



پژوهشگاه علوم و فناوری اطلاعات ایران
(ایرانداک)

بررسی روند توسعه علوم فناوری اطلاعات در جهان

با توجه به نقش ایرانداک در اطلاع‌رسانی علم و فناوری اطلاعات

۲	مقدمه و روش‌شناسی
۵	تحلیل مستندات و کلیدواژه‌های پر کاربرد
۹	بررسی روند کاربرد و تمرکز کلیدواژگان و حوزه‌های مختلف
۱۵	جمع‌بندی و نتیجه‌گیری

تهیه شده در پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)
پژوهشکده فناوری اطلاعات

بیانم خدا

مقدمه و روش‌شناسی

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^۱، به‌عنوان 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) بصورت به‌روز مورد نیاز بوده است. از سوی دیگر ایرانداک با تجمیع داده‌های فعالیت‌های علمی کشور ارزش افزوده فراوانی (مانند همانندجویی و رصد وضعیت پژوهش کشور و ...) ایجاد نموده است. اگرچه امکان مشاهده روند جستجوها درباره یک کلیدواژه خاص در نمودارهای پایگاه اطلاعات علمی ایران (گنج) قابل مشاهده است، اما کشف روندهای یک حوزه علمی مانند فناوری اطلاعات و مقایسه آن با روندهای جهانی، نیازمند بررسی تخصصی و بکارگیری روش‌های متن‌کاوی است.

^۱ Thesaurus

هدف از این گزارش ارائه اجمالی یکی از دستاوردهای پژوهشگاه علوم و فناوری اطلاعات ایران در پشتیبانی از نیاز پژوهش و مبتنی بر تحلیل داده محور است. این گزارش در راستای نقش ایرانداک در توسعه فناوری اطلاعات و پاسخ به این سوال مطرح می‌گردد که تا چه میزان روند توسعه پژوهش‌ها در یک زمینه خاص علمی در کشور با روندهای جهانی آن زمینه علمی تطابق دارد. برای این منظور، نیاز است که روند توسعه پژوهش‌ها در حوزه‌های مشخصی از علم در دنیا، با بررسی مقالات علمی و گزارش‌های پژوهشی منتشر شده، شناسایی و تحلیل شود. بنابراین در این گزارش، با شناسایی حوزه‌های مفهومی پژوهشی فناوری اطلاعات، روند رشد و توسعه پژوهش در این حوزه‌ها نیز تحلیل می‌شود.

برای این منظور، از رویکرد مدل‌سازی موضوعی^۱، فعالیت‌های پژوهشی در زمینه فناوری اطلاعات در دنیا با ابزارهای متن‌کاوی بررسی شده و با پردازش نتایج مدل‌سازی موضوعی، روند پژوهش‌ها در طول زمان (بازه ۱۰ ساله) کشف شده است. رویکرد مدل‌سازی موضوعی، رویکردی مبتنی بر متن‌کاوی و یادگیری ماشین است که بر مبنای آن، امکان شناسایی حوزه‌های موضوعی یک سند یا مجموعه‌ای از اسناد فراهم می‌شود. در این گزارش با انتخاب مجموعه‌ای از مجلات مهم و پیشرو در زمینه فناوری اطلاعات در پایگاه «وب‌آف‌ساینس»^۲، ارتباطات بین واژگان پرکاربرد و تکامل زمانی آنها تجزیه و تحلیل شده و در نهایت یک تحلیل سطح بالا از مفاهیم مرتبط با حوزه‌های فناوری اطلاعات ارائه گردیده است. منظور از ارائه سطح بالا از مفاهیم و حوزه‌های فناوری اطلاعات، شناسایی حوزه‌های اصلی روندهای فناوری اطلاعات و نمایش واژگان کلیدی هر یک از حوزه‌ها، در یک دوره زمانی خاص است.

انتخاب مجلات مهم و معتبر حوزه فناوری اطلاعات و به دنبال آن، مقالات این مجلات، با توجه به دسته‌بندی موضوعی ارائه شده در «وب‌آف‌ساینس» انجام شده است. دسته‌های موضوعی «وب‌آف‌ساینس» که بیشتری ارتباط را با حوزه فناوری اطلاعات دارند عبارتند از:

- Computer Science, Artificial Intelligence
- Computer Science, Cybernetics

^۱ Topic Modeling

^۲ Web of science (WOS)

