۱۳۰	۱۰٪	۰	۰٪	
۱۳۹۰	۱۸۶	۱۵٪	۲۷۴	۲۶٪	۰.۰۷۹۵
۱۳۹۱	۲۱۲	۱۷٪	۲۷۳	۲۶٪	-۰.۰۱۴۵
۱۳۹۲	۱۸۸	۱۵٪	۲۷۴	۲۶٪	۰.۰۶۳۵
۱۳۹۳	۲۲۷	۱۸٪	۲۸۰	۲۷٪	۰.۰۲۱۸
۱۳۹۴	۳۴۳	۲۸٪	۲۸۴	۲۷٪	-۰.۰۱۶۹
۱۳۹۵	۲۴۷	۲۰٪	۲۸۴	۲۷٪	۰.۱۰۹۱
۱۳۹۶	۳۰۶	۲۵٪	۱۹۵	۱۹٪	۰.۱۸۶۲
۱۳۹۷	۳۵۹	۲۹٪	۱۱۳	۱۱٪	
۱۳۹۸	۲۰۲	۱۶٪	۱۴	۱٪	

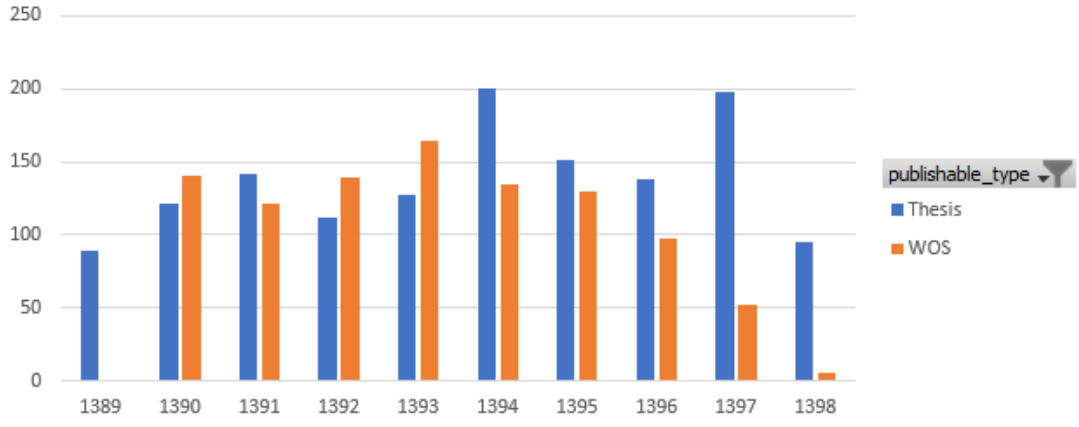


t	m	n	S	n	معیار پذیرش ۹۰ درصد	Result
۱.۹۱۰۷۴۵۸	۰.۰۴۰۴	۰	۰.۰۵۱۸۲۳۸	۶	۱.۹۴۳	Approved

شکل ٤٥- مقایسه تست شکاف زمانی بین مطالعات در موضوع دوم و سوم پایگاه WOS

موضوع چهارم پایگاه WOS مرتبط با موضوع پانزدهم پایگاه گنج و با نامگذاری خبرگان رمزنگاری و امنیت شبکه باز می‌گردد.

اگر مطالعات این دو حوزه در در طول زمان مورد توجه قرار دهیم خواهیم داشت:

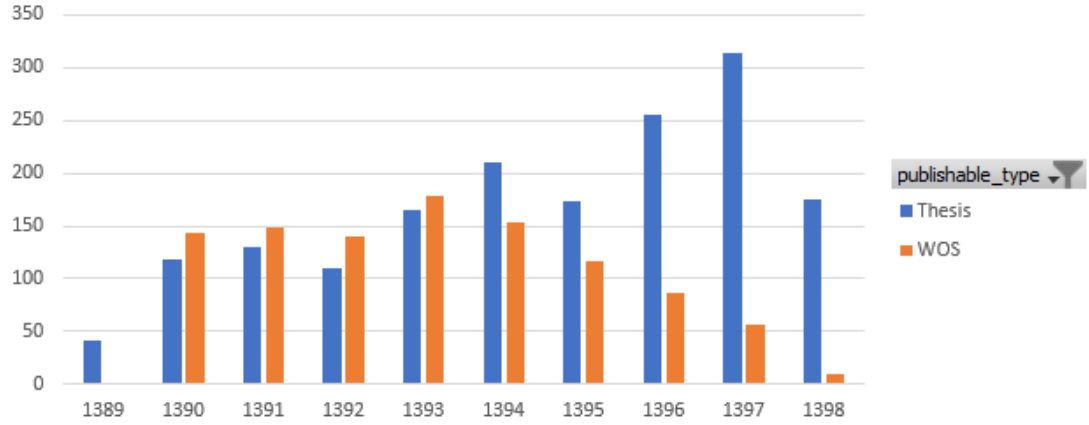


شکل ٤٦- مقایسه تعداد مطالعات در هر دو دسته‌بندی در موضوع چهارم پایگاه WOS

در این حوزه ایجاد فاصله زمانی مورد تایید تست T قرار نگرفت. به‌نظر می‌رسد محققان داخلی در این حوزه فعالیت خوب و متناسب با روند بین‌المللی انجام می‌دهند.

