



## About "Subject" and "Subject Analysis": Viewpoints and Perspectives

**Farzaneh Shadanpour** 

\*Corresponding author, Ph.D of Knowledge and Information Science, Instructor of National Library and Archives of Iran, Faculty of Education and Psychology, Kharazmi University, Tehran, Iran. E-mail: fshadanpoor@yahoo.com

**Nosrat Riahinia** 

Professor, Department of Knowledge and Information Science, Faculty of Education and Psychology, Kharazmi University, Tehran, Iran. E-mail: sara\_purriahi@yahoo.com

**Keivan Borna** 

Assistant Professor, Department of Mathematics and Computer Science, Faculty of Computer Science, Kharazmi University, Tehran, Iran. E-mail: borna@khu.ac.ir

**Gholam Ali Montazer** 

Professor, Department of Information Technology, Faculty of Industrial/ Systems Engineering, University of Tarbiat- Modarres, Tehran, Iran. E-mail: montazer@modares.ac.ir

### Abstract

**Objective:** Since the discovery and representation of subjects contained in information resources is the most important goal in information description, analysis and retrieval systems, and since subject search is the most common type of user search in databases and library catalogs, in this study what is "subject" and how to analyze and extract subject matters in the process of determining subject keywords for an information resource, as well as some research and operational perspectives are discussed.

**Methodology:** This conceptual paper uses documentary research method to examine the concepts of "subject" and "subject analysis", presented in a selection of related research works in the field of Information Science, as well as considering some corresponding concepts in the field of Computer Science and Natural Language Processing.

**Findings:** Despite various interpretations of what "subject" is, which some assume to be an axiom, and its definition unnecessary, there is consensus, explicitly or implicitly, on its meaning of "aboutness". Subject analysis of information resources, as part of the indexing process, is the analysis and identification of stated topics and concepts and/or obvious features of the information source that may involve manpower, computer algorithms designed to identify

textual terms or the combination of the two is done. Each of these methods has its strengths and weaknesses. Search and retrieval problems due to the multiplicity of subject analysis tools and keyword assignment, low usage rate of standard descriptive schemas, inconsistencies between indexed terms assigned by a single indexer at different times, as well as between multiple indexers for the same information resource, human error in general, the inconsistency of users' search terms with the assigned keywords, and the lack of on time description of resources due to the growing production of them are among the weaknesses of the human subject analysis approach. The application of automated methods of artificial intelligence and natural language processing offers promising prospects for increasing speed and consistency in various processes of describing and organizing information, including the extraction of subject keywords. However, evaluating the efficiency of the output of these methods, alone or in comparison with the keywords of human production, specially from the perspective of users, is required.

**Conclusion:** In the competition of machine algorithms with the human mind in the analysis and recognition of resource subjects, the human mind excels; whether it is an indexer who expresses in his own language what the source is about, or a designer who can design a machine that mimics the computational steps of his mind to saves time and resources. Libraries and information centers can, far from rushing to replace conventional processes and procedures, use the various methods and tools available from the Natural Language Processing and Machine Learning fields to design and develop automated systems for indexing and extracting or assigning subject keywords - and in the larger perspective, automated classification- to use resources (financial, human, and time) for the ultimate goal of maximizing productivity, shortening the subject access path to resources for users, and facilitating the description process.

**Keywords:** Subject, Subject Analysis, Subject Indexing, Natural Language Processing

**Article type:** Research

**How to cite:**

Shadanpour, F., Riahi nia, N., Borna, K., & Montazer, G. A. (2022). About "Subject" and "Subject Analysis": Viewpoints and Perspectives. *Library and Information Sciences*, 25(3), 36-64.

#### ARTICLE INFO

---

**Article history:**

Received: 02/06/2022

Received in revised form: 29/06/2022

Accepted: 14/07/2022

Available online: 21/12/2022

Publisher: Central Library of Astan Quds Razavi  
Library and Information Sciences, 2022, Vol. 25, No.3, pp. 36-64.

© The author(s)





## درباره «موضوع» و «تحلیل موضوعی»: نکته‌ها و چشم‌اندازها

### فرزانه شادان‌پور

\* نویسنده مسئول، مربی پژوهشی سازمان اسناد و کتابخانه ملی ایران، دانش آموخته دکتری علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روان‌شناسی، دانشگاه خوارزمی، تهران، ایران. رایانامه: fshadanpoor@yahoo.com

### نصرت ریاحی نیا

استاد گروه علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روان‌شناسی، دانشگاه خوارزمی، تهران، ایران. رایانامه: sara\_purriahi@yahoo.com

### کیوان برنا

استادیار گروه علوم کامپیوتر، دانشکده علوم ریاضی و کامپیوتر، دانشگاه خوارزمی، تهران، ایران. رایانامه: borna@khu.ac.ir

### غلامعلی منتظر

استاد گروه مهندسی فناوری اطلاعات، دانشکده مهندسی صنایع و سیستم‌ها دانشگاه تربیت مدرس، تهران، ایران. رایانامه: montazer@modares.ac.ir

## چکیده

**هدف:** با توجه به اهمیت بازیابی موضوعات منابع اطلاعاتی و جستجوی موضوعی کاربران، در این پژوهش چستی «موضوع» و چگونگی تحلیل و استخراج موضوعات در فرایند تعیین کلیدواژه‌های موضوعی و مسائل و راه‌حل‌های ممکن مورد بحث قرار می‌گیرند.

**روش پژوهش:** این مقاله پژوهشی مفهومی است که با روش کتابخانه‌ای-سندی مفاهیم «موضوع» و «تحلیل موضوعی» را در گزیده‌ای از متون شاخص علم اطلاعات و دانش‌شناسی با نگاهی بر مفاهیم متناظر در حوزه پردازش زبان طبیعی بررسی می‌کند.

**یافته‌ها:** «موضوع» یا «دربارگی» منابع و تحلیل موضوعی عبارتست از: شناسایی موضوعات و مفاهیم مندرج در منبع اطلاعاتی با قوای فکری نیروی انسانی، با الگوریتم‌های کامپیوتری یا ترکیب این دو. تعدد ابزارهای تحلیل موضوعی و انتساب کلیدواژه‌ها، میزان پایین کاربرد طرح‌های استاندارد توصیف، عدم همگونی در تحلیل موضوعی، عدم همخوانی عبارت‌های جستجوی کاربران با کلیدواژه‌های انتساب داده شده و روزآمد نبودن توصیف منابع، از نقاط ضعف رویکرد تحلیل موضوعی توسط انسان است. روش‌های خودکار هوش مصنوعی و پردازش زبان طبیعی توان افزایش سرعت و یکدستی در فرایندهای مختلف توصیف و سازماندهی اطلاعات و از جمله استخراج کلیدواژه‌های موضوعی را دارند اما، ارزیابی کارآمدی خروجی این روش‌ها، به تنهایی یا در مقایسه با کلیدواژه‌های تولید انسان و از دیدگاه‌های مختلف ضروری است.

**نتیجه‌گیری:** کتابخانه‌ها و مراکز اطلاع‌رسانی می‌توانند به دور از شتابزدگی، از روش‌های پردازش زبان طبیعی و هوش مصنوعی در تحلیل موضوعی و انتساب یا استخراج کلیدواژه‌های موضوعی استفاده کنند و نیروی انسان را در ارتقاء کیفیت خدمات و فرایندها به کارگیرند.

**کلیدواژه‌ها:** موضوع، تحلیل موضوعی، نمایه‌سازی موضوعی، پردازش زبان طبیعی

**نوع مقاله:** پژوهشی

**استناد:**

شادان‌پور، فرزانه؛ ریاحی نیا، نصرت؛ برنا، کیوان؛ منتظر، غلامعلی (۱۴۰۱). درباره «موضوع» و «تحلیل موضوعی»: نکته‌ها و چشم‌اندازها. کتابداری و اطلاع‌رسانی، ۲۵(۳)، ۳۶-۶۴

**تاریخچه مقاله:**

تاریخ دریافت: ۱۴۰۱/۴/۱۱ تاریخ ویرایش: ۱۴۰۱/۵/۷ تاریخ پذیرش: ۱۴۰۱/۵/۲۳ تاریخ انتشار: ۱۴۰۱/۹/۳۰

ناشر: کتابخانه مرکزی آستان قدس رضوی

کتابداری و اطلاع‌رسانی، ۱۴۰۱، دوره ۲۵، شماره ۳، شماره پیاپی ۹۹، صص. ۳۶-۶۴

© نویسندگان



## مقدمه

توصیف منابع اطلاعاتی از جنبه‌های گوناگون مادی و محتوایی و تولید فراداده‌هایی که کاربر با کمک آن‌ها بتواند در کوتاه‌ترین زمان مرتبط‌ترین منابع مورد نیاز خود را بازیابی کند، از جمله فعالیت‌های اصلی در حوزه علم اطلاعات و دانش‌شناسی است. جنبه‌های مادی را نوع محمل یا رسانه‌ای که اطلاعات بر آن حمل شده است، تاریخ نشر، پدیدآور، صفحات و مانند آن تشکیل می‌دهد و جنبه‌های محتوایی، موضوعاتی هستند که در منابع اطلاعاتی مطرح شده‌اند. کشف و بازنمایی این موضوعات مهم‌ترین هدف در نظام‌های توصیف، تحلیل و بازیابی اطلاعات به شمار می‌رود و جستجوهای کاربران در خیل منابع اطلاعاتی نیز، بیشتر جستجوی موضوعی است که از رایج‌ترین و در عین حال چالش‌برانگیزترین نوع جستجو توسط کاربران در پایگاه‌های اطلاعاتی و فهرست کتابخانه‌هاست (گلوب، هگلباخ، و آردو<sup>۱</sup>، ۲۰۱۸). اهمیت موضوع در سازماندهی اطلاعات به قدری است که سهم قابل توجهی از فعالیت‌های این حوزه در تمام دوران‌های گذشته و معاصر به تمهیدات مناسب برای بهبود و ارتقای قابلیت کشف و دسترسی به موضوعات منابع اطلاعاتی اختصاص یافته است و می‌توان ادعا کرد که تاریخ آن به قدمت تاریخ نگارش و ظهور نخستین سندهاست (اندرسن، ۱۳۸۵) و در منابع بسیاری مورد بررسی قرار گرفته<sup>۲</sup>؛ هر چند به خود «موضوع» به صورت مستقل بسیار کمتر پرداخته شده است (لنکستر، الیکر و هارکنس<sup>۳</sup>، ۱۹۸۹). در جستجوی موضوعی، کاربر نیاز خود را در قالب واژه‌ها یا عبارت‌هایی که از ذخیره دانشی او منشأ می‌گیرند، به سیستم اطلاعاتی اعلام و سیستم نیز طی فرایندی عبارت‌های جستجوی کاربر را با محتوای نمایه شده در سیستم تطبیق داده و منابعی را بازیابی کرده در اختیار وی قرار می‌دهد. بنابراین، مهم است که این کلیدواژه‌ها به عنوان نقاط دسترسی موضوعی به درستی تعیین و در اختیار کاربر قرار گیرد. کلیدواژه‌های موضوعی محصول فرایند تحلیل موضوعی و استخراج یا تولید کلیدواژه‌های موضوعی هستند. این فرایند با شناسایی موضوعات و مباحث مطرح در مدرک آغاز می‌شود و سپس واژگانی که فرض می‌شود قادرند به درستی موضوع مدرک را بشناسانند و کلیدواژه نامیده می‌شوند با برداشت مستقیم از متن مدرک یا با انتساب کلماتی از یک بانک واژگان کنترل‌شده به صورت دستی یا ماشینی استخراج می‌شوند (گو<sup>۴</sup>، ۲۰۱۸). کلیدواژه‌ها، واسطه میان کاربر و منبع اطلاعاتی هستند پس لازم است به خوبی بیانگر موضوعاتی باشند که در مدرک مورد بحث قرار گرفته‌اند. در حوزه علم اطلاعات و

1. Golube, Hagebach, & Ardo

۲. به عنوان مثال نگاه کنید به: فتاحی و طاهری (۱۳۸۴)، پوراحمد و حسنی (۱۳۸۷)، فتاحی (۱۳۸۵)، Miksa (2021)، Diao (2018)، Guthrie (1992).