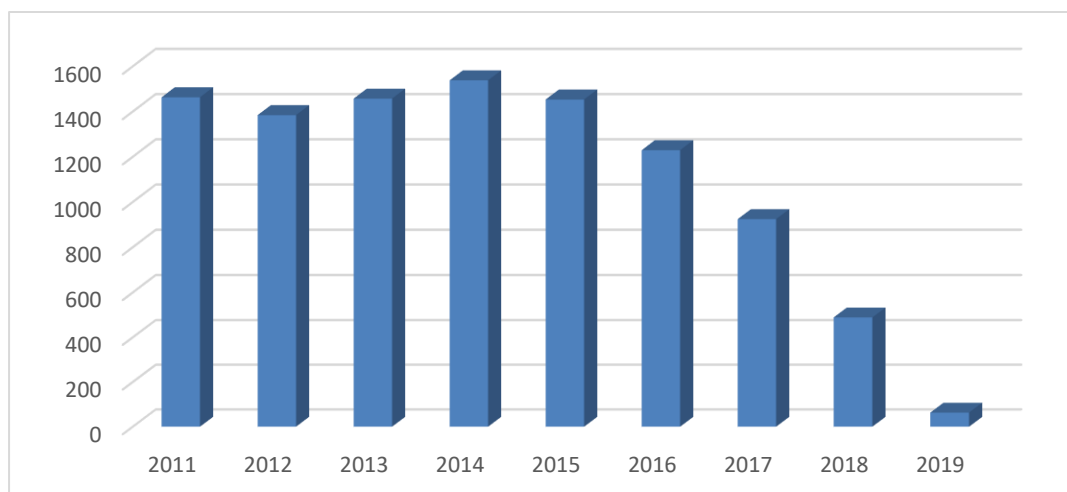
- Computer Science, Hardware & Architecture
- Computer Science, Information Systems
- Computer Science, Interdisciplinary Applications
- Computer Science, Software Engineering
- Computer Science, Theory & Methods
- Information Science & Library Science
- Medical Informatics
- Telecommunications

با در نظر گرفتن این دسته‌بندی موضوعی، مقالات مجلات انتخاب شده‌اند و با رویکردهای متن‌کاوی و مدل‌سازی موضوعی تحلیل شده‌اند تا در نهایت حوزه‌های پژوهشی اصلی در زمینه فناوری اطلاعات و روند تکامل آنها در ۱۰ سال گذشته شناسایی و تحلیل شوند.^۱

^۱ برای توضیحات بیشتر مراجعه شود به طرح پژوهشی باعنوان «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 و همکار دکتر محمد ربیعی، ۱۳۹۹، ایرانداک.

تحلیل مستندات و کلیدواژه‌های پر کاربرد

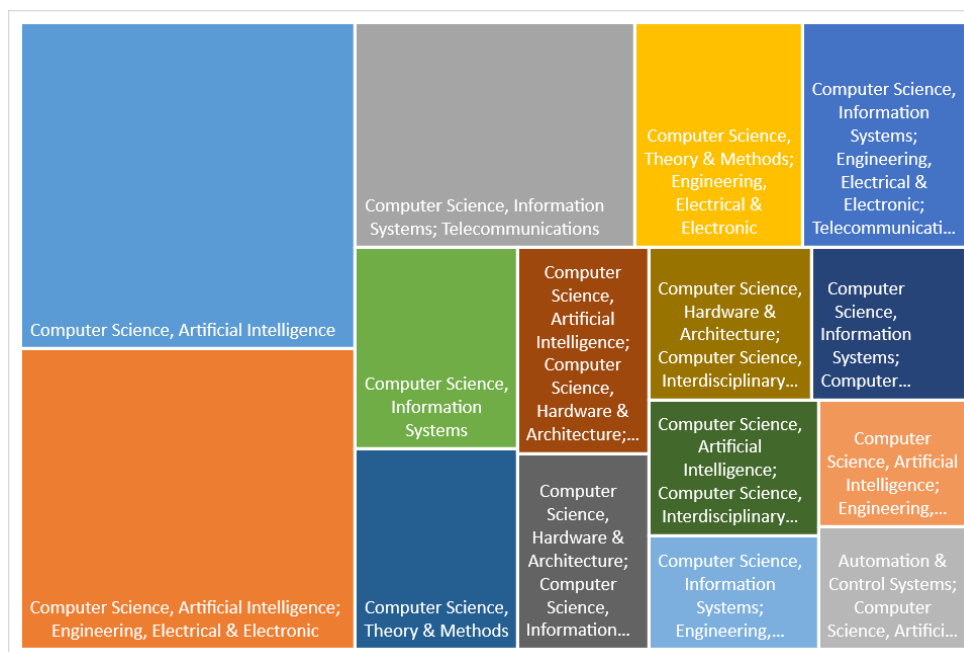
در این گزارش از ۱۰ هزار سند مرتبط با حوزه فناوری اطلاعات و پر استناد استفاده شده است. این انتخاب بر اساس سال انتشار، تعداد کلمات موجود در عنوان در جستجو، میزان ارجاع و شباهت صورت گرفته است. معیار شباهت براساس تعداد کلمات جستجو در عنوان و کلیدواژه با وزن بیشتر و در چکیده و کلیدواژه‌های تکمیلی با وزن کمتر است. توزیع مدارک انتخابی در سال‌های مختلف به صورت زیر است:



شکل ۱- میزان استخراج منبع بر حسب سال

کمتر بودن منابع در سال‌های اخیر بدلیل عدم تکمیل استنادات کافی به این مطالعات بوده است. زیرا معیار اولویت جستجوی و انتخاب، 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.