موضوع پنجم پایگاه WOS مرتبط با موضوعات دوم، دهم و دوازدهم پایگاه گنج و با نامگذاری خبرگان کاربردهای فناوری اطلاعات و شبکه‌های اجتماعی باز می‌گردد.

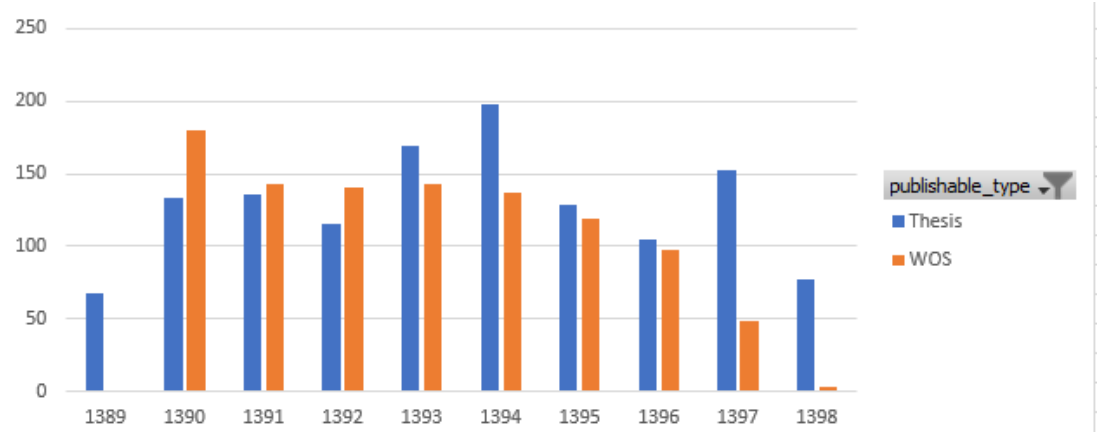
اگر مطالعات این دو حوزه در در طول زمان مورد توجه قرار دهیم خواهیم داشت:



شکل ٤٧- مقایسه تعداد مطالعات در هر دو دسته‌بندی در موضوع پنجم پایگاه WOS

با دقت به نمودار فوق به این نکته می‌رسیم که مطالعات در زمینه جاری به طرز فزاینده‌ای در کشورمان در حال افزایش است که دلیل آن می‌تواند تجمع مطالعا بر کاربردهای اجتماهی شبکه‌های مجازی باشد.

موضوع ششم پایگاه WOS مرتبط با موضوع پنجم پایگاه گنج و با نامگذاری خبرگان پردازش تصویر باز می‌گردد. اگر مطالعات این دو حوزه در در طول زمان مورد توجه قرار دهیم خواهیم داشت:

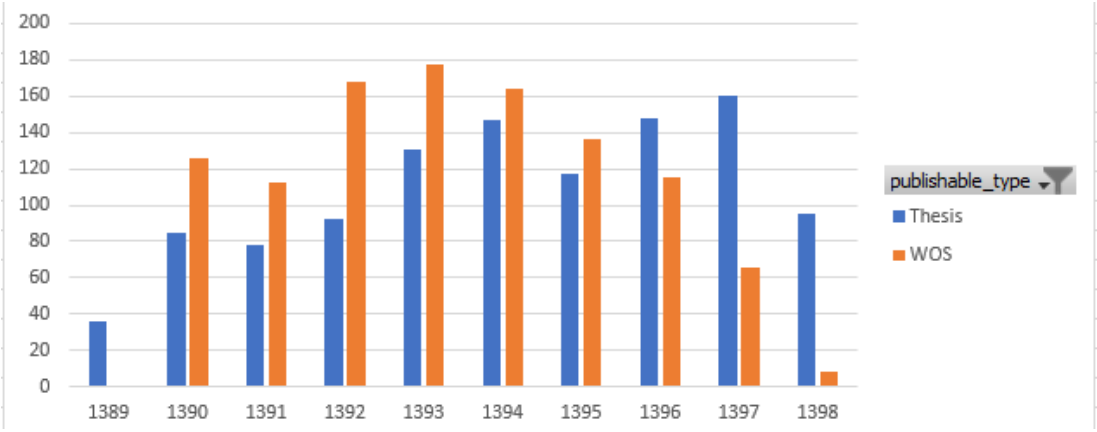


شکل ٤٨- مقایسه تعداد مطالعات در هر دو دسته‌بندی در موضوع ششم پایگاه WOS

در این حوزه نیز اختلاف زمانی میان نوع مطالعات بین دو نمونه مورد اثبات قرار نگرفت.

موضوع هفتم پایگاه WOS مرتبط با موضوعات اول و چهاردهم پایگاه گنج و با نام‌گذاری خبرگان هوش مصنوعی و پردازش ابری باز می‌گردد.