3. Lancaster, Elliker & Harkness

4. Goh

دانش‌شناسی، بررسی منابع اطلاعاتی از حیث موضوعات مندرج در آن‌ها، کشف موضوعات و تعیین واژگانی که نماینده آن‌ها باشند، فهرست‌نویسی تحلیلی یا نمایه‌سازی موضوعی (برای مقالات) نامیده می‌شود. اجرای همین فرایند با استفاده از کامپیوتر و محاسبات در حوزه علوم کامپیوتر و هوش مصنوعی با عبارت «استخراج کلیدواژه» شناخته می‌شود (توشارا، ماونیکا، و مانگامورو<sup>۱</sup>، ۲۰۱۹)، همچنین در حوزه علم اطلاعات و دانش‌شناسی «کلیدواژه» را می‌توان مترادف «عبارت نمایه‌ای»<sup>۲</sup>، «سرعنوان موضوعی»<sup>۳</sup>، «توصیفگر»<sup>۴</sup> و «تگ»<sup>۵</sup> دانست که مفاهیم کلیدی و مهم (همان موضوعات) را در یک متن بازنمایی می‌کند و چنین به نظر می‌رسد که کارکرد اولیه و اصلی آن، توصیف مدارک به منظور بهبود قابلیت یافته شدن<sup>۶</sup> مدارک یا بخش‌هایی از آن‌هاست. کلیدواژه‌ها همچنین برای خلاصه‌سازی متون، ساخت هستان‌شناسی، توسعه سامانه‌های توصیه‌گر، تورق و ساخت عنوان‌های فرعی برای منابع چندرسانه‌ای به کار می‌روند. از کلیدواژه‌ها می‌توان به طور گسترده‌ای برای تعریف پرس‌وجو و بهبود کارکرد در سامانه‌های بازیابی اطلاعات استفاده کرد؛ چرا که به سادگی قابل‌تعیین، بازبینی، یادآوری و به اشتراک‌گذاری هستند و چون مستقل از پیکره‌اند، می‌توان آن‌ها را در پیکره‌های مختلف و متعدد موجود در سامانه‌های بازیابی اطلاعات به کار برد. ضمن این که می‌توان از آن‌ها برای غنی کردن نتایج جستجوها بهره جست (روز، انگر، کرمر و کاوولی<sup>۷</sup>، ۲۰۱۰). اما، «موضوع» چیست و در فرایند تعیین کلیدواژه‌های موضوعی برای یک منبع اطلاعاتی توسط نمایه‌ساز، موضوعات چگونه تحلیل و استخراج می‌شوند؟ مشکلات، مسائل و راه‌حل‌های ممکن در این حوزه کدام‌اند و چه چشم‌اندازهای تحقیقاتی و عملیاتی پیش رو قرار می‌دهند؟

از آنجا که در بسیاری از پژوهش‌های مرتبط با سازماندهی اطلاعات مطالب مربوط به موضوع این مقاله به صورت مکرر ذکر شده‌اند، به منظور پرهیز از تکرار مطالب و مطول شدن فهرست منابع، از مرور سیستماتیک پژوهش‌های موجود صرف نظر شده و مسائل مذکور با مروری برگزیده‌ای از منابع معتبر (مانند استانداردهای حوزه سازماندهی، مرتبط‌ترین پژوهش‌های نظری در داخل و خارج از کشور) با روش کتابخانه‌ای-سندی و انتخاب منبع از استنادهای مرتبط‌ترین پژوهش‌ها، مورد بحث قرار می‌گیرند. همچنین، هرچند در پژوهش‌هایی به «دربارگی»<sup>۸</sup> و «از چیزی بودن»<sup>۹</sup> در منابع اطلاعاتی غیرمتنی (مانند تصویر، صوت و فیلم)

- 
1. Tushara, Mownika & Mangamuru
  2. Index term
  3. Subject heading
  4. Descriptor
  5. Tag
  6. Findability
  7. Rose, Engel, Cramer & Cowley
  8. Aboutness
  9. Ofness

پرداخته شده (ارسطوپور و فتاحی، ۲۰۱۲)، اما با توجه به این که «موضوع» منابع اطلاعاتی غیرمتنی پس از تشخیص این که از چه چیزی هستند و درباره چه هستند، هنوز با استفاده از واژه‌ها نمایه‌سازی می‌شوند، این نوشتار بر منابع اطلاعاتی متنی متمرکز است.

## درباره «موضوع»

واژه «موضوع» در نگاه اول معنایی واضح و بدیهی دارد. «موضوع» چیزی است که درباره آن فکر می‌کنیم، سخن می‌گوییم، بحث می‌کنیم یا می‌نویسیم. در منابع فارسی معانی مختلفی مانند «مقصودی که از آن در عملی بحث کنند» (دهخدا، ۱۳۹۹)؛ «مطلبی که درباره آن بحث می‌شود» (عمید)؛ «امر مورد بحث، چیزی که در مورد آن بحث کنند» (معین) برای واژه «موضوع» ذکر شده است. این واژه در شاخه‌های مختلف علم و اندیشه، معانی اختصاصی تری دارد. واژه «subject» در زبان انگلیسی معادل «موضوع» در زبان فارسی است. برای این واژه در فرهنگستان زبان و ادب فارسی معنای «آنچه درباره آن بحث و آزمایش کنند» (فرهنگستان زبان و ادب فارسی) ذکر شده که در این تعریف اشاره‌ای به معنای واژه «سوژه» در حیطه علم و فلسفه نیز نهفته است. واژه «موضوع» در موقعیت‌های مختلف معانی مختلفی دارد. از مترادف‌ها یا معانی نزدیک به آن می‌توان به «باره»، «پرسمان»، «زمینه»، «نهشته»، «جستار»، «رازی، ۱۳۷۲»، «سوژه»، «مبحث»، «مسئله»، «مشکل»، «مطلب»، «محمول»، «باب»، «خصوص»، «فقره»، «قضیه»، «نهاد»، «گذارده»، «وضع شده»، «قرار داده شده»، «ساختگی»، «مصنوع»، «مکذوب»، «مجعول» (خداپرستی، ۱۳۷۶)، «درون‌مایه» و «مضمون» (پورتر، ۱۳۹۷، ص. ۴۱۳) را نام برد. از دیدگاه معرفت‌شناختی، پرسش اصلی این است که آیا موضوع یک مدرک چیزی «درونی» و «ذهنی»<sup>۱</sup> است یا «بیرونی» و «عینی»<sup>۲</sup>؟ آیا موضوع یک مدرک است؟ یا چیزی است که نمایه‌ساز از تفسیر مدرک تولید می‌کند؟ ماهیت هستی‌شناختی موضوع یک مدرک چیست؟ درک عمیق‌تر از این موارد برای همه نظریه‌ها و عملکردهای سازماندهی دانش و بازیابی اطلاعات لازم است تا بدانیم موضوعات مدارک چگونه باید تعیین شوند و بتوانیم تحلیل موضوعی مفید و بازمنون‌های موضوعی خوبی از مدرک تولید کنیم (دوتا و دوتا، ۲۰۱۳).

واژه «موضوع» علاوه بر اهمیت، سابقه دیرینی نیز در حوزه علم اطلاعات و دانش‌شناسی دارد. در عرصه اطلاعات و منابع اطلاعاتی، «موضوع» همان ایده‌ها و مضامینی است که در منبع اطلاعاتی مورد اشاره قرار گرفته است؛ یعنی آن چیزی است که محتوای منبع درباره آن صحبت می‌کند. اغلب اندیشمندان علم

1. Subjective  
2. Objective  
3. Dutta & Dutta

اطلاعات و دانش‌شناسی نیز موضوع یک اثر را مبحث یا مباحثی می‌دانند که در اثر درباره آن‌ها صحبت شده است. بنابراین، مفهوم «موضوع» برای این حوزه تا حدود زیادی وابسته به مدرک، سند و اثر است. اما، برخی از آن‌ها نیز با الهام از سایر شاخه‌های علمی، برداشت‌های متفاوتی را ارائه کرده‌اند که ابعاد مختلف این واژه را به ویژه در بافت علم اطلاعات و دانش‌شناسی، روشن‌تر کرده است. اولین بار در دهه ۱۹۶۰ رابرت فیرثورن<sup>۱</sup> اصطلاح «دربارگی»، یا «درباره بودن» را به جای آن به کار برد تا از معانی پیچیده فلسفی که واژه «موضوع» یا همان «subject» را در بر گرفته‌اند، پرهیز کرده باشد (فیرثورن، ۱۹۶۹، نقل در هالی و جودری، ۲۰۲۰) ولی در عمل این واژه معادلی برای «موضوع» و هر دو برای توضیح هم به کار رفتند (هالی و جودری، ۲۰۲۰). چارلز امی کاتر<sup>۲</sup> (۱۹۰۴)، متأثر از عملکردهای حرفه‌ای کتابداری و اطلاع‌رسانی، موضوع را «درون‌مایه»<sup>۳</sup> یا «مبحث (جستار)»<sup>۴</sup> مطرح شده در یک مدرک می‌داند، چه به صورت صریح به آن اشاره شده باشد، چه نشده باشد (کاتر، ۱۹۰۴، ص. ۲۳). کیزر،<sup>۵</sup> موضوع را چیزهای کلی و شرایط الصاق شده به آن‌ها (محسوسات و فرایندها) می‌داند (کیزر، ۱۹۱۱، ص. ۳۱۲، ۲۷۶). رانگاناتان<sup>۶</sup>، «موضوع» را مجموعه‌ای از ایده‌های سازمان‌یافته می‌داند (رانگاناتان، ۱۹۶۷، ص. ۸۲) و به پیروی از مقوله‌بندی‌های منطق ارسطو، برای هر موضوع ماهیت، ذات، حقیقت و ابعاد و جنبه‌هایی قائل شده و آن‌ها را «چهریزه»<sup>۷</sup> نامیده است. چهریزه‌ها ویژگی‌های مختلف هر موضوع هستند و عبارتند از: چهریزه ماهیت، چهریزه ماده، چهریزه انرژی، چهریزه مکان و چهریزه زمان. البته همه موضوع‌ها ممکن است حاوی همه چهریزه‌های پنج‌گانه نباشند. او همچنین موضوعات را به سه گروه اصلی، مرکب و پیچیده تقسیم کرد (سلطانی، ۱۳۵۱). ویکری<sup>۸</sup> «موضوع» را درون‌مایه‌ها یا توده‌های پیچیده‌ای از جنبه‌های خاص می‌داند که آثار، اعم از کتاب، مقالات نشریات یا بخش‌هایی از آن‌ها، درباره‌شان نوشته شده‌اند (ویکری، ۱۹۶۸). گوپینات<sup>۹</sup> موضوع را مجموعه‌ای نظام‌مند از ایده‌ها می‌داند (گوپینات، ۱۹۷۶). بورکو و برنیر<sup>۱۰</sup> موضوع را کانون<sup>۱۱</sup> اثر و درون‌مایه‌های اصلی آن می‌دانند که کوشش و توجه نویسنده بر آن‌ها معطوف بوده است. مباحث یا جستارها مفاهیمی هستند که از آن‌ها برای معرفی، توجیه، اثبات و تعمیم موضوع مورد مطالعه و گزارش شده در اثر استفاده می‌شود (بورکو و برنیر، ۱۹۷۸، ص. ۱۴۲). هانس ولیش<sup>۱۲</sup>

1. Robert Fairthorne
2. Charles Amy Cutter
3. Theme
4. Topic
5. Kaiser
6. Ranganathan
7. Facet
8. Vickery
9. Gopinath
10. Borko & Bernier
11. Foci
12. Hans Wellisch

(۱۹۹۱) «مبحث» را مفهوم یا درون‌مایه‌ای می‌داند که در مدرک بیان شده یا از آن گرفته شده است؛ ولی از نظر او واژه «موضوع» شناخته‌شده‌تر است و بیشتر از آن استفاده می‌شود. او این دو را مترادف می‌داند ولی معتقد است در سنت علم اطلاعات و دانش‌شناسی «موضوع» اعم از «مبحث» است (ولیش، ۱۹۹۱، ص. xxiv). جودری و تیلور<sup>۱</sup> «دربارگی» را موضوعی که اثر حامل آن است و ابعاد اثر و نوع و شکل آن را نیز شامل می‌شود، تعریف کرده‌اند (جودری و تیلور، ۲۰۱۸، ص. ۶۱۷، ۶۴۶، ۴۵۴). در استاندارد معیارهای پایه برای نمایه‌ها<sup>۲</sup> موضوع (مباحث یا ویژگی‌ها<sup>۳</sup>) مباحث طرح شده در مدرک و/یا ویژگی‌های هر مدرک (مانند نوع<sup>۴</sup>، قالب، رویکرد روش‌شناختی) تعریف شده است (مؤسسه استاندارد ملی آمریکا<sup>۵</sup>، ۱۹۹۷، ص. ۹).

در مدل مفهومی FRSAD<sup>۶</sup> «دربارگی» را ماده موضوعی<sup>۷</sup> یک اثر و مربوط به درون‌مایه‌های<sup>۸</sup> آن دانسته است. «درون‌مایه» نیز در این مدل مفهومی «هر موجودیتی که به عنوان موضوع یک اثر به کار رود» تعریف شده است. درون‌مایه‌ها (و بنابراین دربارگی) می‌تواند هر چیزی باشد - از یک مفهوم انتزاعی گرفته تا شیء محسوس، شخص یا دوره زمانی - هر چیزی که وجود دارد یا در تصور می‌آید. بر اساس این مدل مفهومی، درون‌مایه نوعاً توسط یک نام<sup>۹</sup> یا برجستگی که برای یک ایده به کار می‌رود تا آن را بدون ابهام بشناساند، شناخته می‌شود (الزامات کاربردی مدل مفهومی داده‌های مستند موضوعی (فرساد)<sup>۱۰</sup>، ۲۰۱۱، ص. ۶، ۱۰). این ایده در مفاهیم سایر رشته‌ها مانند فلسفه و نشانه‌شناسی قابل ردگیری است (هالی و جودری، ۲۰۲۰).