بسیاری از مطالعات صورت گرفته در زمینه فناوری اطلاعات، منحصر به یک حوزه در پایگاه «وب آف ساینس» نبوده و حوزه‌های ترکیبی از علوم مرتبط با فناوری اطلاعات و مهندسی را در بر می‌گیرند. بر این اساس پانزده حوزه ترکیبی پر کاربرد در زمینه فناوری اطلاعات که از ترکیب دو یا چند حوزه دسته‌بندی شده در این پایگاه است به شرح شکل زیر قرار گرفته است:



شکل ۲- حوزه‌های ترکیبی پرکاربرد

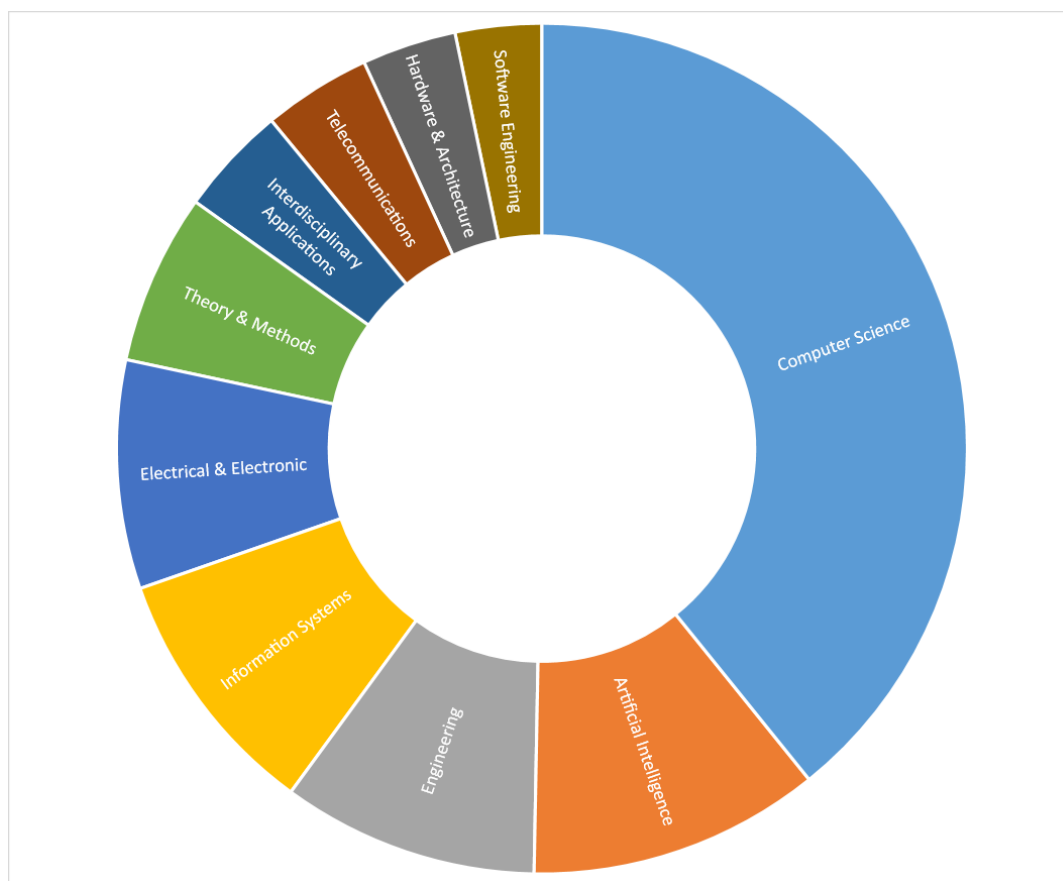
بنابراین ۷ حوزه ترکیبی که بیش از دو سوم درصد مطالعات در زمینه فناوری اطلاعات را به خود اختصاص داده‌اند، شامل حوزه‌های ترکیبی زیر هستند:

- حوزه‌ی ترکیبی علوم کامپیوتر و هوش مصنوعی
- حوزه ترکیبی علوم کامپیوتر، هوش مصنوعی، مهندسی و الکترونیک
- حوزه ترکیبی علوم کامپیوتر، سیستم‌های اطلاعاتی و ارتباطات
- حوزه ترکیبی علوم کامپیوتر، تئوری و روش، مهندسی و الکترونیک
- حوزه ترکیبی علوم کامپیوتر، سیستم‌های اطلاعاتی، مهندسی، الکترونیک و ارتباطات
- حوزه ترکیبی علوم کامپیوتر و سیستم‌های اطلاعاتی
- حوزه ترکیبی علوم کامپیوتر و تئوری و روش

قابل ذکر است ۴ حوزه ترکیبی اول بیش از نیمی از تمرکز مطالعات را پوشش می‌دهند. همچنین نکته قابل توجه دیگر، حضور همزمان علوم کامپیوتر و هوش مصنوعی در صدر این مطالعات است. علاوه بر آن، در ۱۵ حوزه برتر نمایش داده شده که بیش از نیمی از مطالعات را پوشش می‌دهند، زمینه علمی علوم کامپیوتر موجود است.

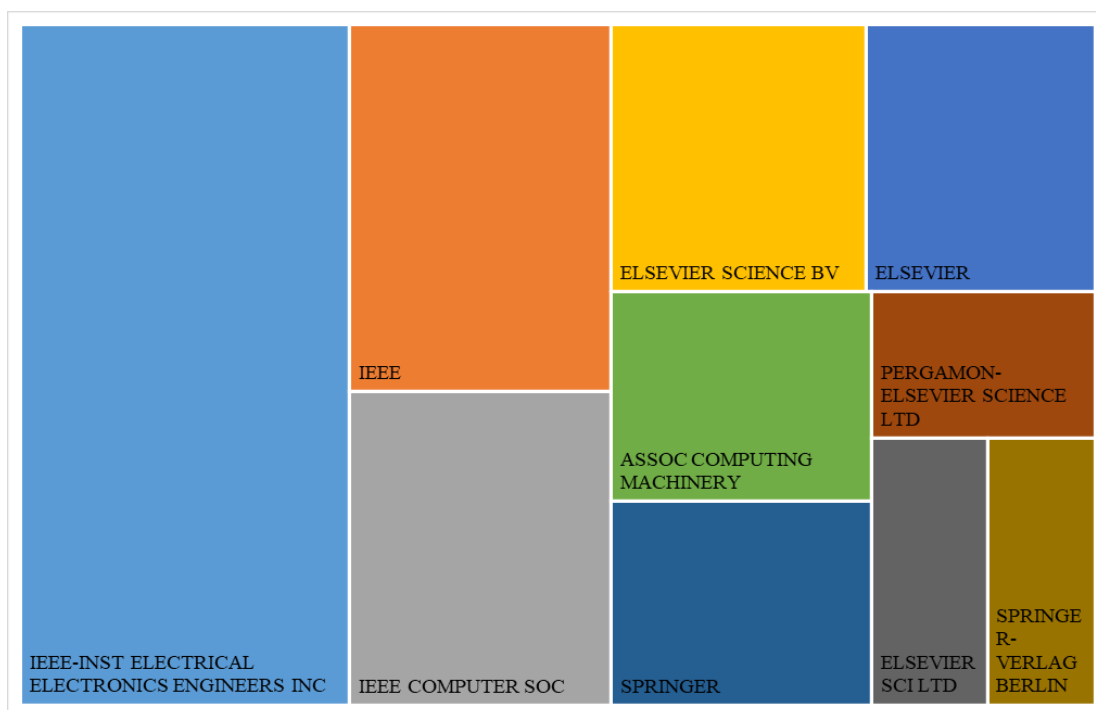
لازم به ذکر است دو حوزه مهندسی (در حوزه‌های عمومی و کاربردی) و الکترونیک نیز جزو دسته‌بندی‌های «وب‌آف‌ساینس» بوده‌اند که با جستجوی مطالعات مرتبط با فناوری اطلاعات به فهرست استخراجی اضافه شدند.

در شکل ۳ نسبت حوزه‌های ذکر شده مستقل (و غیر ترکیبی) مصور شده است. تفاوت این تحلیل با تحلیل قبلی جدا کردن مطالعات در حوزه‌های مختلف به صورت مستقل است. بطور مثال اگر مطالعه‌ای، دو حوزه ترکیبی متفاوت را پوشش داده است، برای هر دو حوزه محاسبه شده و درصد فراوانی حوزه‌های مستقل، تصویر شده است.



شکل ۳- درصد تمرکز بر حوزه‌ها در مطالعات فناوری اطلاعات

همچنین از بُعد 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 الزویر^۱ و آی‌تریپل‌ای^۲ بوده است. ده دسته‌بندی انتشارات برتر که بیش از ۸۰ درصد فعالیت‌ها را منتشر نموده‌اند، در شکل ۴ نمایش داده شده‌اند.