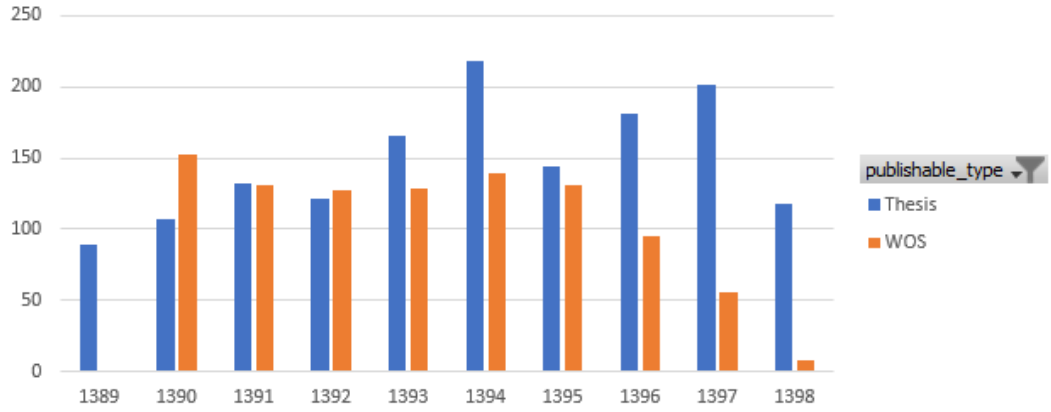
اگر مطالعات این دو حوزه در در طول زمان مورد توجه قرار دهیم خواهیم داشت:



شکل ٤٩- مقایسه تعداد مطالعات در هر دو دسته‌بندی در موضوع هفتم پایگاه WOS

لازم به ذکر است اختلاف سه سال در میزان مطالعات، تنها با اختلاف اندکی (از معیار پذیرش ۹۵ درصد) مورد تایید قرار نگرفت.

و نهایتا موضوع هشتم پایگاه WOS مرتبط با موضوع سوم پایگاه گنج و با نامگذاری خبرگان یادگیری ماشین باز می‌گردد. اگر مطالعات این دو حوزه در در طول زمان مورد توجه قرار دهیم خواهیم داشت:



شکل ٥٠- مقایسه تعداد مطالعات در هر دو دسته‌بندی در موضوع هشتم پایگاه WOS

فارغ از کم شدن تعداد مطالعات در انتهای بازه (سال ۱۳۹۸) به دلایل ذکر شده، به نظر می‌رسد این حوزه در میان پژوهشگران داخلی و بین‌المللی در حال توسعه است.

۳-۵-۲. اعتبارسنجی مدل

همانگونه که پیشتر اشاره شد بیشینه درجه تخصیص هر فعالیت علمی به دسته‌های مختلف استخراج شده، مبنای تخصیص آن مطالعه بر دسته مورد نظر خواهد بود (نمونه جدول ۱۸). لازم به ذکر اینکه هرچقدر ضریب مربوطه میزان بالاتری (نزدیک به ۱) داشته باشد، تخصیص آن مجموعه کلیدواژگان به دسته موضوعی یاد شده بالاتر است. در این مطالعه از مجموعه داده‌های ثبت پایان‌نامه‌های کشور در پایگاه گنج با مشخصات فناوری اطلاعات (حدود ۱۰هزار) رکورد استفاده شد. برای تست نیکویی

برازش این تخصیص بر داده‌های آتی، از کلیدواژه‌های ثبت شده برای رساله‌های دکترای همین پایگاه استفاده می‌شود. با توجه به زمان تقریبی کمتر از سه سال برای تکمیل یک فرایند پژوهشی دکترا و جلوگیری از تکراری بودن مدرک (پیشنهاده و رساله)، داده‌های پیشنهاده‌های ثبت شده از انتهای سال ۱۳۹۵ تا انتهای سال ۱۳۹۷ برای تست استخراج می‌شوند. مجموعه این داده‌ها پس از پاکسازی معادل تقریبا ۱۵ درصد کل داده‌های می‌باشند:

تعداد رکورد پیشنهاده	سال ثبت
۳۸۶	۱۳۹۵
۶۰۹	۱۳۹۶
۵۱۶	۱۳۹۷

تست نرمال برای سنجش یکسانی میانگین دو مجموعه تست و مدل مورد استفاده قرار می‌گیرد. به دلیل زیاد بودن تعداد دومجموعه، آزمون مربوط به تفاضل دو میانگین به شکل زیر تشکیل می‌شود:

$$z = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

نوع	تعداد	میانگین	واریانس
مدل	10907	0.127852259	3.1999E-06
تست (پیشنهاده)	1598	0.127403362	2.30774E-06
		z	1.077
		z 95%	1.650

با توجه به محاسبات فوق تفاوت میان میانگین‌های توزیع دو مجوعه تست و مدل اصلی با احتمال ۹۵ درصد اثبات نشد.

۳-۶. جمع‌بندی و نتیجه‌گیری

گسترش علم و فناوری و امکان تعامل و ارتباط بین پژوهشگران و دانشمندان موجب شده است تا از یکسو شاهد تعدد و گوناگونی پژوهش‌های صورت پذیرفته در حوزه‌های مختلف علوم باشیم و از سوی دیگر با تنوع نیازهای و تقاضاهای پژوهش مواجه شویم. به منظور مدیریت نظام پژوهش، داشتن شناخت کافی نسبت به تقاضاها و نیازهای پژوهشی حال و آینده و توانمندی‌ها و منابع پژوهشی