در ویرایش و درهم‌کرد سال ۲۰۱۷، ایفلا این مدل مفهومی را با سایر مدل‌ها در یک چارچوب تحت عنوان مدل مرجع کتابخانه<sup>۱۱</sup> گردآورده که ضمن حفظ ایده‌های اصلی در ویرایش‌های قبلی، قرابت بیشتری با مفاهیم و اصطلاحات حوزه علوم کامپیوتر دارد. در این ویرایش نیز از اصطلاح «نام» برای همه آنچه در محتوای یک اثر مورد اشاره قرار گرفته، شامل نام اشخاص، تنالگان، اماکن جغرافیایی و موضوع استفاده می‌کند ولی از واژه «درون‌مایه» اثری در آن به چشم نمی‌خورد (مدل مرجع کتابخانه ایفلا<sup>۱۲</sup>، ۲۰۱۷، ص. ۳۱-۳۵).

دوتا و دوتا «موضوع» را پیوستاری در نظر می‌گیرند که بر پایه مدرک عمل می‌کند و آن را از جنبه

1. Joudrey & Taylor

2. Basic criteria for indexes

3. Topics or features

4. Genre

5. American National Standards Institutes

6. Functional Requirements for Subject Authority Data

7. Subject matter

8. Thema

9. Nomen

10. Functional requirements for subject authority data (FRSAD): a conceptual model.

11. Library Reference Model

12. IFLA Library Reference Model (LRM)



زبان‌شناختی در نظر گرفته‌اند. آن‌ها «موضوع» را مجموعه‌ای از کلمات که به خوبی تعریف شده‌اند، از حیث معنایی با هم ارتباط دارند و می‌توان آن‌ها را اتم یا کوچک‌ترین بخش تجزیه‌ناپذیر از گروه موضوعات مندرج در متن دانست (دوتا و دوتا، ۲۰۱۳) تعریف کرده‌اند. اما به هر حال، همچنان که میکسا<sup>۱</sup> نیز اذعان می‌دارد، موضوعات مندرج در اثر از موجودیت آن، که کلمات و عبارات هستند، جدا نیست. ما به موضوعات با توجه به کلماتی که برای وصف آن‌ها به کار می‌روند، فکر می‌کنیم (یعنی یک درون‌مایه از طریق نام یا برجسب آن) و این موضوعات «فهم‌هایی نامیده می‌شوند که هویت‌های جداگانه خود را دارند». آن‌ها اندیشه‌های مردم هستند که به طریقی بیان می‌شوند (میکسا، ۱۹۸۳، ص. ۷، ۵۷).

لنکستر با گرایشی عمل‌گرایانه معتقد است که «درباره چیزی بودن» و عبارت‌های مترادف آن عبارت‌هایی هستند که به نظر می‌رسد برای اکثر مردم قابل قبول و قابل درک‌اند. او اضافه می‌کند که «من قصد ندارم وارد بحث‌های فلسفی درباره معنای «درباره بودن» شوم. بعضی از نویسندگان تاکنون این کار را کرده‌اند. آن‌ها با این کار در توضیح وضعیت، حداقل تا جایی که به کار نمایه‌سازی موضوعی مربوط می‌شود، شکست خورده‌اند.» (لنکستر، ۲۰۰۳، ص. ۱۳) و استدلالی دارد مبنی بر این که «اگر کسی بخواهد در مورد تعریف دقیق عبارات پیش از انجام هر کاری توافق کند، کار زیادی انجام نخواهد داد، چه در نمایه‌سازی یا در هر فعالیت دیگر» (لنکستر، ۲۰۰۳، ص. ۱۵). در واقع، اغلب این مفاهیم تقریباً مترادف هستند؛ یعنی یک واژه با مترادف‌ها یا نزدیک به مترادف‌ها توضیح داده می‌شود. اگر کسی عبارت اصلی مانند «موضوع»، «درون‌مایه»، «مبحث» و «مضمون» را نفهمد، مفهوم «دربارگی» ممکن است برای او دست نیافتنی باقی بماند (هالی و جوردی، ۲۰۲۰). ویلسون پس از مباحثات فراوان درباره موضوع و درباره بودن نهایتاً به این نتیجه رسید که «موضوع» قابل تعریف نیست (ویلسون، ۱۹۶۸، ص. ۸۸-۹۰). متکالف<sup>۲</sup> نیز معتقد است که اصطلاح «موضوع» به قدری سرشار از ابهام است که هرچند در زبان روزمره رایج است، ولی در معانی فنی نمی‌توان آن را به کار برد (متکالف، ۱۹۷۳). «موضوع» به عنوان پتانسیل معرفتی مدارک نیز تعریف شده است (یورلند<sup>۳</sup>، ۲۰۱۸ الف). «در ایزو ۵۱۲۷ نیز «مفهوم» واحدی از دانش تعریف شده که با ترکیب منحصر به فردی از ویژگی‌ها ساخته شده است (ایزو، ۲۰۱۷، شماره ۳. ۱. ۱. ۰۲) و حوزه موضوعی حوزه‌ای است از یک دانش خاص (ایزو، ۲۰۱۷، شماره ۳. ۱. ۲. ۰۷). در استاندارد ایزو ۵۹۶۳ «موضوع» هر مفهوم یا ترکیبی از مفاهیم است که یک درون‌مایه را در یک مدرک بیان می‌کند و «مفهوم» عبارت است از واحد فکر (ایزو، ۱۹۸۵، ص. ۱). این تعریف به تعبیر علوم فلسفه و علوم شناختی نزدیک می‌شود که مفاهیم (اجزای تشکیل‌دهنده موضوع) را سازنده

1. Miksa  
2. Metcalfe  
3. Hjørland

افکار می‌داند که برای فرایندهای روان‌شناختی مانند مقوله‌بندی، استنباط، حافظه، یادگیری و تصمیم‌گیری ضروری است و میان زبان طبیعی - که متن نیز نوعی از آن است - و مفهوم، قائل به چنان رابطه تنگاتنگی است که بعضی از فیلسوفان تا جایی پیش رفته‌اند که معتقدند لازمه داشتن مفاهیم در ذهن، زبان طبیعی است (مارگولیس و لارنس<sup>۱</sup>، ۲۰۱۹). یورلند «موضوع» یک مدرک را «توانایی‌های بالقوه اطلاعاتی و معرفتی آن مدرک می‌داند که این قابلیت را دارد که به کاربر اطلاعات بدهد و توسعه دانش را به پیش براند». او همچنین «موضوع» را مترادف با «مبحث» می‌داند و درون‌مایه متن را ایده محوری و اصلی آن می‌داند (یورلند، ۲۰۱۸ ب)

### تحلیل موضوعی

در حوزه بازیابی اطلاعات - و نه فقط حوزه علم اطلاعات و دانش‌شناسی - «موضوع» از نقاط دسترسی است که کار جستجو و بازیابی اطلاعات توسط کاربر با آن صورت می‌گیرد. تحلیل موضوعی منابع اطلاعاتی بخشی از فرایند نمایه‌سازی است (چو و اوبراین<sup>۲</sup>، ۱۹۹۳). در تعبیری ساده و ابتدایی، کارهایی که انجام می‌شود تا محتوای یک مدرک و موقعیت مباحث در آن نمایانده شود، تحلیل موضوعی<sup>۳</sup> یا نمایه‌سازی موضوعی<sup>۴</sup> آن مدرک است که بخشی از فرایند توصیف مدرک و تولید فراداده کتاب‌شناختی است و فهرست‌نویسی و نمایه‌سازی نامیده می‌شوند. در زبان فارسی، فتاحی (۱۳۸۵) فهرست‌نویسی موضوعی را بررسی آثار از نظر محتوای آن‌ها، تشخیص موضوع یا موضوع‌های مناسب برای آن‌ها و بیان موضوع‌ها در قالب اصطلاح‌ها و عبارت‌هایی که به بهترین وجه نشان‌دهنده محتوا باشد می‌داند و از نظر او اهمیت و ظرافت این فرایند از آن روست که به توانایی تحلیل محتوا، تخصص نسبی موضوعی، مطالعه زیاد، آشنایی با آثار مختلف، دانش زبان‌شناسی و جز آن نیاز دارد. سلطانی و راستین (۱۳۷۹، ص. ۲۴۲) فهرست موضوعی را فهرستی تعریف کرده‌اند که بر مبنای موضوع تهیه شده باشد، خواه الفبایی و یا رده‌ای، به نحوی که خواننده را از طریق موضوع به مدرک برساند. پورا احمد و حسنی (۱۳۸۷، ص. ۲۱) نیز از ارائه اطلاعات تحلیل منبع سخن گفته‌اند که بخشی از این اطلاعات تحلیلی اصطلاح‌های مربوط به فهرست‌نویسی موضوعی است و این که فهرست‌نویسی موضوعی با توجه به تحلیل محتوا و مفاهیم موجود در اثر انجام می‌شود.

در محیط اینترنت نیز همین عمل بر محتواهای موجود در اینترنت توسط جویشرها که منابع اینترنت را جستجو و اطلاعات آن‌ها را قابل بازیابی می‌کنند، صورت می‌گیرد؛ با این تفاوت که جز در معدود سال‌های

1. Margolis & Laurence  
2. Chu & O'Brien  
3. Subject analysis  
4. Subject indexing

اولیه پیدایش اینترنت، این عملکرد ماشینی است. در تعیین موضوعات، به ویژه در متون قدیم‌تر علم اطلاعات و دانش‌شناسی که هنوز نمایه‌سازی به صورت دستی وجه غالبی داشت، توصیه بر این بوده که گرایش اصلی یا هدف مشخص کتاب و منظور نویسنده از نوشتن آن در نظر گرفته شود (برویک سایر<sup>۱</sup>، ۱۹۵۴، ص. ۱۷۹) و در صورتی که موضوع کتاب نمایان نیست، باید آن را به طور دقیقی بررسی کرد (بلیس<sup>۲</sup>، ۱۹۳۹، ص. ۱۱۵-۱۱۸) ولی چگونگی این بررسی دقیق توضیح داده نشده است. هاچینز<sup>۳</sup> اشاره می‌کند که در حوزه علم اطلاعات و دانش‌شناسی مطالب فراوانی درباره ساخت زبان‌های نمایه، نظام‌های رده‌بندی و اصول آن، صورت‌بندی صحیح مداخل نمایه‌ای، ارزیابی نمایه‌ها و سیستم‌های اطلاعاتی وجود دارد، ولی پژوهش‌های اندکی درباره این که یک نمایه‌ساز چگونه موضوع یک مدرک را تعیین می‌کند و چگونه درمی‌یابد که یک مدرک درباره چیست، وجود دارد (هاچینز، ۱۹۷۸، ص. ۱۷۲).

می<sup>۴</sup> با دیدگاهی مبتنی بر فلسفه زبان و با تأکید بر نقش محوری متن و مدرک، معتقد است که نمایه‌سازی موضوعی عمل پیچیده تفسیر و انتخاب توسط نمایه‌ساز است و باید با استفاده از روش‌های پژوهشی که متون و نحوه استفاده از آن‌ها را بررسی می‌کنند، مورد مطالعه قرار گیرد (می، ۲۰۰۰). وی با رویکرد تحلیل حوزه<sup>۵</sup> معتقد است که مفهوم تحلیل باید از حد تحلیل متن فراتر رفته و حتی پیش از آن، موقعیت<sup>۶</sup> (اهداف و مقاصد سازمان‌هایی که نمایه‌سازی برای آن‌ها صورت می‌گیرد) و نیازهای اطلاعاتی کاربران بالقوه نمایه را نیز دربرگیرد (می، ۲۰۰۵)، ولی از چالش‌های تحلیل حوزه، از جمله این که چه حد و مرزی میان وفاداری به متن مورد تحلیل و تحلیل حوزه قائل است و چگونه این حد و مرز تعیین می‌شود، بحثی به میان نمی‌آورد. فارادین<sup>۷</sup> اهمیت تمایز میان مفاهیم محسوس<sup>۸</sup> و انتزاعی<sup>۹</sup> را یادآور می‌شود و معتقد است که «دربارگی» یک منبع را تنها یک شخص (معمولاً نمایه‌ساز یا فهرست‌نویس) می‌تواند تعیین کند و اشخاص متعدد ممکن است معانی متعددی از یک مدرک دریافت کنند ولی تنها یک معناست که می‌تواند نمایه شود (فارادین، ۱۹۸۰).

در آثاری که حاوی قواعد برای توصیف مدرک - و از جمله تحلیل موضوعی - هستند، برای شناسایی موضوعات مدرک، علاوه بر بخش‌هایی از هر مدرک می‌توان مفاهیم کلیدی را نیز در آن‌ها یافت (مانند مقدمه

- 
1. Berwick Sawyer
  2. Bliss
  3. Hutchins
  4. Mai
  5. Domain-centred approach
  6. Context
  7. Farradane
  8. Concret
  9. Abstract

و فهرست‌مندرجات و غیره). بدین منظور مراجعه به منابع بیرون از خود مدرک (مانند نقدها، معرفی منابع و متخصصان موضوعی) برای دریافت موضوعات توصیه شده است (دیویی<sup>۱</sup>، ۲۰۱۱، مقدمه؛ رینولدز<sup>۲</sup>، ۱۹۸۹). البته این موضوع نشان‌دهنده آن است که نمایه‌ساز جامع‌المعلومات نیست و نمی‌تواند اشراف بر همه موضوعات مطرح در منابع در حوزه‌های مختلف علوم داشته باشد. هاچینز نیز بعضی ویژگی‌های ساختاری مدرک را ذکر می‌کند که فهرست‌نویس باید به آن‌ها توجه کند تا ایده‌ای از این که مدرک درباره چیست را شکل دهد، مانند هم‌وقوعی، کلمات کلیدی، راهنماهای ساختاریافته، چکیده‌ها و تیترا فصل‌ها (هاچینز، ۱۹۷۵، بخش‌های ۳، ۷-۴، ۶).