شکل ۴- انتشارات فعال در زمینه فناوری اطلاعات

^۱ Elsevier

^۲ IEEE

بررسی روند کاربرد و تمرکز کلیدواژگان و حوزه‌های مختلف

در گام بعدی از این تحلیل، این سوال مطرح می‌شود که هر حوزه 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 شده است. در جدول ۱ زیر کلیدواژه‌های مورد استفاده (بالاتر از ۵ درصد) برای هر دسته‌بندی پایگاه وب‌آف‌ساینس استخراج شده است.

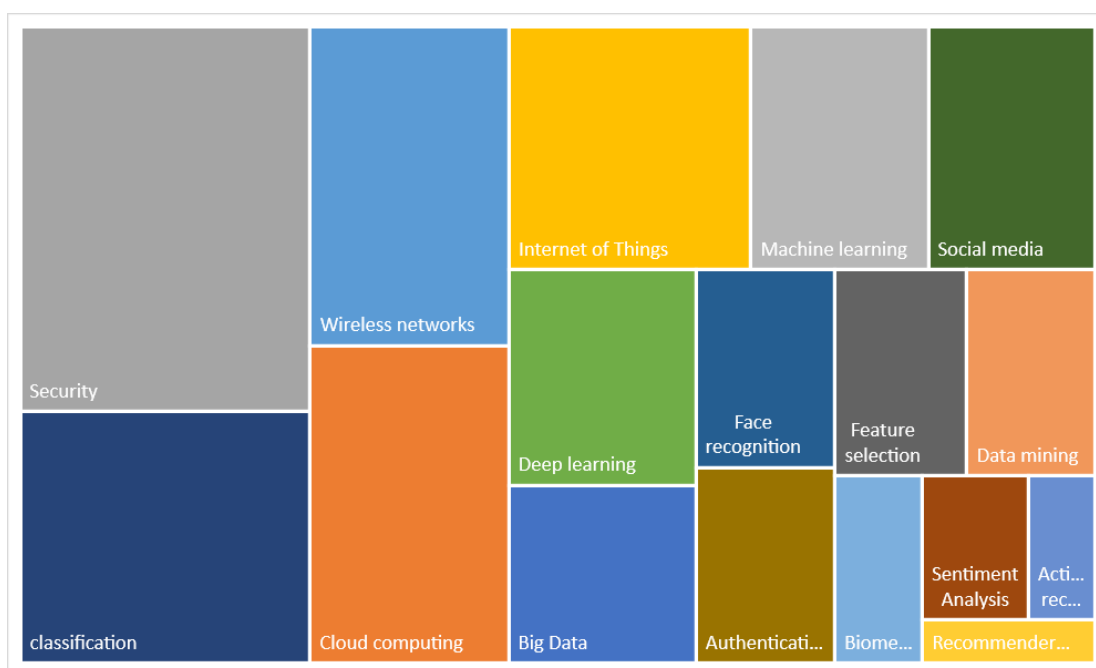
جدول ۱- کلیدواژه‌های مورد استفاده هر دسته پایگاه «وب‌آف‌ساینس»

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
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)

کلیدواژه‌های پر تکرار	دسته‌بندی فناوری اطلاعات	ردیف
- Adaptive Neural Control - Cloud Computing - Neural Network (Nn) - Optimal Control - Reinforcement Learning (Rl) - Adaptive Dynamic Programming - Approximate Dynamic Programming - Attack Graph - Internet of Things	Hardware & Architecture	۵
- Security - Experimentation - Internet of Things (IOT) - Privacy - Wireless Sensor Networks - Algorithms - Interference Alignment - Social Media	Information Systems	۶
- Face Recognition - Physical-Layer Security - Biometrics - Interference Alignment - Network Coding - Physical Layer Security - Wiretap Channel	Electrical & Electronic	۷
- 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	Interdisciplinary Applications	۸
- Cloud Computing - Defect Prediction - Information Visualization - Software Defect Prediction - Visual Analytics - Attack Graph - Feature Detection - Multimedia	Software Engineering	۹