موجود ضروری است. پیش از این پژوهشگران برای بررسی و تحلیل آثار یک دوره، ملزم به استفاده از روش‌های دشوار کیفی بودند ولی با ظهور فناوری‌های نوین تحلیل اطلاعات و نیز توسعه پایگاه‌های اطلاعات علمی این امکان برای پژوهشگران پدید آمد تا با پژوهش‌های کارآمد و متنوعی روی داده‌های وسیع انجام دهند و بتوانند بینش مناسبی از تعاریف، مرزبندی‌ها، روندها، کاستی‌ها و نیازهای پژوهشی یک حوزه علمی ارائه نمایند (خاصه و سهیلی ۱۳۹۷). استفاده پژوهشگران از این پایگاه‌ها به‌عنوان مرجعی برای شناسایی آخرین تحقیقات علمی انجام شده در حوزه‌های مختلف علوم گسترش یافته است. ازاین‌رو کاربرد داده‌کاوی و متن‌کاوی به‌عنوان زیرشاخه‌ی مهمی از آن به‌منظور کشف دانش از منابع موجود و ابزاری برای به دست آوردن الگوهای جدیدی از اطلاعات روزبه‌روز گسترده‌تر شده است. از این رو، پژوهش‌های علم‌سنجی متعددی با هدف شناسایی روندهای پژوهش، تعیین کاستی‌های پژوهشی و نیازسنجی و اولویت‌بندی پژوهش‌ها انجام شده است. سیاست‌گذاران پژوهش پس از اطلاع از خلأ پژوهشی و نیز پژوهش‌هایی که تقاضای قابل‌توجهی در جامعه پژوهشی برای آن‌ها وجود ندارد و با بهره‌گیری از دانش و اطلاعات دیگری که در اختیار دارند و نیز در نظر گرفتن سیاست‌های کلان پژوهشی، نیازهای پژوهشی را شناسایی و تعیین می‌کنند. در این مرحله سیاست‌گذار ممکن است سیاست‌ها و برنامه‌هایی را اعمال کند که تقاضای پژوهشی خاصی در جامعه پژوهشی ایجاد شود و این نیاز به پژوهشگران منتقل گردد. درنهایت سیاست‌گذاران و مدیران بر اساس معیارهای مختلف و با توجه به منابع موجود اولویت‌های پژوهشی را در حوزه‌های مختلف تعیین می‌کنند. این پژوهش در صدد است راهکار مناسبی را مبتنی بر تحلیل محتوای پژوهش‌های موجود و از طریق کاربست تکنیک‌های متن‌کاوی ارائه نماید که از طریق آن کاستی‌های پژوهشی به صورت مستقل از نظر خبرگان و در فرایندی ماشینی شناسایی شود. برای این منظور حوزه فناوری اطلاعات به عنوان مورد مطالعه انتخاب شد و منابع پژوهشی موجود در پایگاه ایرانداک و نیز نمونه داده‌های پایگاه WOS به عنوان داده‌های پژوهش مورد استفاده قرار گرفت.

به منظور شناسایی متون پژوهشی مرتبط با حوزه مورد نظر و نیز شناسایی و تحلیل موضوعات موجود در این حوزه از تکنیک مدل‌سازی موضوعی استفاده شد. در نهایت موضوعات مختلف این حوزه شناسایی شده و تحلیل مقایسه‌ای آنها ارائه گردید. پژوهش ارائه شده می‌تواند سیاست‌گذاران و مدیران پژوهشی را در اتخاذ تصمیمات کلان پژوهشی و تعیین نیازها و اولویت‌بندی پژوهشی یاری رساند.

در این فصل به بررسی موضوعات فناوری اطلاعات در مطالعات ملی و در پایگاه گنج پرداخته شد. در بخش اول از سبد واژگان کلیدی اولیه که در فصل قبل با جستجوی اولیه و سپس نظر خبرگان

تشکیل شد، برای جستجو در پایگاه گنج مورد استفاده قرار گرفت. با توجه به مدارک بازیابی شده، کلیدواژه‌های پر کاربرد استخراج و روند پرداختن به آنها در طول سال مورد بررسی قرار گرفت. همچنین مقایسه زمانی میان رشد کلیدواژه‌ها در پایگاه‌های بین‌المللی و صورت پذیرفت. عدم تطابق این دو الگو از نظر خبرگان مورد بررسی قرار گرفت که در جدول زیر خلاصه شده است:

جدول ۲۲- علت فاصله کلیدواژه‌ها از نظر خبرگان خبرگان

کلیدواژه	دلایل انحراف
Wireless Network(s)	عدم تناسب میزان انتشار به‌دلیل وابستگی مطالعات به زیرساخت
Cloud Computing	اختلاف چند ساله به‌دلیل پیرو بودن پژوهشی (محاسبه شده در فصل) کم بودن تجاری شدن این علم در کشور به‌صورت گسترده
Secuirity	برخی کلیدواژه‌ها در موضوعات فناوری اطلاعات نیست و به همین دلیل تناسب موجود است از نظر عملی به‌نظر می‌رسد فاصله میزان مطالعات وجود دارد
Internet of Things	اختلاف چند ساله به‌دلیل پیرو بودن پژوهشی
Big Data	اختلاف چند ساله به‌دلیل پیرو بودن پژوهشی (محاسبه شده در فصل) عدم وجود داده‌های مناسب و عملیاتی و زیرساخت فناوری مناسب
Deep Learning	اختلاف چند ساله به‌دلیل پیرو بودن پژوهشی (محاسبه شده در فصل) عدم وجود زیرساخت فناوری مناسب
Face Recognition	اختلاف چند ساله به‌دلیل پیرو بودن پژوهشی (محاسبه شده در فصل) عدم وجود زیرساخت فناوری مناسب
Sentiment Analysis	اختلاف چند ساله به‌دلیل پیرو بودن پژوهشی (محاسبه شده در فصل) عدم وجود زیرساخت فناوری مناسب
Feature Selection	اختلاف چند ساله به‌دلیل پیرو بودن پژوهشی (محاسبه شده در فصل) عدم وجود زیرساخت فناوری مناسب
Authentication	اختلاف چند ساله به‌دلیل پیرو بودن پژوهشی (محاسبه شده در فصل) عدم وجود زیرساخت فناوری مناسب
Classification	اختلاف چند ساله به‌دلیل پیرو بودن پژوهشی (محاسبه شده در فصل) عدم وجود زیرساخت فناوری مناسب
Deep Learning	دارای فاصله زمانی و در حال رشد در مقاطع زمانی قعلی
Social Network	دارای فاصله زمانی به‌دلیل تفاوت در نوع مطالعات از توسعه (بین‌المللی) تا کاربرد (مطالعات ملی)
Machine Learning	اختلاف چند ساله به‌دلیل پیرو بودن پژوهشی (محاسبه شده در فصل) کم بودن تجاری شدن این علم در کشور به‌صورت گسترده