اندرسن<sup>۳</sup> (۱۳۸۵) معتقد است مقرّ دانش، ذهن انسان است و در بستر علوم کتابداری و اطلاع‌رسانی، سازماندهی دانش (که اغلب سازماندهی اطلاعات نامیده می‌شود) در داخل ذهن یا مغز صورت می‌گیرد. وی معتقد است در مورد چگونگی توصیف محتوا در میان بخش اعظم جامعه کتابداری و اطلاع‌رسانی، برخلاف توصیف ظاهری، هیچ اجماعی وجود ندارد و بسیاری از روش‌ها در این زمینه با هم رقابت کرده یا یکدیگر را کامل می‌کنند. اندرسن ضمن تفکیک دو رویکرد بنیادین در نمایه‌سازی-تحلیل ذهنی توسط انسان و الگوریتم رایانه‌ای-نمایه‌سازی توسط انسان را نمایه‌سازی ذهنی می‌داند که معمولاً به اشتباه نمایه‌سازی دستی نامیده می‌شود. این دو شیوه تحلیل پیام، به منظور تقویت یکدیگر و رفع متقابل کاستی‌ها هر روز بیش از پیش با هم ترکیب می‌شوند. از نظر او توصیف فرایند نمایه‌سازی توسط انسان بسیار مشکل است. روالی که انسان برای درک و شناسایی موضوعات، ویژگی‌های مهم و معنی‌دار و کاربردهای احتمالی متون نوشتاری، تصاویر یا دیگر انواع پیام‌ها به کار می‌گیرد، ریشه در فرایندهای ادراکی ذهن و مغز دارند. این فرایندها هنوز به خوبی درک نشده‌اند و مرکز توجه حوزه‌هایی نظیر روان‌شناسی‌شناختی، علوم شناختی و هوش مصنوعی هستند. او نیز دو مرحله ذهنی در نمایه‌سازی توسط انسان ذکر می‌کند: (۱) شناسایی موضوع، جنبه‌ها یا کاربردها و (۲) توصیف این موضوعات در قالب واژه‌های مناسب برای بازیابی و معتقد است که نمایه‌سازها، خواه متخصص و باتجربه و خواه مبتدی و بی‌تجربه، نمایه‌سازی را با شیوه‌های بسیار گوناگون اجرا می‌کنند و استفاده از واژگان کنترل شده تنها کمی از این تنوع و تغییرپذیری می‌کاهد (اندرسن، ۱۳۸۵).

نمایه‌سازی را می‌توان اعم از تحلیل موضوعی در نظر گرفت زیرا کل فرایند مربوط به نمایه‌سازی موضوعی را تا تعیین کلیدواژه‌ها و افزوده شدن آن‌ها به نمایه و افزوده شدن نمایه موضوعی به فراداده کتاب‌شناختی در بانک اطلاعاتی دربرمی‌گیرد. در تعریف ایزو نیز چنین تفکیکی را می‌توان دید. در *استاندارد ایزو ۵۹۶۳*، که

1. Dewey  
2. Reynolds  
3. Anderson

برای نمایه‌سازی انسانی تهیه شده است، علاوه بر تعاریف، راهنمایی‌هایی برای نمایه‌سازان در مراحل تحلیل و شناسایی مفاهیم ارائه شده است. در این استاندارد، نمایه به عنوان «فهرست نظام‌مندی از موضوعات با اشاره به مکان آن در یک مدرک یا مجموعه‌ای از مدارک» و نمایه‌سازی «عمل توصیف یا شناسایی موضوعات مندرج در یک مدرک» دانسته شده و توصیف موجودیت فیزیکی خارج از حیطه نمایه‌سازی در نظر گرفته شده است. ایزو ۵۱۲۷ نمایه‌سازی را «تحلیل فکری موضوع یک مدرک برای شناسایی مفاهیم بازنمایی شده در آن و تخصیص اصطلاحات نمایه‌ای مربوط برای این که اطلاعات قابل بازیابی شود»، نمایه‌سازی موضوعی را «نمایه‌سازی که در آن تحلیل موضوعی محتوای مدارک و توصیف‌گرهای موضوعی مدیریت می‌شوند» و نمایه‌سازی خودکار را «نمایه‌سازی با استفاده برنامه‌های کامپیوتری» تعریف کرده است (ایزو<sup>۱</sup>، ۲۰۱۷). در همین استاندارد تحلیل موضوعی زیرمجموعه تحلیل محتوا و «تحلیل یک مدرک برای تعیین مؤلفه‌های محتوای آن و روابط میان آن‌ها» تعریف شده است (همان، شماره ۳.۸.۱.۱). تحلیل موضوعی دقیقاً مترادف تحلیل اطلاعات قلمداد شده است (ویکری، ۲۰۰۴، ص. ۱۱۹). هالی و جودری تعیین «دربارگی» مدرک را معادل تحلیل مفهومی در نظر گرفته‌اند (هالی و جودری، ۲۰۲۰).

در علم اطلاعات و دانش‌شناسی فرایند تحلیل موضوعی، به عنوان یکی از مهم‌ترین مراحل سازماندهی اطلاعات (اعم از فهرست‌نویسی، نمایه‌سازی، چکیده‌نویسی، تهیه نمایه انتهای کتاب)، شامل دو مرحله اصلی تحلیل مفهومی و شناسایی مفاهیم<sup>۲</sup> و ترجمه مفاهیم به زبان نمایه<sup>۳</sup> است که در هر دو این مراحل «موضوع» یا «دربارگی» مدرک مسئله‌ای اصلی است (می، ۲۰۰۰). ترجمه مفاهیم به زبان نمایه مستلزم فهم این است که مدرک درباره چه چیزهایی نوشته شده است و بعد این ادراک و شناسایی به یک زبان موضوعی مصنوعی، مانند طرح‌های رده‌بندی و واژگان کنترل شده استاندارد، مانند یک فهرست از سر عنوان‌های موضوعی یا یک اصطلاح‌نامه، برگردانده می‌شوند. ویلسون چهار روش برای تحلیل مفهومی تعیین کرده است: ۱. تشخیص مقاصد نویسنده از نگارش اثر؛ ۲. وزن دهی به اهمیت نسبی یا برتری بعضی از عناصر در اثر؛ ۳. شمارش اشاره‌هایی که به ایده‌ها و عناصر مختلف می‌شود با رویکردی تجربی؛ و ۴. انتخاب درون‌مایه یا عناصری که اثر را به عنوان یک کل نگه می‌دارد (و در واقع انتخاب آنچه باید نگه داشته شود و کنار گذاشتن آنچه اهمیت زیادی ندارد). سومین مورد ذکر شده توسط ویلسون رویکردی عینی و نزدیک به روش‌های نمایه‌سازی ماشینی دارد، چرا که او کمیت و تکرار واژگان و مفاهیم در مدرک را معیاری برای درباره بودن به کار می‌برد و نمایه‌ساز را به جستجوی عبارات یا مفاهیمی که مکرراً در متن دیده می‌شوند، توصیه می‌کند و صریحاً اظهار

1. International Organization for Standardization (ISO)  
2. Conceptual analysis  
3. Translation

می‌کند که یکی از روش‌های این که مطلبی در یک نوشته دارای غلبه و سلطه به نظر برسد، اشاره مداوم به آن است. او کاربرد هر چهار روش را با هم توصیه می‌کند؛ هرچند باز هم چه یک نمایه‌ساز هر چهار روش را به کار ببرد، چه چند نمایه‌ساز، نتایج ممکن است متفاوت باشند (ویلسون، ۱۹۶۸، ص. ۷۸-۸۹)، امری که اشاره‌ای است به همگونی میان چند نمایه‌ساز یا میان نمایه‌سازی یک نفر از اثر واحد در چند زمان مختلف<sup>۱</sup>. همچنین او «در پی یک موضوع واحد و مشخص در اثر بودن» را بی‌معنا دانسته، اذعان می‌دارد که حقیقت تحلیل مفهومی ناشناخته است (ویلسون، ۱۹۶۸، ص. ۷۰-۷۱). لنگریج<sup>۲</sup> قائل به سه مرحله در فرایند تحلیل مفهومی است:

۱. پوشش<sup>۳</sup> مدرک مانند عنوان، عنوان فرعی، اطلاعات پدیدآور، فهرست مندرجات و عناوین فصول و از این قبیل؛

۲. تحلیل مدرک از حیث رشته علمی، تعیین مباحث، تعیین ماهیت مدرک مانند: زبان، نوع ادبی و غیره؛

۳. تهیه خلاصه در زبان طبیعی (لنگریج، ۱۹۸۹، ص. ۵۵-۵۷، ۱۳۶).

لنکستر تحلیل مفهومی را چیزی بیش از شناسایی درون‌مایه‌های مورد بحث قرار گرفته، یا به عبارت دیگر، بازنمایی شده در مدرک نمی‌داند (لنکستر، ۲۰۰۳، ص. ۱۵). جودری و تیلور با اندکی تفاوت تقریباً همین مراحل را در بررسی منابع پیشنهاد می‌کنند و درباره رویکرد آن‌ها در مورد چگونگی تعیین مهم‌ترین موضوعات چیزی نمی‌گویند. ویژگی رویکرد آن‌ها علاوه بر بررسی ویژگی‌های کتاب‌شناختی و فیزیکی مدرک، تحلیل محتوای اثر برای شناسایی موضوعات و مباحث و در نظر گرفتن چهریزه‌هایی مانند: انواع نگارشی، رشته علمی، زمان و اسامی از هر نوع را توصیه می‌کنند که برای مرحله ترجمه موضوعات به زبان نمایه و رده‌بندی اهمیت پیدا می‌کنند (جودری و تیلور، ۲۰۱۸، ص. ۵۶-۴۶۲) که به نظر می‌رسد تلفیق رویکردهای پیشین موجود در ادبیات این حوزه باشد. آن‌ها نیز معتقدند یک فرد نمایه‌ساز ممکن است با کاربرد روش‌های مختلف برای تعیین موضوع اثر به نتایج مختلفی برسد، همچنان که چندین فرد نیز که یک روش را به کار می‌برند ممکن است به نتایج مختلفی برسند (جودری و تیلور، ۲۰۱۸، ص. ۴۵۵).

در استاندارد معیارهای پایه برای نمایه‌ها، تحلیل مدارک به سه نوع تقسیم شده است: الف. تحلیل فکری توسط انسان و شناسایی موضوعات و مفاهیم بیان‌شده و/یا ویژگی‌های آشکار؛ ب. الگوریتم‌های کامپیوتری طراحی شده برای شناسایی اصطلاحات، عبارات و ویژگی‌های مفید؛ ج. ترکیب تحلیل انسانی و تحلیل مبتنی

1. Inter/ intra indexer consistency

2. Langridge

3. Scan

بر کامپیوتر (مؤسسه استاندارد ملی آمریکا<sup>۱</sup>، ۱۹۹۷، ص. ۹-۱۰). در فرایند تحلیل مفهومی منابع اطلاعاتی کتابخانه‌ای هر چیزی وارد نمی‌شود. طبق قواعد فهرست‌نویسی موضوعی کتابخانه کنگره آمریکا فقط مباحثی که بیست درصد یا بیشتر از منبع درباره آن‌هاست در موضوعات وارد می‌شوند و این با هدف بالا بردن میزان ربط در بازیابی نیز هست (کتابخانه کنگره<sup>۲</sup>، ۲۰۲۰).