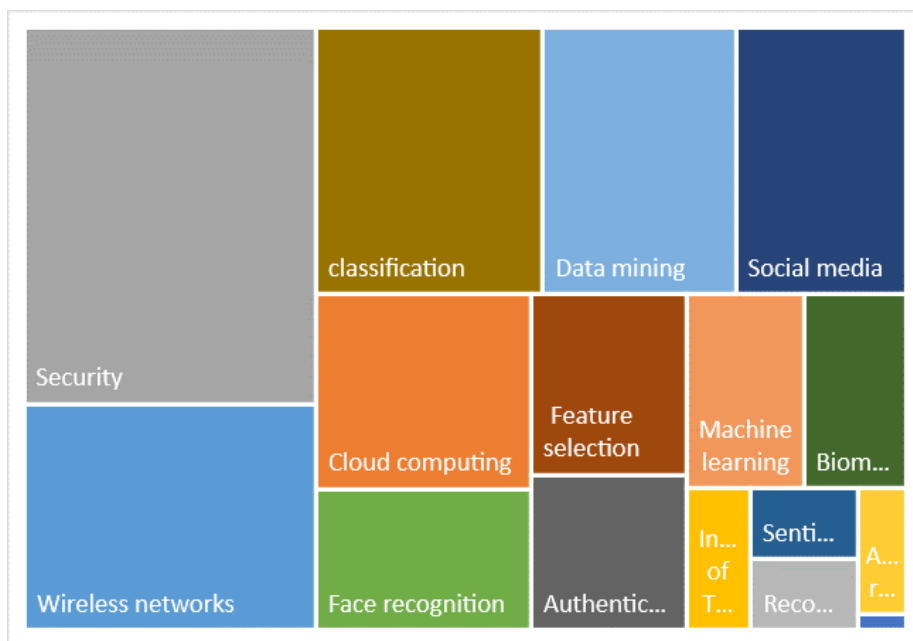
ردیف	دسته‌بندی فناوری اطلاعات	کلیدواژه‌های پر تکرار
۱۰	Telecommunications	• Wireless Sensor Networks • Internet of Things (IOT) • Security • Authentication • Edge Computing

با توجه به فراوانی کلیدواژه‌های مقالات در ۱۰ هزار مجله مرتبط با جستجو، برای ترسیم ۱۵ کلیدواژه برتر به شکل زیر می‌رسیم.



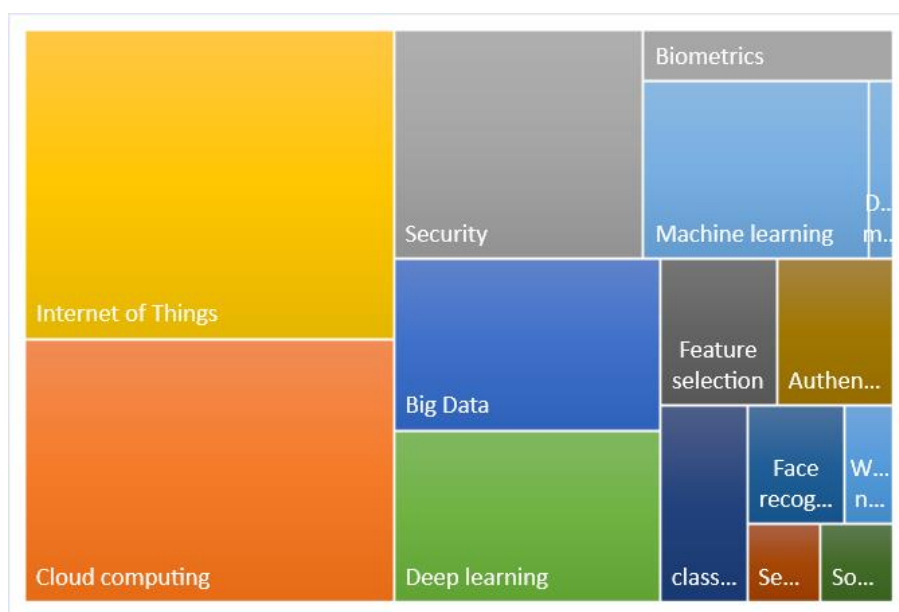
شکل ۵- کلیدواژه‌های پر کاربرد حوزه فناوری اطلاعات در ده سال

با توجه به شکل فوق می‌توان مشاهده نمود که مباحث «امنیت»، «طبقه‌بندی (در حوزه یادگیری ماشین)» و «پردازش ابری» جایگاه ویژه‌ای را در میان کلیدواژه‌گان پر کاربرد در زمینه فناوری اطلاعات دارند. با توجه به اینکه کلیدواژه‌های پر کاربرد در زمینه فناوری اطلاعات با توجه به پویایی زیاد این زمینه علمی ممکن است در طول زمان دچار تغییرات گردد، می‌توانیم حوزه‌ها را در ابتدا و انتهای بازه مورد بررسی یعنی سال ۲۰۱۱ و سال ۲۰۱۹ مقایسه کنیم.