در ادامه، روش مدل‌سازی موضوعی مورد توجه قرار گرفت و محققین فارغ از نتایج بدست آمده در توزیع کلیدواژه‌ها در دسته‌بندی‌های بین‌المللی، مجدد دسته‌بندی را بر اساس داده‌های ملی به انجام رساندند. با توجه به افت ضریب سرگشتگی در عدد ۱۵، این تعداد دسته‌بندی برای موضوعات ترکیبی کلیدواژه‌های مرسوم انتخاب شده و این دسته‌بندی در این فصل به نمایش درآمد. برخی از دسته‌ها به علت عدم تقسیم‌بندی داده‌های پایگاه گنج و استخراج بر مبنای کلیدواژه‌ها شکل گرفت که باعث شد ۵ دسته از دسته‌های فوق در حوزه فناوری اطلاعات قرار نگیرد. ۸ دسته به‌دست آمده در مطالعات بین‌المللی با ۱۰ دسته کشف شده در مطالعات ملی مورد مقایسه قرار گرفتند. در این میان با وجود تطبیق این دسته بندی‌ها، برخی کلیدواژگان در هر دستهٔ، در مطالعات ملی کمتر مورد توجه قرار گرفته است.

نقاط خلا یا عقب ماندگی پژوهشها در این کلیدواژگان در هر دسته از نظر خبرگان به‌دلایل زیر بوده است:

جدول ۲۳- مقایسه نقص کلیدواژه دسته‌بندی پایگاه WOS و دسته‌بندی پایگاه گنج

موضوع پایگاه گنج	عنوان	کلیدواژه‌های کم کاربردتر نسبت به دسته‌بندی بین‌المللی	دلایل خبرگان
موضوع ۱	هوش مصنوعی و سیستم‌های خبره	cloud computing big data privacy	کاربردهای عملی مطرح شده و کارایی بررسی نشده است
موضوع ۲	شبکه‌های مجازی	data mining authentication network security	استفاده کمتر از داده‌های واقعی یا عدم تجمیع این نوع داده در کشور پرداختن کمتر به موضوعات امنیت
موضوع ۳	یادگیری و انتخاب ویژگی	security deep learning privacy	فاصله مطالعات ملی از تحقیقات کاربردی کمبود تقاضا پرداختن کمتر به موضوعات امنیت کمبود سرمایه‌گذاری بر تجهیزات
موضوع ۴	اینترنت اشیا	big data face recognition privacy	استفاده بیشتر از الگوریتم‌های ابتکاری برای بهبود و کاربرد کم بدلیل کمبود تقاضا
موضوع ۵	پردازش تصویر	pattern recognition biometrics computer vision	در حال توسعه تحقیقات در این زمینه
موضوع ۹	یادگیری ماشین	online learning wireless sensor networks (wsns)	عدم توسعه شبکه اجتماعی ملی موفق
موضوع ۱۰	داده‌کاوی	cloud computing data mining	استفاده در مباحث عملیاتی در دنیا و مطالعات تئوریک در کشور

موضوع پایگاه گنج	عنوان	کلیدواژه‌های کم کاربردتر نسبت به دسته‌بندی بین‌المللی	دلایل خبرگان
موضوع ۱۲	کاربرد فناوری در پزشکی	feature selection feature extraction authentication sentiment analysis	تئوریک و عدم کاربردی‌سازی کمبود تقاضا
موضوع ۱۴	پردازش ابری	privacy deep learning security	عدم کاربردی بودن در کشور پرداختن کمتر به موضوعات امنیت
موضوع ۱۵	رمزنگاری و امنیت شبکه	face recognition pattern recognition deep learning	در حال توسعه تحقیقات در این زمینه

با توجه به جدول فوق می‌توان دریافت که در زمینه فناوری اطلاعات موضوعات در حال ترکیب و در بین دسته‌های مختلف فناوری اطلاعات در حال تغییر است. همچنین کلیدواژه‌های داغ بصورت پیوسته و با سرعت درحال تغییر هستند (درزمینه فناوری اطلاعات). با توجه به نتایج بررسی شده این تاخیر با فاصله زمان چندساله به مطالعات ملی می‌رسد و نیازمند سیاست‌گذاری بلند مدت برای رفع این نیاز و تاخیر در کشور احساس می‌شود. در بعد کمبود تقاضا برای بسیاری از تحقیقات عملی در کشور باید اشاره نمود که در نقشه جامع علمی کشور، ویژگی‌های اصلی نظام علم و فناوری و نوآوری بیان شده است که اولین ویژگی این نظام به «ترکیب عرضه‌محوری و تقاضامحوری» اختصاص دارد. با توجه به این تعریف لازم است تا توجه ویژه‌ای به پژوهش‌های تقاضامحور شده و از انجام پژوهش‌های غیرمولد جلوگیری کرد. برای دستیابی به این هدف لازم است تا از یک‌سو شناخت کاملی از تقاضاهای پژوهشی کشور داشته و از سوی دیگر پژوهش‌های موجود در کشور را به‌درستی ارزیابی و تحلیل نمود تا با مقایسه این دو، کاستی‌های پژوهشی کشور مشخص شده و پژوهشگران را برای پر کردن آن راهبری نمود.