از بحث‌های مهم در تحلیل مفهومی مسئله عینی بودن و ذهنی بودن تحلیل است که ابعاد متکثری دارد و در زمان حاضر آنچه در ادبیات علم اطلاعات و دانش‌شناسی مورد توافق به نظر می‌رسد این است که عینیت ایده‌آل هم از لحاظ ارزش و هم از حیث امکان تحقق زیر سؤال رفته است؛ به ویژه این که تحلیل مفهومی اطلاعات توسط شخص نمایه‌ساز عملی است که در ذهن صورت می‌گیرد و به شدت ذهنی و درونی شده است؛ اما آگاهی از این واقعیت و علم به تفاوت‌های تحلیلی افراد اهمیت زیادی دارد. خودآگاهی، ذهن باز، اندیشه‌ورزی و ملاحظه دیدگاه‌های دیگران از ویژگی‌های مطلوب فردی در تعیین موضوع و دربارگی مدارک است (هالی و جودری، ۲۰۲۰). یورلند تحلیل موضوعی را فرایندی تعریف می‌کند که هدف آن توصیف مدارک برای بهبود قابلیت بازیابی آن‌هاست و ممکن است توسط انسان یا ماشین صورت پذیرد و ضمن اشاره به بحث‌های پیرامون مزایا، معایب و حد و مرز میان نمایه‌سازی ماشینی و دستی، نمایه‌سازی ماشینی با استفاده از روش‌های یادگیری ماشین را نیز نوعی نمایه‌سازی انسانی می‌داند، چرا که طراحی الگوریتم‌ها بر عهده متخصصان انسانی است که به منظور صرفه‌جویی در وقت و تسهیل در عملکردهای پیچیده و زمان‌بر محاسباتی در نمایه‌سازی حجم‌های بالای اطلاعات، همان الگوها و مدل‌های فکری نمایه‌سازی را در طراحی الگوریتم‌ها پیاده می‌کنند. او درباره عینیت و ذهنیت در نمایه‌سازی ماشینی و دستی معتقد است که ماشین‌ها و الگوریتم‌ها لزوماً در نمایه‌سازی دارای ویژگی عینیت نیستند و تکرارپذیری در این مورد (تکرار ورودی‌ها در الگوریتم نتایج یکسان می‌دهد) را نباید با عینیت اشتباه گرفت. وی معتقد است الگوریتم‌ها و جویشرها مبتنی بر ذهنیت برنامه‌نویسان آن‌ها و ابزارهای در دسترس هستند. این که کامپیوترها عینی و نمایه‌سازی انسانی ذهنی است، یک افسانه است. همه انواع نمایه‌سازی‌ها بر پایه فرضیات و علائق بنا شده‌اند، اما هرچه قدرت انتخاب بیشتری به کاربران داده شود، احتمال اصلاح استراتژی‌های جستجو برای یافتن مهم‌ترین مدارک بیشتر است (یورلند، ۲۰۱۸ الف).

در انتهای این بخش از مطالب ذکر بخش‌هایی از استاندارد ایزو ۵۹۶۳ که هم اکنون نیز دارای اعتبار

است و بیشتر، توصیه‌هایی درباره نمایه‌سازی و تحلیل موضوعی است، مفید به نظر می‌رسد:

1. American National Standards Institutes  
2. Library of Congress

«... در بررسی مدرک، بهترین کار خواندن کل مدرک است، ولی این عملی نیست و نیاز هم نیست... بخش‌های مهم متن باید به دقت بررسی شوند، به ویژه عنوان، چکیده، فهرست مندرجات، مقدمه و نتیجه، تصاویر، جداول، نمودارها، عباراتی که برجسته نشان داده شده‌اند ... در تعیین تعداد اصطلاحات نمایه که به یک مدرک منتسب می‌شوند، نباید حد و مرز خودسرانه تعیین کرد و این به میزان اطلاعات موجود در مدرک و نیز به نیازهای احتمالی کاربران بستگی دارد. تعیین چنین حد و مرزهایی از دست دادن عینیت و از دست دادن اطلاعاتی ارزشمند در بازایی است. نمایه‌ساز باید بداند که این ابزارها [واژگان کنترل‌شده] ممکن است محدودیت‌هایی در بازنمایی مفاهیم موجود در متن ایجاد کنند ... کیفیت و همگونی نمایه‌سازی به عواملی مانند: الف) صلاحیت‌ها و تخصص نمایه‌ساز و ب) کیفیت ابزارهای نمایه‌سازی بستگی دارد ... داوری ذهنی نمایه‌ساز در شناسایی مفاهیم و انتخاب اصطلاحات نمایه خواه‌ناخواه بر عملکرد سیستم نمایه‌سازی اثر می‌گذارد. همگونی در گروه‌های بزرگ نمایه‌ساز و گروه‌های نمایه‌سازی که در مکان‌های مختلف کار می‌کنند دشوارتر است ... نمایه‌ساز باید دانش مناسبی از حوزه تحت پوشش منابع داشته باشد و بر فنون و زبان نمایه‌سازی نیز مسلط باشد ... کیفیت نمایه‌سازی همچنین به نگهداری زبان نمایه‌سازی مورد استفاده بستگی دارد و باید اصطلاحات جدید وارد و تغییرات ضروری صورت گیرد و به نیازهای جدید کاربران توجه شود. روزآمدسازی منظم این ابزارها ضروری است.» (ایزو، ۱۹۸۵).

در متن فوق، همه نکاتی که از علل گرایش به استفاده از ماشین‌های هوشمند در فرایندهای توصیف و سازماندهی مدارک هستند، مانند دشواری حفظ همگونی، دقت و صحت در کار انسانی - که در مقادیر معتدله اطلاعاتی که محصول فناوری‌های روز است دشوارتر نیز می‌شود - به خوبی قابل ملاحظه است. اگر تحلیل موضوعی و نمایه‌سازی موضوعی ماشینی و به طور کلی استفاده از ماشین‌ها را، شبیه‌سازی عملکردهای انسانی بدانیم که طراحی آن‌ها توسط هوش انسان و با هدف کوتاه کردن زمان و آسان کردن اجرای فرایندها با استفاده از ابزار ماشینی صورت می‌گیرد، افق دیگری پیش رو گشوده می‌شود که در آن عملکرد ماشینی نه صرفاً مکانیکی، بلکه تقلیدی از هوش انسان و زیر نظر انسان است و می‌تواند برای پردازش مقادیر بزرگ داده کارآمدی، صرفه‌جویی در منابع و سرعت عمل بالاتری را به همراه داشته باشد. خواه‌ناخواه، چه در سیستم دستی و چه در سیستم ماشینی، این فکر و اندیشه انسان و توان تحلیل اوست که رویکرد به متن ایجاد می‌کند تا آن را به گونه‌های متفاوت مورد بررسی و تحلیل قرار دهد، درک کند و آنچه را درک کرده در اختیار دیگران قرار دهد تا به سهولت موضوعات و مفاهیم مورد نظر خود را بیانند. در مجموع می‌توان گفت چگونگی تعیین موضوعات در علوم مربوط به اطلاعات و پردازش آن بسیار مهم است و به نظر می‌رسد به رغم مباحثات فراوان، درک مشترکی از چیستی آن وجود دارد. در هنگام تحلیل موضوعی چه نیاز کاربر را در نظر



بگیریم و چه منظور نویسنده را از نگارش اثر، واقعیت این است که تنها خود متن، یعنی واژگان در نحو جملات برای ایفاد معانی، بهترین مرجع برای تعیین مباحث و موضوعات آن هستند.

### مسائل و راه‌حل‌های ممکن

در مسئله بازیابی موضوعی منابع و کیفیت نقاط دسترسی موضوعی، یا همان کلیدواژه‌های موضوعی روش‌های سنتی توصیف منابع و مدارک در محیط اطلاعاتی سنتی با استفاده از نیروی انسانی بسیار مناسب و کارآمد بوده‌اند و هم اینک نیز مفاهیم منطقی، الگوهای عملی و مبانی عقلانی آن‌ها همچنان در حیطه ذخیره و بازیابی اطلاعات پابرجا و در حال خدمت‌رسانی است. اما، مشکلاتی نیز در این خصوص وجود دارد که شاید بتوان گفت مهم‌ترین آن‌ها رشد انتشارات و افزایش روند دیجیتالی کردن مجموعه‌ها و لزوم سازماندهی به موقع منابع است که ماشینی کردن همه یا بخشی از روندها را به ذهن متبادر می‌سازد. علاوه بر این، مشکلاتی در حوزه سازماندهی و توصیف منابع، به ویژه در کتابخانه‌ها و انبارهای دیجیتالی وجود دارد که نمی‌توان آن‌ها را نادیده انگاشت. به عنوان مثال، در کتابخانه‌های دیجیتالی به علل گوناگون، از جمله: تعدد ابزارهای تحلیل موضوعی و انتساب کلیدواژه (سر عنوان‌های موضوعی، اصطلاح‌نامه‌ها و نمایه‌سازی تمام متن)، جستجو و بازیابی با مسائل عدیده‌ای روبه‌روست و رضایت کاربران را موجب نمی‌شود (علیپورحافظی، ۱۳۹۴).

قابلیت‌های جستجو در نرم‌افزارهای کتابخانه‌های دیجیتالی در ایران خوب ارزیابی شده اما آشفتگی در ابعاد مختلف سازماندهی موجود در این کتابخانه‌ها موجب عدم موفقیت بازیابی است. به عنوان مثال، نوروزی، غلامی و جعفری (۱۳۹۶) ذکر کرده‌اند که استفاده از طرح‌های استاندارد توصیف و تحلیل موضوعی و به‌کارگیری استاندارد واحد برای مبادله اطلاعات بین کتابخانه‌های رتبه پایینی در کتابخانه‌های دیجیتالی داراست. مرادی، حاجی زین العابدینی و دلیلی (۱۳۹۲) اشاره کرده‌اند که از کتابداران متخصص در سازماندهی منابع اطلاعاتی استفاده نمی‌شود و علیپورحافظی، حری، اسفندیاری مقدم و علیپور حافظی (۱۳۹۲) معتقدند که کتابداران به ثبت اطلاعات در بخش محدودی از فیلدهای اطلاعاتی آن هم به صورت سلیقه‌ای اقدام می‌کنند که در بسیاری از موارد برای بازیابی اطلاعات کافی نیستند. علاوه بر این، ناسازگاری میان عبارات نمایه‌ای انتساب یافته توسط نمایه‌سازهای مختلف برای مدرک واحد و عبارات انتساب یافته توسط نمایه‌ساز واحد در زمان‌های مختلف و تأثیر حجم کار بالا و زمان اندک در کاهش دقت و افزایش خطا در نمایه‌سازی توسط انسان، از دیگر معضلات در سازماندهی و پردازش این منابع است که پوکورنی (۲۰۱۸) و خوشیان و میرزاییان (۱۳۹۹) نیز بر این نکته تأکید داشته‌اند. این همان بحث مهم همگونی در تحلیل موضوعی است، بدین معنا که اشخاص، از جمله نمایه‌سازان موضوعی، برداشت‌های متفاوتی از موضوعات

مدارک دارند و حتی یک نمایه‌ساز در زمان‌های مختلف یک مدرک واحد را به صورت یکدست و همگونی نمایه نمی‌کند و تجارب شخصی، علائق فکری، تحصیلات و پیش‌فرض‌های اولیه درباره مدرک مورد نمایه در این خصوص تأثیرگذارند (جودری و تیلور، ۲۰۱۸، ص. ۴۵۴؛ ویلسون، ۱۹۶۸، ۷۸-۸۹؛ فارادین، ۱۹۸۰). برخی حتی نقش گذشت زمان و تغییر مکان را بر نمایه‌ساز واحد مؤثر دانسته‌اند (چان، ۲۰۰۷، ص. ۲۰۸-۲۰۹).

جنبه دیگری که قابل طرح است، مسئله کاربر و رفتار وی در هنگام جستجوی اطلاعات است؛ زیرا کاربران هنگام جستجو از زبان طبیعی استفاده می‌کنند (شهبازی، نوروزی و علیپور حافظی، ۱۳۹۳) و نسبت زیادی از جستجوی موضوعی کاربران با سرعنوان‌های موضوعی همخوانی ندارد (فتاحی و نیک زمان، ۱۳۹۱). دیگر این که اصلاحات و به‌روزرسانی اصطلاحات زبانی در فهرست واژگان موضوعی از پیش آماده، مانند سرعنوان‌های موضوعی و اصطلاح‌نامه‌ها، معمولاً به صورت گذشته‌نگر انجام نمی‌شوند. در واقع هر کاربر «کلکسیون‌ی از سرعنوان‌های موضوعی با ساختاری طبیعی در ذهن خود تشکیل داده که بازخوانی آن با ترکیباتی متفاوت امکان‌پذیر است. آنچه این سرعنوان‌های طبیعی را از سرعنوان‌های ساختاریافته و کنترل‌شده جدا می‌کند وجود ترتیبی سازمان‌یافته از واژه‌ها و اصطلاحات با ترکیب‌های متفاوت است که کاربر با این‌گونه ساختار آشنایی ندارد ... و نمی‌توان گفت که سرعنوان‌های موضوعی مقصود کاربر را به طور کامل با ارائه منابعی مشخص پاسخگو هستند و علت آن هم عدم تغییر سرعنوان‌های موضوعی در گذر زمان است» (لطیفی، زارعی و غفاری، ۱۳۹۴). سپهر و ابادری (۱۳۹۵) نیز به همین دلیل استفاده از فراداده‌های معنایی را که از پردازش زبان طبیعی و فناوری‌های ماشینی به دست می‌آیند، لازم دانسته‌اند.