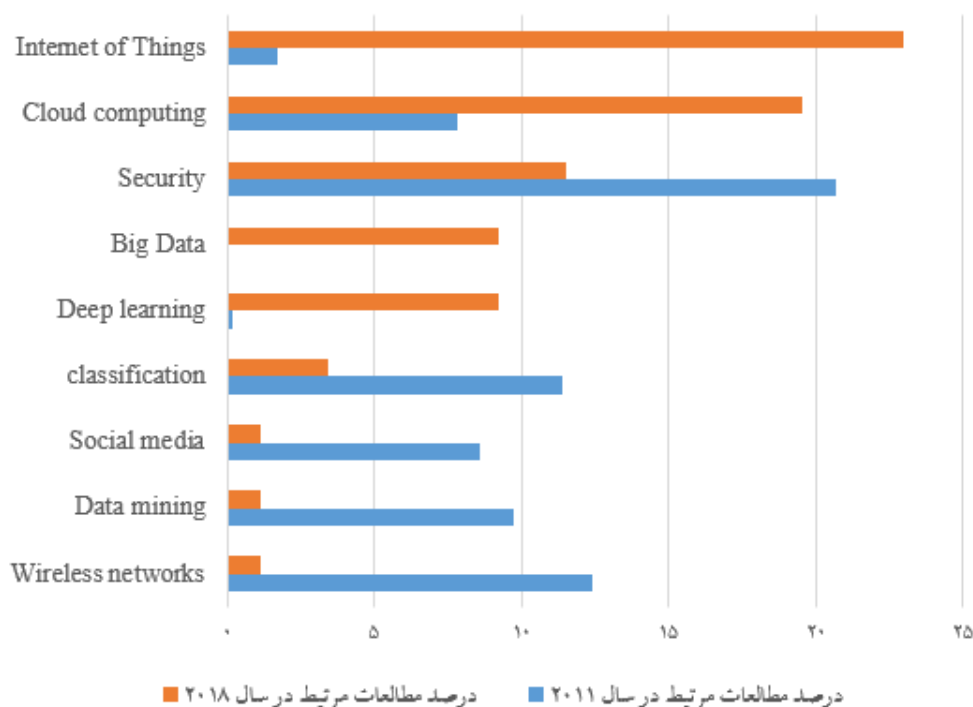
شکل ۶- کاربرد و تکرار کلیدواژه‌ها در سال ۲۰۱۱

همانگونه که در شکل فوق دیده می‌شود، ۴ حوزه «امنیت»، «شبکه‌های بی‌سیم»، «طبقه‌بندی» (در حوزه یادگیری ماشین) و «داده‌های عظیم» بزرگترین حوزه‌های تمرکز کلیدواژه‌های فناوری اطلاعات در سال ۲۰۱۱ هستند. در حالیکه در سال ۲۰۱۹ شکل حوزه‌های زیر را داریم.



شکل ۷- کاربرد و تکرار کلیدواژه‌ها در سال ۲۰۱۹

با یک مقایسه در دو نمودار اخیر می‌توان روند تغییر مسیر مطالعات و پژوهش‌ها را در سال‌های اخیر در زمینه فناوری اطلاعات مشاهده نمود که در شکل ۸ آمده است.



شکل ۸- تغییر تکرار کلیدواژه‌ها از سال ۲۰۱۱ تا سال ۲۰۱۹

همانگونه که در شکل ۸ ملاحظه می‌شود حوزه‌های نوظهوری همانند «اینترنت اشیا»، «محاسبات ابری» و «داده‌های عظیم» در کنار حوزه‌های مانند «یادگیری عمیق» تعریف حوزه‌های فناوری اطلاعات را در بازه زمانی جدید تشکیل داده‌اند و حوزه‌هایی مانند «امنیت» و «شبکه‌های بی‌سیم» در مطالعات جایگاه‌شان کاهش یافته است.