فهرست منابع

٤-١. فهرست منابع انگلیسی

Astrom, F. (2002). Visualizing library and information science concept spaces through keyword and citation based maps and clusters Emerging frameworks and methods. CoLIS4: Proceedings of the fourth international conference on conceptions of library and information 185–197. Libraries Unlimited: Greenwood Village, CO.

Al-khamayseh, Shadi, and Elaine Lawrence. "Towards Understanding Success Factors in Interactive Mobile Government." 2007.

Banerjee, Rohit . "Initiating GEIT Using COBIT 5 at the Oman Ministry of Manpower." 2016.

Battisti, F., Ferrara, A., Salini, S. (2015). Scientometrics 103:413–433

Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, 55(4), 77–84.

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. The Annals of Applied Statistics, 1(1), 17–35.

Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. The Journal of Machine Learning Research, 3, 993–1022.

Blei, David M. "Probabilistic topic models." Communications of the ACM 55, 2012: 77-84.

Blei, David, Y. Andrew, and Michael Jordan. "Latent dirichlet allocation." Journal of machine Learning research 3, 2003: 993-1022.

Buckland, M. (2012). What kind of science can information science be? Journal of the American Society for Information Science and Technology, 63, 1–7.

Burke, C. (2007). History of information science. Annual Review of Information Science and Technology, 41, 3–53.

Chang, J., & Blei, D. M. (2009). Relational topic models for document networks. In AIStats (9) . 81–88 http://www.jmlr.org/proceedings/papers/v5/chang09a/chang09a.pdf .

De Battisti, Francesca, Alfio Ferrara, and Silvia Salini. "A decade of research in statistics: a topic model approach." Scientometrics 103, 2015: 413-433.

Figuerola, C. G., Javier F., Marco G., Pinto M., (2012). Mapping the evolution of library and information science (1978–2014) using topic modeling on LISA. Siecntometrics.

Furner, J. (2015). Information science is neither. Library Trends, 63(3), 362–377.

Gilchrist, A. (1969). Library and information science abstracts: A brief progress report. Aslib Proceedings, 21(8), 325–327.

Gru¨n, B., & Hornik, K. (2011). Topicsmodels: An R package for fitting topic models. Journal of Statistical Software, 40(13), 1–30.

Hall, David, Daniel Jurafsky, and Christopher Manning. "Studying the history of ideas using topic models." Proceedings of the conference on empirical methods in natural language processing. 2008. 363-371.

Harter, S. P., & Hooten, P. A. (1992). Information science and scientists: ‘‘JASIS’’, 1972–1990. Journal of the American Society for Information Science, 43(9), 583–593.

Hofmann, T. (1999). Probabilistic latent semantic indexing. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval (pp. 50–57.(

John W, Mohr, and Bogdanov Petko. "Introduction-Topic models: What they are and why they matter ." Poetics 41, 2013: 545-569.

Kim, S. J., & Jeong, D. Y. (2006). An analysis of the development and use of theory in library and information science research articles. Library and Information Science Research, 28(4), 548–562.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. In Discourse processes (Vol. 25, pp. 259–284). http://net.pku.edu.cn/*course/cs220/reading/intro_to_LSA.pdf .

Li, W., & McCallum, A. (2006). Pachinko allocation: DAG-structured mixture models of topic correlations. In Proceedings of the 23rd international conference on machine learning ICML ‘06, 2006 577–584. New York: ACM.

Mukherjee, B. (2009). Scholarly research in LIS open access electronic journals: A bibliometric study. Scientometrics, 80(1), 167–194.

Newman, N.C., Porter, A.L., Newman, D., Trumbach, C.C., & Bolan, S.D. (2012). Comparing methods to extract technical content for technological intelligence. In Technology Management for Emerging Technologies (PICMET), 2012 Proceedings of PICMET’12 (p. 12791285.(

Shawn, G., & Milligan, I. (2012). Review of MALLET, produced by Andrew Kachites McCallum. Journal of Digital Humanities, 2(1). http://journalofdigitalhumanities.org/2-1/review-mallet-by-ian-milliganand-shawn-graham ./

Smeaton, A. F., Keogh, G., Gurrin, C., McDonald, K., & Sødring, T. (2002). Analysis of papers from twenty-five years of SIGIR conferences: What have we been doing for the last quarter of a century? ACM SIGIR Forum, 36(2), 39–43.

Steyvers, M., Smyth, P., Rosen-Zvi, M., & Griffiths, T. (2004). Probabilistic author-topic models for information discovery. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 306–315.(

Sugimoto, C. R., & McCain, K. W. (2010). Visualizing changes over time: A history of information retrieval through the lens of descriptor tri-occurrence mapping. Journal of Information Science, 36(4), 481–493.