علاوه بر این که در وضعیت فعلی تولید محتواهای اطلاعاتی متأثر از تحولات فناورانه دیجیتال و غیر آن به سرعت رو به افزایش است، سازماندهی و توصیف منابع با روش‌های دستی مستلزم صرف وقت و هزینه زیاد و عملاً غیرممکن است (حمید، ۲۰۱۶، ص. ۱؛ سون، هو، لی، لی و چی<sup>۱</sup>، ۲۰۲۰)، استانداردهای توصیف کتاب‌شناختی نیز به طور مداوم در شرف پیچیده‌تر شدن و تغییرات هستند، و همین موضوع موجب می‌شود به‌روزرسانی و نگهداری ابزارهای جانبی (مانند اصطلاح‌نامه‌ها و سرعنوان‌های موضوعی و سایر بانک‌های مستند) و تطبیق فرایند توصیف منابع با آن‌ها زمان‌بر و پرهزینه شود. در این شرایط اگر انتظار داشته باشیم که نیروی انسانی محدود منابع بی‌شماری را پردازش کنند، عملاً همواره منابع اطلاعاتی فراوانی را شاهد خواهیم بود که در صف توصیف مانده‌اند و این مانعی در دسترسی به موقع کاربران به منابع مورد نیاز آن‌هاست. چاره کار همگامی با تحولات فناورانه و استفاده از امکانات بالقوه همین تحولات در یافتن راه‌حلهایی است که بتوانند فرایندها را ماشینی کرده، نیروی انسانی را در مسیر کنترل کیفیت، تکامل و

بهینه کردن ماشین‌ها به کار بگیرند (شورت<sup>۱</sup>، ۲۰۱۹؛ یونگر<sup>۲</sup>، ۲۰۱۸؛ تچوا<sup>۳</sup>، ۲۰۱۹). این راه‌حل‌ها در حوزه علوم کامپیوتر و پردازش زبان طبیعی حدود دو دهه است که توجه پژوهشگران را به خود جلب کرده است. البته ابتکارها و آزمون‌ها برای کاربرد ماشین، چه به صورت کمکی و چه به صورت اصلی، در فرایندهای توصیف، ذخیره و جستجوی ماشینی منابع اطلاعاتی پدیده تازه‌ای نبوده و از دهه ۱۹۵۰ میلادی با ساخت کامپیوترها آغاز شده است (گیل لیوا<sup>۴</sup>، ۲۰۱۷)؛ اما، در دو دهه گذشته روند استفاده از روش‌های علوم کامپیوتر و هوش مصنوعی در حوزه‌های مختلف، از جمله داده‌کاوی و متن‌کاوی، رو به فزونی نهاده است. استخراج کلیدواژه، هرچند در حوزه کتابداری و علم اطلاعات و اطلاع‌رسانی معمولاً به صورت دستی یا با کمک ماشین انجام می‌شده است. امروزه از مطالعات و فعالیت‌های مهم مربوط به متن‌کاوی و پردازش زبان طبیعی است. در واقع، بدون کاربرد فنون متن‌کاوی جستجو و یافتن مطالب مورد نظر در مقادیر عظیم داده‌هایی که به صورت متن و در قالب نشریات و کتاب‌های الکترونیکی، صفحات وب، نامه‌های الکترونیکی و شبکه‌های اجتماعی گوناگون ارائه می‌شوند، وقت‌گیر و گیج‌کننده است (لازمی، ابراهیم پور، کومله و نوروزی<sup>۵</sup>، ۲۰۱۹).

متن‌کاوی فرایند استخراج الگوهای مفید و جالب، استخراج دانش و استخراج اطلاعات با ربط از اسناد و مدارک متنی با استفاده از کامپیوتر و محاسبات است و آن را از انواع داده‌کاوی می‌دانند (صدیقی و شرام، ۲۰۱۵). متن‌کاوی با عنوان‌هایی از قبیل «تجزیه و تحلیل هوشمند متن»<sup>۶</sup>، «کاوش داده‌های متنی»<sup>۷</sup>، «کشف دانش از متن»<sup>۸</sup> و «تحلیل متن»<sup>۹</sup> شناخته شده و حوزه‌ای میان‌رشته‌ای است (وانگچاسووات<sup>۱۰</sup>، ۲۰۱۷). استخراج کلیدواژه همچنین حوزه‌ای بسیار مهم و مطالعه شده در پردازش زبان طبیعی و حوزه‌ای از علوم کامپیوتر و به طور مشخص از زیرشاخه‌های متن‌کاوی است که با مجموعه‌ای از فنون و عملیاتی مانند طراحی الگوریتم‌ها و بازنمایی‌های محاسباتی، زبان انسان را برای کامپیوتر قابل دستیابی می‌کند تا بتواند آن را تحلیل و پردازش کرده، معنا و دانش‌هایی را که تا پیش از این ناشناخته مانده بوده‌اند، استخراج کند (منگ و دیگران<sup>۱۱</sup>، ۲۰۱۸). حتی گفته شده که متن‌کاوی کاری فراتر از ایجاد دسترسی به اطلاعات بوده و در اصل

- 
1. Short
  2. Junger
  3. Tchoua
  4. Gil-Leiva
  5. Lazemi, Ebrahimpour-Komleh & Noroozi
  6. Intelligent Text Analysis
  7. Text Data Mining
  8. Knowledge Discovery in Text
  9. Text Analysis
  10. Wongchasuwat
  11. Meng

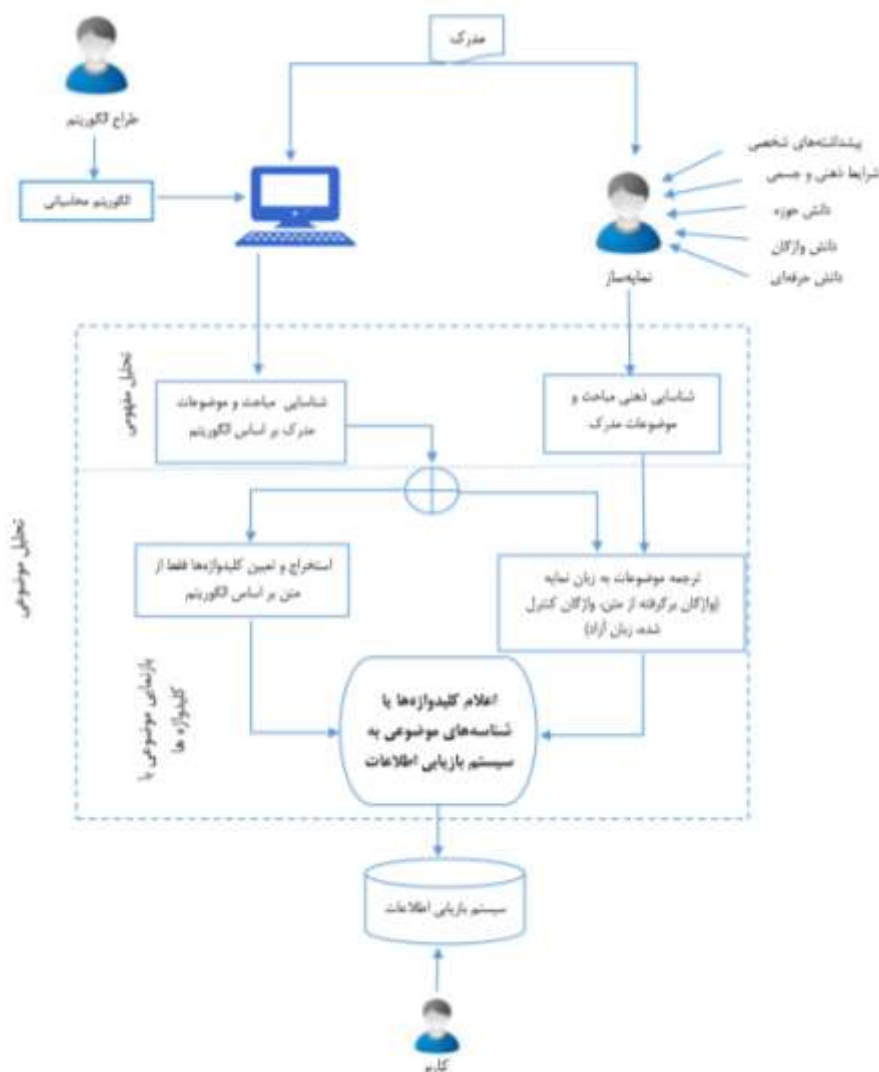
یاری رساندن به کاربر در تحلیل و درک اطلاعات برای تصمیم‌گیری است (بانو و چیترا<sup>۱</sup>، ۲۰۱۵؛ الهیاری<sup>۲</sup> و دیگران، ۲۰۱۷).

در پردازش زبان طبیعی مدل‌های مختلفی برای استخراج کلیدواژه وجود دارد که دسته‌ای از آن‌ها مدل‌سازی موضوعی<sup>۳</sup> نام دارند و روش‌هایی را برای سازماندهی، درک، جستجو و خلاصه‌سازی خودکار منابع الکترونیکی فراهم می‌آورند. این کار با شناسایی موضوعات مدارک و کلمات درون هر موضوع صورت می‌گیرد و ضمن استخراج کلیدواژه‌های موضوعی، می‌توان مدارک مجموعه را به موضوعات مختلف مرتبط کرد (هورتادو، ۲۰۱۶). استفاده از روش‌های گفته شده روند رو به رشدی در اقدامات و مطالعات مربوط به تحلیل، پردازش و سازماندهی متون دارد و می‌توان آن‌ها را راه‌حلی نسبتاً نوظهور دانست که در حیطه‌های تحلیل موضوعی مربوط به توصیف منابع در علم اطلاعات و دانش‌شناسی نیز کاربرد دارند.

کتابخانه‌های بسیاری در جهان، سیاست‌گذاری جدیدی در توصیف مجموعه‌هایشان در پیش گرفته‌اند که مبتنی بر ایجاد نقاط دسترسی موضوعی با استفاده از روش‌های ماشینی، به عنوان روش اصلی (و نه تنها روش) برای انواع منابع، از جمله کتاب‌های الکترونیکی - و حتی برای منابع چاپی - است (یونگر<sup>۴</sup>، ۲۰۱۸). استخراج کلیدواژه با استفاده از روش‌های کارآمد ماشینی متناسب با ویژگی‌های هر دسته از منابع اطلاعاتی می‌تواند کمکی در کوتاه کردن مسیر دسترسی به محتوای آن‌ها باشد؛ اما ارزیابی کارآمدی این دسته از کلیدواژه‌ها به تنهایی یا در مقایسه با کلیدواژه‌های تولید انسان به گونه‌ای که علاوه بر معیارهای سنتی دقت و سنج اف<sup>۵</sup> در ارزیابی کارآمد، از دیدگاه کاربران در ارزیابی مطلوب و مؤثر ارزیابی شوند، ضروری است. در شکل ۱ مفاهیم مورد بررسی در فرایند نمایه‌سازی انسانی و ماشینی مدل شده‌اند.

---

1. Banu & Chitra  
2. Allahyari  
3. Topic modeling algorithms  
4. Junger  
5. F-measure



شکل ۱. مدل مفهومی نمایه‌سازی انسانی و ماشینی

### جمع‌بندی

هر منبع اطلاعاتی موجودیتی متشکل از اطلاعات و حامل آن است. نوع حامل‌های اطلاعاتی تحت تأثیر مظاهر تمدن، توسعه و فناوری‌های روز در هر دوران، به نوبه خود بر شیوه‌های تولید محتوا و سازماندهی آن تأثیر می‌گذارد. منابع اطلاعاتی از حیث سازماندهی برای بازیابی مؤثر و سریع، دستمایه و مصالح کار و تجربه برای دو حوزه علم اطلاعات و دانش‌شناسی از یک سو و علوم کامپیوتر، پردازش زبان طبیعی و یادگیری ماشین از سوی دیگر هستند که با روش‌های متفاوت اهداف مشترکی را در بازیابی دنبال می‌کنند که این دومی در پی شبیه‌سازی توانایی‌ها و عملکرد ذهن نمایه‌ساز در فرایندهای مختلف توصیف منابع اطلاعاتی است و وابستگی زیادی به عناصر تشکیل دهنده منبع اطلاعاتی در توصیف دارد. تحلیل موضوعی و بازنمودن موضوعات و

مباحث موجود در منابع اطلاعاتی در قالب واژه‌هایی که در یک نظام ذخیره و بازیابی اطلاعات رابط میان متن و کاربر قرار می‌گیرند، بخش مهمی از فرایند توصیف به شمار می‌رود. در مؤسسات و نهادهای اطلاع‌رسانی که روزانه با مقادیر فراوانی از انواع منابع اطلاعاتی برای توصیف و پردازش مواجه هستند، خودکار کردن همه یا بخش‌هایی از این فرایند می‌تواند نقش مهمی در ارتقاء سطح بهره‌وری از منابع انسانی و مالی و کوتاه کردن زمان توصیف منابع داشته باشد. نتیجه این سرعت و بهره‌وری ارائه خدمات بهتر و دسترسی سریع‌تر او به منابع اطلاعاتی مورد نیازش است. روش‌های یادگیری ماشین، متن‌کاوی، تحلیل متن و پردازش زبان طبیعی، قابلیت کاربرد در حوزه تحلیل موضوعی و استخراج خودکار کلیدواژگان موضوعی را دارا هستند و کتابخانه‌ها و مراکز اطلاع‌رسانی می‌توانند از روش‌ها و ابزارهای مختلف و مناسبی که پردازش زبان طبیعی و یادگیری ماشین در اختیار قرار می‌دهند، برای طراحی و توسعه سیستم‌های ماشینی نمایه‌سازی و استخراج یا انتساب کلیدواژه‌های موضوعی - و در چشم‌اندازی بزرگ‌تر رده‌بندی خودکار - با هدف نهایی و کلی بهره‌وری بیشتر از منابع (مالی، انسانی و زمان) و کوتاه کردن مسیر دسترسی موضوعی به منابع برای کاربران و تسهیل و تسریع فرایند توصیف استفاده کنند. قدر مسلم در رقابت الگوریتم‌های ماشینی با ذهن انسان در تحلیل و تشخیص موضوعات متن، ذهن انسان پیشی می‌گیرد؛ حال این انسان چه نمایه‌سازی باشد که به زبان خود بیان کند که متن درباره چیست و چه طراحی باشد که بتواند ماشینی طراحی کند که گام‌های محاسباتی ذهن او را تقلید و موجب صرفه‌جویی در زمان و منابع شود. اما همکاری این دو حوزه از علم، به دور از شتابزدگی در جایگزینی روندها و رویه‌های مرسوم، رشد و توسعه متقابل در هر دو حوزه و در نهایت ارائه خدمات بازیابی بهینه را به همراه خواهد داشت.