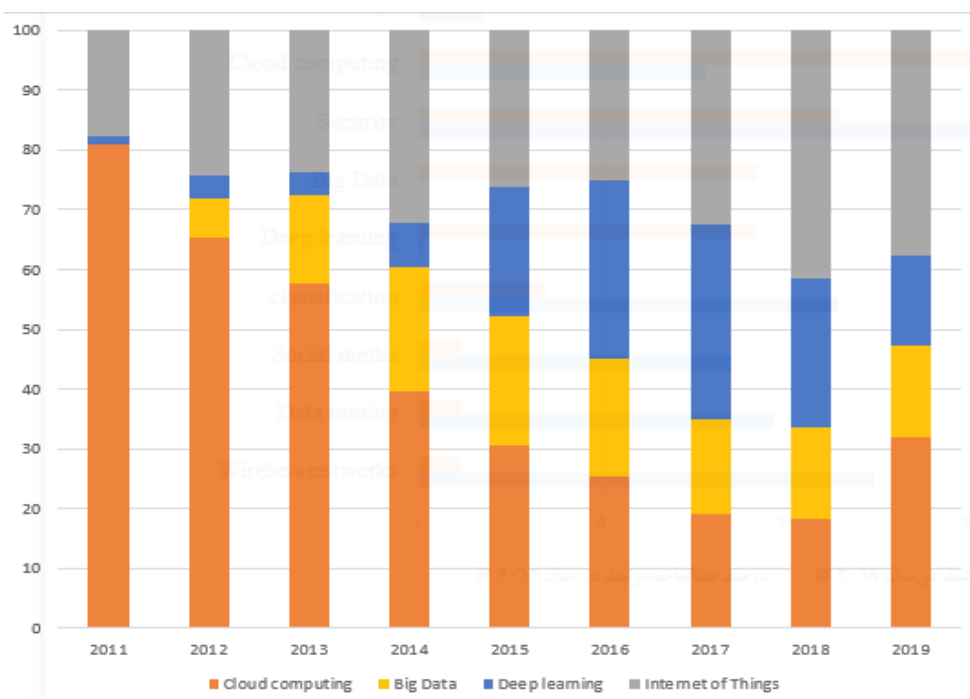
جمع‌بندی و نتیجه‌گیری

در این گزارش به بررسی و مطالعه حوزه‌های (موضوعات) فناوری اطلاعات در مطالعات بین‌المللی با استفاده از روش‌های متن‌کاوی و یادگیری ماشین پرداخته شد.

همانگونه که مشاهده شد رشد زمینه‌های علمی بویژه در حوزه‌های «اینترنت اشیاء» و «یادگیری عمیق» بسیار مشهود است در حالیکه رشد در زمینه‌های علمی مانند اینترنت اشیاء همچنان ادامه دارد. لازم به ذکر است که تعداد کمتر مشاهدات علمی در سال ۲۰۱۹ به علت کامل نشدن تمام انتشارات (یا آنلاین نشدن آنها) است. در نهایت می‌توان نتیجه گرفت چهار حوزه زیر نقش و جایگاه بالاتری در آینده فناوری اطلاعات خواهند داشت:

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

همچنین می‌توان روند رشد کلیدواژه‌های این چهار حوزه را نیز طی دوره ۱۰ ساله با هم مقایسه کرد که نتایج آن در شکل ۹ نمایش داده شده است. با توجه به این شکل که میزان رشد چهار حوزه «اینترنت اشیاء»، «محاسبات ابری»، «داده‌های عظیم» و «یادگیری عمیق» را در مقیاس ۱۰۰ نشان می‌دهد، در میان چهارحوزه اصلی، «اینترنت اشیاء» و «یادگیری عمیق» همچنان پتانسیل بالایی برای توسعه در سال‌های پیش‌رو خواهند داشت.



شکل ۹- روند رشد ۴ حوزه نوظهور کشف شده در پژوهش

این نتیجه را به صورت دقیق در جدول زیر نیز می توان مشاهده کرد. در این میان «اینترنت اشیاء» بیشترین رشد و «شبکه های بی سیم» بیشترین نزول را داشته اند.

جدول ۲- مقایسه کلیدواژه های پر کاربرد در سال ۲۰۱۱ و ۲۰۱۹

کلیدواژه	درصد در بین کلیدواژگان پر کاربرد در سال ۲۰۱۹	درصد در بین کلیدواژگان پر کاربرد در سال ۲۰۱۱	رشد
Internet of Things	۲۲/۹۹	۱/۷۱	۲۱/۲۸
Cloud Computing	۱۹/۵۴	۷/۸۵	۱۱/۶۹
Big Data	۹/۲۰	۰	۹/۲۰
Deep Learning	۹/۲۰	۰/۱۴	۹/۰۵
Machine Learning	۱۰/۳۴	۴/۲۸	۶/۰۷
Sentiment Analysis	۱/۱۵	۱/۴۳	-۰/۲۸
Authentication	۳/۴۵	۴/۵۶	-۱/۱۲
Action Recognition	۰	۱/۱۴	-۱/۱۴
Recommender Systems	۰	۱/۴۳	-۱/۴۳
Feature Selection	۳/۴۵	۵/۲۸	-۱/۸۳
Face Recognition	۲/۳۰	۵/۷۱	-۳/۴۱
Biometrics	۰	۳/۷۱	-۳/۷۱

کلیدواژه	درصد در بین کلیدواژگان پر کاربرد در سال ۲۰۱۱	درصد در بین کلیدواژگان پر کاربرد در سال ۲۰۱۹	رشد
Social Media	۸/۵۶	۱/۱۵	-۷/۴۱
Classification	۱۱/۴۱	۳/۴۵	-۷/۹۶
Data Mining	۹/۷۰	۱/۱۵	-۸/۵۵
Security	۲۰/۶۸	۱۱/۴۹	-۹/۱۹
Wireless Networks	۱۲/۴۱	۱/۱۵	-۱۱/۲۶