Swain, D. K., Jena, K. L., & Mahapatra, R. K. (2012). Interlending and document supply: A bibliometric study from 2001 to 2010. Webology, 9(2), 51–60.

Wallach, H. (2006). Topic modeling: Beyond bag-of-words. In In Proceedings of the 23rd International Conference on Machine Learning (p. 977984). Pittsburgh, Pennsylvania, U.S.

Web of Science. Keywords Plus® (Web of Science Core Collection and Current Contents Connect only). 01 29, 2020. http://images.webofknowledge.com/WOKRS534DR1/help/WOS/hp_advanced_search.html (accessed 01 29, 2020.(

Wei, X., & Croft, W.B. (2006). Lda-based document models for ad-hoc retrieval. In Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval 178-185

Yan, E., Ding, Y., & Jacob, E.K. (2012). Overlaying communities and topics: An analysis on publication networks. Scientometrics pp. 1–15.

Yau, C., Porter A., Newman N., Suominen A., (2014). Clustering scientific documents with topic modeling, Scientometrics

٤-۲. فهرست منابع فارسی

ایرانداک. *درباره*. ۱۳۹۸. www.irandoc.ac.ir/about (دستیابی در ۰۹ ۰۱, ۲۰۲۰).

پژوهشگاه علوم و فناوری اطلاعات ایران. *پژوهشگاه علوم و فناوری اطلاعات ایران» (ایرانداک)*. ۱۳۹۵. http://irandoc.ac.ir/about-us/about-us.html (دستیابی در ۱۳۹۵).

خاصه, علی‌اکبر, و فرامرز سهیلی. "ترسیم چشم‌انداز پژوهش در علم‌سنجی و حوزه‌های سنجشی وابسته." *فصلنامه علمی پژوهشی پردازش و مدیریت اطلاعات (۳۳)*, ۱۳۹۷: ۹۴۱-۹۶۶.

ساجدی نژاد, آرمان . *طرح پژوهشی بررسی کارکردهای مدیریت اطلاعات علمی و فناورانه در بافت دولت الکترونیکی و دولت همراه*. تهران: پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک), ۱۳۹۵.

فتاحی, رحمت الله. "حکایت استفاده کاربردی از پژوهش‌ها در ایران,۴۶." *فصلنامه علوم و فناری اطلاعات*, ۱۳۹۰: ۷۷۷-۷۷۹.

قنادینژاد , فرزانه, غلامرضا حیدری, و رحیم چینی پرداز. "تحلیل محتوای متون مربوط به اولویت‌های پژوهشی در علم اطلاعات و دانش‌شناسی." *پژوهشنامه کتابداری و اطلاع رسانی*, ۸, ۱۳۹۷: ۵۵-۷۴.

محبی, آزاده. *طرح پژوهشی تدوین نقشه راه فناوری اطلاعات پژوهشگاه با تمرکز بر پیشرفت‌های نوین در حوزه تجزیه و تحلیل اطلاعات*. پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک), ۱۳۹۶.

نامداریان, ل. *طرح پژوهشی بررسی و تبیین جایگاه پژوهشگاه علوم و فناوری اطلاعات در نظام سیاستگذاری علم و فناوری کشور*. پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک), ۱۳۹۶.

پیوست

دسته‌بندی پایگاه WOS از انتشارات علوم

Acoustics	Emergency Medicine	Literature, American	Physics, Nuclear
Agricultural Economics & Policy	Endocrinology & Metabolism	Literature, British Isles	Physics, Particles & Fields
Agricultural Engineering	Energy & Fuels	Literature, German, Dutch, Scandinavian	Physiology
Agriculture, Dairy & Animal Science	Engineering, Aerospace	Literature, Romance	Plant Sciences
Agriculture, Multidisciplinary	Engineering, Biomedical	Literature, Slavic	Poetry
Agronomy	Engineering, Chemical	Logic	Political Science
Allergy	Engineering, Civil	Management	Polymer Science
Anatomy & Morphology	Engineering, Electrical & Electronic	Marine & Freshwater Biology	Primary Health Care
Andrology	Engineering, Environmental	Materials Science, Biomaterials	Psychiatry
Anesthesiology	Engineering, Geological	Materials Science, Ceramics	Psychology
Anthropology	Engineering, Industrial	Materials Science, Characterization & Testing	Psychology, Applied
Archaeology	Engineering, Manufacturing	Materials Science, Coatings & Films	Psychology, Biological
Architecture	Engineering, Marine	Materials Science, Composites	Psychology, Clinical
Area Studies	Engineering, Mechanical	Materials Science, Multidisciplinary	Psychology, Developmental
Art	Engineering, Multidisciplinary	Materials Science, Paper & Wood	Psychology, Educational
Asian Studies	Engineering, Ocean	Materials Science, Textiles	Psychology, Experimental
Astronomy & Astrophysics	Engineering, Petroleum	Mathematical & Computational Biology	Psychology, Mathematical
Audiology & Speech-Language Pathology	Entomology	Mathematics	Psychology, Multidisciplinary
Automation & Control Systems	Environmental Sciences	Mathematics, Applied	Psychology, Psychoanalysis
Behavioral Sciences	Environmental Studies	Mathematics, Interdisciplinary Applications	Psychology, Social
Biochemical Research Methods	Ergonomics	Mechanics	Public Administration
Biochemistry & Molecular Biology	Ethics	Medical Ethics	Public, Environmental & Occupational Health
Biodiversity Conservation	Ethnic Studies	**Medical Informatics**	Quantum Science & Technology