**یادداشت:** این مقاله مستخرج از رساله است.

## سپاسگزاری

از معاونت محترم پژوهشی دانشگاه خوارزمی به خاطر حمایت معنوی در اجرای پژوهش حاضر سپاسگزاری می‌شود. همچنین نگارندگان بر خود لازم می‌دانند از داوران محترم به خاطر مطالعه متن مقاله حاضر و ارائه نظرهای ارزشمند سپاسگزاری نمایند.

## منابع

اندرسن، جیمز دی (۱۳۸۵). سازماندهی دانش. ترجمه حمیدرضا جمالی مهمویی. در *دایره‌المعارف کتابداری و اطلاع‌رسانی*، ج ۲. سرویراستار ابراهیم افشار، با همکاری سودابه نوذری. بازیابی از: <http://portal.nlai.ir/daka/Wiki%20Pages/%D8%B3%D8%A7%D8%B2%D9%85%D8%A7%D9%86%D8%AF%D9%87%D9%8A%20.aspx>

پوراحمد، علی اکبر؛ حسنی، محمدرضا (۱۳۸۷). *فهرست‌نویسی موضوعی: اصول و روش‌ها*. ویراسته رحمت الله فتاحی. تهران: کتابدار.

پورتر، ابوت اچ (۱۳۹۷). *سواد روایت*. ترجمه رویا پورآدر و نیما اشرفی. تهران: اطراف.

خداپرستی، فرج الله (۱۳۷۶). *فرهنگ جامع واژگان مترادف و متضاد زبان فارسی*. شیراز: دانشنامه فارسی. بازیابی از: <https://vajehyab.com/motaradef/%D9%85%D9%88%D8%B6%D9%88%D8%B9>  
خوشیان، ناهید؛ میرزاییان، وحیدرضا (۱۳۹۹). *پراگماتیک عملکردهای پردازش زبان طبیعی در حوزه علوم کتابداری و اطلاع رسانی*. *دانش‌شناسی*، ۶(۲۳)، ۱۱۷-۱۵۱.

دهخدا، علی اکبر (۱۳۹۹). *لغت‌نامه دهخدا* (نسخه دیجیتال) براساس نسخه فیزیکی ۱۵ جلدی انتشار سال ۱۳۷۷. موسسه لغت‌نامه دهخدا و مرکز بین‌المللی آموزش زبان فارسی دانشگاه تهران. بازیابی از: <https://dehkhoda.ut.ac.ir/fa/dictionary/309078/%D9%85%D9%88%D8%B6%D9%88%D8%B9>

رازی، فریده (۱۳۷۲). *فرهنگ واژه‌های فارسی سره: برای واژه‌های عربی در فارسی معاصر*. تهران: نشر مرکز. بازیابی از: <https://vajehyab.com/sareh/%D9%85%D9%88%D8%B6%D9%88%D8%B9>  
سپهر، فرشته؛ ابذری، زهرا (۱۳۹۵)، ۸ و ۹ اردیبهشت). *امکان سنجی استفاده از هستی‌شناسی در کتابخانه‌های دیجیتالی ایران*. مقاله ارائه شده در دومین کنفرانس بین‌المللی وب پژوهی، تهران.

سلطانی، پوری (۱۳۵۱). *رده‌بندی کولن*. *خبرنامه انجمن کتابداران ایران*، ۵(۳)، ۱۱۵-۱۲۰.

سلطانی، پوری (۱۳۸۱). *سرعنوان‌های موضوعی فارسی* (کتاب). در عباس حری و نرگس نشاط (ویراستاران)، *دایره‌المعارف کتابداری و اطلاع‌رسانی* (ویراست نخست). تهران: سازمان اسناد و کتابخانه ملی ایران. بازیابی از: [http://portal.nlai.ir/daka/Wiki%20Pages/%D8%B3%D8%B1%D8%B9%D9%86%D9%88%D8%A7%D9%86%20%D9%81%D8%A7%D8%B1%D8%83%D8%AA%D8%A7%D8%A8\).aspx](http://portal.nlai.ir/daka/Wiki%20Pages/%D8%B3%D8%B1%D8%B9%D9%86%D9%88%D8%A7%D9%86%20%D9%81%D8%A7%D8%B1%D8%83%D8%AA%D8%A7%D8%A8).aspx)  
سلطانی، پوری؛ راستین، فروردین (۱۳۷۹). *دانشنامه کتابداری و اطلاع‌رسانی*. ویراست ۳. تهران: فرهنگ معاصر.

شهبازی، فرزاد؛ نوروزی، یعقوب؛ علیپور حافظی، مهدی (۱۳۹۳). *بررسی میزان استفاده نرم‌افزارهای کتابخانه دیجیتالی ایران از ویژگی‌های نظام‌های خبره در بازیابی اطلاعات*. *پژوهشنامه پردازش و مدیریت اطلاعات*، ۳۰(۳)، ۸۲۳-۸۵۱. بازیابی از: <http://jipm.irandoc.ac.ir/article-1-2589-fa.html>

علیپورحافظی، حامد؛ حری، عباس؛ اسفندیاری مقدم، علیرضا؛ علیپور حافظی، مهدی (۱۳۹۲). *تحلیل وضعیت سازماندهی منابع اطلاعاتی در کتابخانه‌های دیجیتالی ایران*. *پژوهشنامه پردازش و مدیریت اطلاعات*، ۲۹(۱)، ۵۹-۳۵. بازیابی از: <http://jipm.irandoc.ac.ir/article-1-2589-fa.html>

عمید، حسن (۱۳۸۹). *فرهنگ فارسی عمید*. تهران: آشجع. بازیابی از:

<http://www.vajehyab.com/?q=%D9%85%D9%88%D9%88%D8%B9&f=amid>

فتاحی، رحمت الله (۱۳۸۵). *فهرست‌نویسی*. در *دایره‌المعارف کتابداری و اطلاع‌رسانی*. ج ۲. بازیابی از: <http://portal.nlai.ir/daka/Wiki%20Pages/%D9%81%D9%87%D8%B1%D8%B3%8A.aspx>

فتاحی، رحمت‌الله؛ طاهری، مهدی (۱۳۸۴). *فهرست‌نویسی: اصول روش‌ها*. ویرایش ۴. تهران: کتابدار.

فتاحی، رحمت‌الله؛ نیک زمان، امیر (۱۳۹۱). *تحلیل جستجوهای موضوعی کاربران در فهرست رایانه‌ای به لحاظ نوع و چگونگی همخوانی آن‌ها با سرعنوان‌های موضوعی فارسی* (با استفاده از ثبت داد و گرفت). *پژوهشنامه پردازش و مدیریت اطلاعات*،

۲۸(۱)، ۲۷۱-۲۵۱. بازیابی از: <http://jipm.irandoc.ac.ir/article-1-1802-fa.html>

- لطیفی، سعید؛ زارعی، عاطفه؛ غفاری، سعید (۱۳۹۴). نقش سرعنوان‌های موضوعی فارسی در بازیابی منابع اطلاعاتی کودکان از دیدگاه کتابداران کتابخانه‌های عمومی استان همدان. *نظامها و خدمات اطلاعاتی*، ۴(۳۰۳)، ۸۱-۹۲.  
[https://iss.srbiau.ac.ir/article\\_9720.html?lang=fa](https://iss.srbiau.ac.ir/article_9720.html?lang=fa)
- مرادی، خدیجه؛ حاجی زین العابدینی، محسن؛ لیلی، حمید (۱۳۹۲). داده‌های توصیفی و روابط کتاب‌شناختی در پیشینه‌های کتاب‌شناختی کتابخانه‌های دیجیتال ایران بر اساس استاندارد آر. دی. ای. *فصلنامه مطالعات ملی کتابداری و سازماندهی اطلاعات*، ۲۴(۲)، ۲۲-۴۰.  
[http://nastinfo.nlai.ir/article\\_69\\_3048cadf46d39230306736859f2821cf.pdf](http://nastinfo.nlai.ir/article_69_3048cadf46d39230306736859f2821cf.pdf)  
<https://vajehyab.com/moein/%D9%B9> بازبازی از: معین، محمد. *فرهنگ فارسی*.
- نوروزی، یعقوب؛ غلامی، طاهره؛ جعفری‌فر، نیره (۱۳۹۶). وضعیت کتابخانه‌های دیجیتال ایران بعد از گذشت یک دهه چگونه است؟ *فصلنامه مطالعات ملی کتابداری و سازماندهی اطلاعات*، ۲۸(۴)، ۱۴۷-۱۷۰. بازبازی از:  
[http://nastinfo.nlai.ir/article\\_2195\\_3914aa395fac32e1e99b78effcfdd081.pdf](http://nastinfo.nlai.ir/article_2195_3914aa395fac32e1e99b78effcfdd081.pdf)

## References

- Allahyari, M., Pouriyeh, S., Asefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering, and Extraction techniques. In *Proceedings of KDD Bigdata, Halifax, Canada, August 2017*. 13p.
- American National Standards Institutes. (1997). *Basic Criteria for Indexes*. Bethesda, Maryland: NISO Press.
- Amid, Hassan (2010). Amid Persian dictionary. Tehran: Ashja. Retrieved from <http://www.vajehyab.com/?q=%D9%85%D9%88%D8%D8%B9&f=amid>. (in Persian)
- Anderson, J. D. (2006). Knowledge organization. Translated to Persian by Hamid Reza Jamali Mahmouei. In *Encyclopedia of Library and Information Science*, vol. 2. Afshar, E. (Editor in chief) and Nozari, S. (Co-editor). Retrieved from <http://portal.nlai.ir/daka/Wiki%20Pages/%D8%B3%D8%A7%D9%86%D8%AF%D9%87%D9%8A%20%20%D8%B4.aspx>. (in Persian)
- Arastoopoor, S., & Fattahi, R. (2012). Users' perception of aboutness and ofness in images: An approach towards subject indexing based on Ervin Panofsky's theory and users' view. *12th International ISKO Conference*.
- Banu, G. R., & Chitra, V. K. (2015). A survey of text mining concepts. *International Journal of Innovation in Engineering and Technology (IJJET)*, 5(2), 121-127.
- Berwick Sayer, W. C. (1954). *An introduction to library classification: Theoretical, historical, and practical*. 9<sup>th</sup> ed. London: Grafton.
- Bliss, H. E. (1939). *The organization of knowledge in libraries: And the subject approach to books*. New York: Wilson.
- Borko, H., & Bernier, C. L. (1978). *Indexing: Concepts and methods*. New York: Academic Press Inc.
- Chan, L. M. (2007). *Cataloging and Classification: An Introduction*, 3rd ed. Lanham, MD: Scarecrow Press.



- Chu, C., & O'Brien, A. (1993). Subject analysis: The critical first stage in indexing. *Journal of Information Science*, 19(6), 439-54.
- Cutter, C. A. (1904). *Rules for a Dictionary Catalog*. 4th ed. Washington, DC: Government Printing Office.
- Dehkhoda, Ali Akbar (2020). *Dehkhoda Dictionary* (Digital version) Based on the 15-volume physical version, published in 1998. Dehkhoda Dictionary Institute and International Center for Persian Language Teaching, University of Tehran. Retrieved from <https://dehkhoda.ut.ac.ir/fa/dictionary/309078/%D9%85%D9%88%D8%B9>. (in Persian)
- Dewey, M. (2011). *Dewey Decimal Classification*. 23rd ed. Dublin, OH: OCLC Online Computer Library Center.
- Diao, J. (2018). Evolution of Western Library Catalogs: The Rising Expectations of Users. *International Journal of Librarianship*, 3(1), 78-93.
- Dutta, B., & Dutta, C. (2013). Concept of 'subject' in the context of library and information science from a new angle. *Annals of Library and Information Studies*, 2013, 60(2), 78-87. Retrieved from <http://eprints.rclis.org/19813/>
- Fairthorne, R. A. (1969). Content Analysis, Specification, and Control. *Annual Review of Information Science and Technology*, 4, 73-109.
- Farradane, J. (1980). Relational Indexing, Part II. *Journal of Information Science*, 1(6), 313-14.
- Fattahi, R., & Nikzaman, A. (2012). An Analysis of the Match between the Forms and Vocabulary of Users' Subject Queries with the Persian Subject Headings in a University OPAC. *Iranian Journal of Information Processing and Management*, 28(1), 251-271. Retrieved from <http://jipm.irandoc.ac.ir/article-1-1802-fa.html>. (in Persian)
- Fattahi, R., & Taheri, M. (2005). *Cataloging: Principles and methods*. 4<sup>th</sup> ed. Tehran: Ketabdar. (in Persian)
- Fattahi, R. (2006). Cataloging In *Encyclopedia of Library and Information Science*, vol. 2. Afshar, E. (Editor in chief) and Nozari, S. (Co-editor). Retrieved from <http://portal.nlai.ir/daka/Wiki%20Pages/%D9%81%D9%87%D8%B1%D8%B3%D8%AA%D9%86%D9%88%D9%8A%D8%B3%D9%8A.aspx>. (in Persian)
- Functional requirements for subject authority data (FRSAD): a conceptual model*. (2011). Edited by Marcia Lei Zeng, Maja Žumer and Athena Salaba. München: De Gruyter Saur. (IFLA series on bibliographic control; vol. 43). Retrieved from <http://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf>
- Gil- Leiva, I. (2017). SISA- automatic indexing system for scientific articles: Experiments with location heuristics rules versus TF.IDF rules. *Knowledge Organization*, 44(3), 139-162 .
- Guthrie, L. S. (1992). An Overview of Medieval Library Cataloging, *Cataloging & Classification Quarterly*, 15(3), 93-100. DOI: 10.1300/J104v15n03\_07
- Hamid, F. (2016). *Evaluation techniques and graph-based algorithm for automatic summarization and keyphrase extraction*. (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global database. (UMI No. 10307512)

- Hjørland, B. (2018a). Indexing: Concepts and theory. In *ISKO Encyclopedia of Knowledge Organization*, eds. Birger Hjørland, coed. Claudio Gnoli. Retrieved from <http://www.isko.org/cyclo/indexing>
- Hjørland, B. (2018b). Subject (of documents). *Knowledge Organization*, 44(1), 55-64. Also available in *ISKO Encyclopedia of Knowledge Organization*, ed. Birger Hjørland, coed. Claudio Gnoli. Retrieved from <http://www.isko.org/cyclo/subject>
- Holley R. M., & Joudrey, D. N. (2020). Aboutness and conceptual analysis: A review. *Cataloging & Classification Quarterly*.
- Hutchins, W. J. (1978). The concept of 'aboutness' in subject indexing. *Aslib Proceedings*, 30(5), 172-181. Retrieved from <https://doi.org/10.1108/eb050629>
- Hutchins, W. J. (1975). *Languages of Indexing and Classification*. Herts, England: Peter Peregrinus.
- Goh, R. (2018). Using Named Entity Recognition for Automatic Indexing. Paper presented at the *IFLA WLIC, 2018*, Kuala Lumpur, Malaysia.
- Golube, K., Hagelbach, J., & Ardo, A. (2018). Automatic classification using DDC on the Swedish Union Catalogue. [CEUR-WS.org/vol-2200/paper1.pdf](http://www.ceur-ws.org/vol-2200/paper1.pdf)
- Gopinath, M. (1976). Colon classification. In *Classification in the 1970s: A second look*. edited by Arthur Meltby (rev. ed. Pp. 51-80). London: Clive Bingley.
- Hurtado, J. L. (2016). *Text mining and topic modeling for social and medical decision support*. (Doctoral dissertation). Available from ProQuest Dissertations & Thesis Global database. (UMI No. 10583055)
- IFLA Library Reference Model: A conceptual model for bibliographic information* (2017). Edited by Pat Riva, Patrick Le Boeuf, and Maja Zumer. Retrieved from [https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017\\_rev201712.pdf](https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf)
- International Organization for Standardization (ISO). (1985). *Documentation - Methods for Examining Documents, determining their Subjects and Selecting Indexing Terms (ISO 5963-1985)*. Geneva: International Organization for Standardization.
- International Organization for Standardization (ISO). (2017). *Documentation - Information and documentation — Foundation and vocabulary (ISO 5127-2017)*. Geneva: International Organization for Standardization. Retrieved from <https://www.iso.org/obp/ui/#iso:std:iso:5127:ed-2:v1:en>
- Joudrey, D. N. & Taylor, A. G. (2018). *The Organization of Information*. 4th ed. Santa Barbara, CA: Libraries Unlimited.
- Junger, U. (2018). *Automated first- The subject cataloguing policy of the Deutsche Nationalbibliothek*. Paper presented at: IFLA WLIC 2018- Kuala Lumpur, Malaysia- Transform Libraries, Transform Societies in session 115- Subject Analysis and Access. Retrieved from <http://library.ifla.org/2213/1/115-junger-en.pdf>
- Kaiser, J. (1911). *Systematic indexing*. London: Isaac Pitman & Sons.
- Khodaparasti, Farajollah (1997). *A comprehensive dictionary of persian synonyms & antonyms*. Shiraz: Persian Encyclopedia. Retrieved from <https://vajehyab.com/motaradef/%D9%85%D9%88%D8%B6%D9%88%B9>. (in Persian)

- Khoshian, N., & Mirzaeian, V. (2020). The Most Widely Used Functions of Natural Language Processing in the Field of Library Science and Information Science. *Knowledge Retrieval and Semantic Systems*, 6(23), 117-151. (in Persian)
- Lancaster, F. W. (2003). *Indexing and Abstracting in Theory and Practice*. 3rd ed. Champaign, IL: University of Illinois).
- Lancaster, F. W., Elliker, C., & Harkness, T. (1998). Subject analysis. *Annual Review of Information Science and Technology (ARIST)*, 24, 35-84.
- Latifi, S., zarei, A., & Ghaffari, S. (2015). Studying the role of Persian Subject Headings in retrieving the children information resources from the Perspective of Librarians of Hamadan Province public libraries. *Journal of Information Systems and Services*, 4(4), 81-92. Retrieved from [https://iss.srbiau.ac.ir/article\\_9720.html?lang=en](https://iss.srbiau.ac.ir/article_9720.html?lang=en). (in Persian)
- Lazemi, S., Ebrahimpour-Komleh, H., & Noroozi, N. (2019). PAKE: A supervised approach for Persian automatic keyword extraction using statistical features. *Springer Nature Applied Science*, 1, 1574. DOI: 10.1007/s42452-019-1627-5
- Library of Congress (2020). *H 180: Assigning and Constructing Subject Headings, List of the Subject Heading[s] Manual*. Retrieved from <https://www.loc.gov/aba/publications/FreeSHM/H0180.pdf>
- Mai, J. E. (2000). *The Subject Indexing Process: an investigation of problems in knowledge representation* (Doctoral dissertation). Retrieved from [http://jenserikmai.info/Papers/2000\\_PhDdiss.pdf](http://jenserikmai.info/Papers/2000_PhDdiss.pdf)
- Margolis, E., & Laurence, S. (2019). Concepts, In *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.). Retrieved from <https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=concepts>
- Metcalfe, J. (1973). When is a Subject Not a Subject? In *Towards a Theory of Librarianship*, ed. by C. H. Rawski. 303-338. New York: Scarecrow Press.
- Meng, R., Zhao, S., Han, Sh., He, D., Brusilovski, P., & Chi, Y. (2017). Deep keyphrase generation. *ACL 2017*. DOI: 10.18653/v1/P17-1057
- Miksa, F. (1983). *The Subject in the Dictionary Catalog from Cutter to the Present*. Chicago: ALA.
- Miksa, S. D. (2021). Cataloging Principles and Objectives: History and Development. *Cataloging & Classification Quarterly*, 59(2-3), 97-128.
- Mo'in, M. *Persian Dictionary* [online]. Retrieved from <https://vajehyab.com/moein/%D9%85%D9%88%D8%B6%D9%88%D8%B9/>. (in Persian)
- Moradi, K., Haji Zainul Abedini, M., & Dalil, H. (2013). Descriptive data and bibliographic relationships in the bibliographic records of digital libraries of Iran based on Resource Description and Access (RDA) Standard. *Librarianship and Information Organization Studies*, 24(2), 22-40. Retrieved from [http://nastinfo.nlai.ir/article\\_69.html?lang=en](http://nastinfo.nlai.ir/article_69.html?lang=en). (in Persian)
- Norouzi, Y., Gholami, T., & Jafarifar, N. (2018). What is the Situation of Iranian Digital Libraries after a Decade? *Librarianship and Information Organization Studies*, 28(4), 147-170. Retrieved from [http://nastinfo.nlai.ir/article\\_2195.html?lang=en](http://nastinfo.nlai.ir/article_2195.html?lang=en). (in Persian)

- Pokorny, J. (2018). Automatic subject indexing and classification using text recognition and computer based analysis of the table of contents. In Chau, L., & Mounier, P. *ELPUB* 2018. June 2018, Toronto, Canada. DOI: 10.4000/proceedings.elpub.2018.19.
- Porter, A. H. (2018). *The Cambridge introduction to narrative*. Translated by Roya PourAzar and Nima M. Ashrafi. Tehran: Atraf. (in Persian)
- Pourahmad, A. A., & Hassani, M. R. (2008). *Subject cataloging: Principles and methods*. Edited by R. Fattahi. Tehran: Ketabdar. (in Persian)
- Ranganathan, S. R. (1967). *Prolegomena to library classification*. 3<sup>rd</sup> ed. London: Asia Publishing House.
- Razi, Farideh (1993). *A dictionary of pure Persian words for the Arabic loan-words in contemporary Persian*. Tehran: Center Publication. Retrieved from <https://vajehyab.com/sareh/%D9%85%D9%88%D8%B6%D9%88%D8%B9>. (in Persian)
- Reynolds, S. J. (1989). In theory there is no solution: The impediments to a subject cataloging code. *Library Quarterly*, 59(3), 231–32.
- Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010). Automatic keyword extraction from individual documents. In *Text mining: Application and theory*, edited by Berry, M. W. and Koga, J. Wiley & Sons, pp. 3-20. DOI: 10.1002/9780470689446.ch1
- Sepehr, Fereshteh, and Abazari, Zahra (2016). Using ontology in Iranian digital libraries: A Feasibility study. Paper presented at the Second International Conference on Web Research, Tehran. (in Persian)
- Shahbazi, F., Norouzi, Y., & Alipour-Hafezi, M. (2015). Evaluation of using expert system features in information retrieval in Iranian digital library software. *Iranian Journal of Information Processing and Management*, 30(3), 823-851. Retrieved from <http://jipm.irandoc.ac.ir/article-1-2589-fa.html>. (in Persian)
- Short, M. (2019). Text mining and subject analysis for fiction; or, using machine learning and information extraction to assign subject headings to dime novels. *Cataloging and Classification Quarterly*, 57(5), 315-336. DOI: 10.1080/01639374.2019.1653413
- Siddiqi, S., & Sheram, A. (2015). Keyword and keyphrase extraction techniques: A literature review. *International Journal of Computer Applications*, 109(2). DOI: 10.5120/19161-0607
- Soltani, P. (1972). *Colon Classification*. *Iranian Librarians Association Newsletter*, 5(3), 115-120. (in Persian)
- Soltani, Pouri (2002). Persian subject headings (book). In Abbas Hori and Narges Neshat (Editors), *Encyclopedia of Library and Information Science* (first edition). Tehran: National Archives and Library of Iran. Retrieved from [http://portal.nlai.ir/daka/Wiki%20Pages/%D8%B3%D8%B1%D8%B9%D9%86%D9%88%D8%A7%D9%86%D9%8A\(%D9%83%D8%AA%D8%A7%D8%A8\).aspx](http://portal.nlai.ir/daka/Wiki%20Pages/%D8%B3%D8%B1%D8%B9%D9%86%D9%88%D8%A7%D9%86%D9%8A(%D9%83%D8%AA%D8%A7%D8%A8).aspx). (in Persian)
- Soltani, P., & Rastin, F. (2000). *Encyclopedia of Library and Information Science*. 3<sup>rd</sup> Ed. Tehran: Farhang-e Mo'aser.
- Sun, C., Hu, L., Li, S., Li, T., Li, H., & Chi, L. (2020). A review of unsupervised key phrase extraction methods using within-collection resources. *Symmetry*, 12(1864). DOI:10.3390/sym12111864

- Tchoua, R. B. (2019). *Hybrid human-machine scientific information extraction*. (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global database. (UMI No. 13904924)
- Tushara, M. G., Mownika, T., & Mangamuru, R. (2019). A comparative study on different keyword extraction algorithms. In *Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019)*, Erode, India. 969-973
- Vickery, B. C. (1968). Analysis of information. In *Encyclopedia of library and information science*, vol. 1, edited by A. Kent and H. Lancour. New York: Dekker, 355-384.
- Vickery, B. C., & Vickery, A. (2004). *Information science in theory and practice*. 3<sup>rd</sup> ed. München: K. G. Saur.
- Wongchaisuwat, P. (2018). *Topics in document classification*. (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global database. (UMI No. 10749400)
- Wellisch, H. H. (1991). *Indexing from A to Z*. New York: Wilson.
- Wilson, P. (1968). *Two kinds of power: An essay on bibliographic control*. Berkeley, CA: University of California Press.