Biology	Evolutionary Biology	Medical Laboratory Technology	Radiology, Nuclear Medicine & Medical Imaging
Biophysics	Family Studies	Medicine, General & Internal	Regional & Urban Planning
Biotechnology & Applied Microbiology	Film, Radio, Television	Medicine, Legal	Rehabilitation
Business	Fisheries	Medicine, Research & Experimental	Religion
Business, Finance	Folklore	Medieval & Renaissance Studies	Remote Sensing
Cardiac & Cardiovascular Systems	Food Science & Technology	Metallurgy & Metallurgical Engineering	Reproductive Biology
Cell & Tissue Engineering	Forestry	Meteorology & Atmospheric Sciences	Respiratory System
Cell Biology	Gastroenterology & Hepatology	Microbiology	Rheumatology
Chemistry, Analytical	Genetics & Heredity	Microscopy	Robotics
Chemistry, Applied	Geochemistry & Geophysics	Mineralogy	Social Issues
Chemistry, Inorganic & Nuclear	Geography	Mining & Mineral Processing	Social Sciences, Biomedical
Chemistry, Medicinal	Geography, Physical	Multidisciplinary Sciences	Social Sciences, Interdisciplinary
Chemistry, Multidisciplinary	Geology	Music	Social Sciences, Mathematical Methods
Chemistry, Organic	Geosciences, Multidisciplinary	Mycology	Social Work
Chemistry, Physical	Geriatrics & Gerontology	Nanoscience & Nanotechnology	Sociology
Classics	Gerontology	Neuroimaging	Soil Science
Clinical Neurology	Green & Sustainable Science & Technology	Neurosciences	Spectroscopy
Communication	Health Care Sciences & Services	Nuclear Science & Technology	Sport Sciences
Computer Science, Artificial Intelligence	Health Policy & Services	Nursing	Statistics & Probability
Computer Science, Cybernetics	Hematology	Nutrition & Dietetics	Substance Abuse
Computer Science, Hardware & Architecture	History	Obstetrics & Gynecology	Surgery
Computer Science, Information Systems	History & Philosophy of Science	Oceanography	**Telecommunications**

Computer Science, Interdisciplinary Applications	History of Social Sciences	Oncology	Theater
Computer Science, Software Engineering	Horticulture	Operations Research & Management Science	Thermodynamics
Computer Science, Theory & Methods	Hospitality, Leisure, Sport & Tourism	Ophthalmology	Toxicology
Construction & Building Technology	Humanities, Multidisciplinary	Optics	Transplantation
Criminology & Penology	Imaging Science & Photographic Technology	Ornithology	Transportation
Critical Care Medicine	Immunology	Orthopedics	Transportation Science & Technology
Crystallography	Industrial Relations & Labor	Otorhinolaryngology	Tropical Medicine
Cultural Studies	Infectious Diseases	Paleontology	Urban Studies
Dance	**Information Science & Library Science**	Parasitology	Urology & Nephrology
Demography	Instruments & Instrumentation	Pathology	Veterinary Sciences
Dentistry, Oral Surgery & Medicine	Integrative & Complementary Medicine	Pediatrics	Virology
Dermatology	International Relations	Peripheral Vascular Disease	Water Resources
Development Studies	Language & Linguistics	Pharmacology & Pharmacy	Women's Studies
Developmental Biology	Law	Philosophy	Zoology
Ecology	Limnology	Physics, Applied	
Economics	Linguistics	Physics, Atomic, Molecular & Chemical	
Education & Educational Research	Literary Reviews	Physics, Condensed Matter	
Education, Scientific Disciplines	Literary Theory & Criticism	Physics, Fluids & Plasmas	
Education, Special	Literature	Physics, Mathematical	
Electrochemistry	Literature, African, Australian, Canadian	Physics, Multidisciplinary	

ABSTRACT

Is the song of the research in a specific scientific field in the country is in harmony line with global trends? Or where are the conceptual areas of a scientific field such as information technology? Or what is its borders in the scientific world? And if we hit all boarders? Countiually?

These questions always raise in scientific debates. Therefore, in this research project, with the use of methodological approach of Topic Modeling and with text analysis tools and also by processing the results of Topic Modeling, we discover the boundries of research in the field of Information Technology over time (10 years). In this research, a collection of important and leading journals in the field of Information Technology in WOS database is selected and extracted. It should be noted that the criterion for determining the leadingness of these journals in this project is considered from two viewpoints which are their history and their impact factor, tunedby the oponion of experts. Expert keyword portfolios also have been used to extract both national and international data to provide uniformity and we can possibility compare these two sets of sample data (WOS for global and Ganj for local). With the proposed approach, we try to analyze the connections between commonly used words and their temporal evolution and finally extracts the field of researches. Then, compare areas of global research with national trends.

It should be noted that in this research project, research conducted in the field of information technology in the country is limited to the researches registered in the Ganj database and their alignment with international trends which extracted from WOS database.

Keywords: Topic Modeling, Text Mining, Ganj Database, WOS, Information Technology, Irandoc

Iranian Research Institute
for Information Science and Technology (Irandoc)

Research project

Comparison between information technology research trends in theses and dissertations of the country and global trend using Text Analysis Method

By:

Arman Sajedinejad

And

Mohammad Rabeie

Feb 2021