

# Studying the Patterns of Users' Tagging to Knowledge and Information Science Field Articles in Scientific Social Networks

## Rasul Saadat

PhD Candidate in Knowledge and Information Science; University of Isfahan; Isfahan, Iran; Email: saadat.rasul@gmail.com

## Ahmad Shabani\*

PhD in Knowledge and Information Science; Professor; Department of Knowledge and Information Science; University of Isfahan; Isfahan, Iran Email: shabania@edu.ui.ac.ir

## Asefeh Asemi

PhD in Knowledge and Information Science; Faculty Member at the Institute of Data Analytics and Information Systems; Doctoral School of Economics, Business, & Informatics; Corvinus University of Budapest; Budapest, Hungary Email: asefeh@uni-corvinus.hu

## Mehrdad CheshmehSohrabi

PhD in Information and Communication Sciences; Associate Professor; Department of Knowledge and Information Science; University of Isfahan; Isfahan, Iran Email: mo.sohrabi@edu.ui.ac.ir

## Mohammad TavakoliZadeh Ravari

PhD in Knowledge and Information Science; Associate Professor; Department of Knowledge & Information Science; Yazd University; Yazd, Iran Email: tavakoli@yazd.ac.ir

Iranian Journal of  
**Information  
Processing and  
Management**

Received: 18, Apr. 2020 | Accepted: 24, Jan. 2021

**Abstract:** This research aims to verify the patterns of users' tagging to the articles of knowledge and information science field in academia.edu. The research method is quantitative and based on text mining and applicable typically. The population includes 6086 bibliographic articles and their abstracts extracted from 159 English journals of knowledge and information science field in Scopus database that are core journals in LISTA as well. In order to gather these data, 194337 articles were searched in academia.edu then every tagged article was chosen. Examining the relationship between the growth of different types of tags (one-word, two-word, three-word, and four-word and more) and increasing of documents showed a linear correlation between them. Among the different groups of tags, the highest growth rate was related to two-word tags (0.609%) and the lowest growth rate was related to four-word tags and more (0.143%). The total

\* Corresponding Author

Iranian Research Institute  
for Information Science and Technology  
(IranDoc)

ISSN 2251-8223

eISSN 2251-8231

Indexed by SCOPUS, ISC, & LISTA

Vol. 38 | No. 1 | pp. 109-138

Autumn 2022

<https://doi.org/10.35050/JIPM010.2022.015>



growth rate of the tags (new and duplicate) was also 5.52 (i.e. 5.52 tags per document). It was also found that two-word tags had the most matching (54.92%) and four-word tags had the least matching (1.76%) with different sections of articles (title, abstract, and authors' keywords). The total tags were matched 7.5% with the title, 76.61% with the abstract, and 15.89% with the authors' keywords. Regarding the reuse of tags, it was revealed that in general, 38.8% of the tags have been reused. On the other hand, two-word tags with 57.59% had the most reuse and four-word tags and more with 7.54% had the least. Another point is that 16% of the tags were reused in the first year and more than 50% of the tags were reused in the first 3 years. Finally, it can be said that the existence of a significant user consensus on certain terms indicates that the new patterns of user tags are at least partially compatible with professional indexing concepts about document content, and by focusing on the most widely used tags and their sustainable distribution, weight formulation and even classification schemes may be achieved. Also, users' activities on social networks can be used to increase the quality of suggestions in collective tagging systems. Another point is that there is a connection between professional indexing and user tagging, and the two are not alien to each other. This connectivity can be the basis for a complementary subject access system that enriches professional indexing.

**Keywords:** Social Tagging, Indexing, Knowledge Organization, Folksonomy, Scientific Social Networks, Knowledge and Information Science, Academia.edu.

# مطالعه الگوهای برچسب‌گذاری کاربران به مقالات مرتبط با حوزه علم اطلاعات و دانش‌شناسی در شبکه‌های اجتماعی علمی

رسول سعادت

دانشجوی دکتری علم اطلاعات و دانش‌شناسی؛  
دانشگاه اصفهان؛ اصفهان، ایران؛  
saadat.rasul@gmail.com

احمد شعبانی

دکتری علم اطلاعات و دانش‌شناسی؛ استاد؛  
دانشگاه اصفهان؛ اصفهان، ایران؛  
shabania@edu.ui.ac.ir

عاصفه عاصمی

دکتری علم اطلاعات و دانش‌شناسی؛ دکتری بیزینس  
انفورماتیک؛ عضو هیئت علمی و محقق دانشگاه  
کوروینوس بوداپست؛ بوداپست، مجارستان؛  
asemi.asefeh@uni-corvinus.hu

مهرداد چشمه‌سهرابی

دکتری علوم اطلاعات و ارتباطات؛ دانشیار؛ دانشگاه  
اصفهان؛ اصفهان، ایران mo.sohrabi@edu.ui.ac.ir

محمد توکلی‌زاده راوری

دکتری علم اطلاعات و دانش‌شناسی؛ دانشیار؛  
دانشگاه یزد؛ یزد، ایران tavakoli@yazd.ac.ir



مقاله برای اصلاح به مدت ۸ روز نزد پدیدآوران بوده است.

پدیدوش: ۱۴۰۰/۰۹/۲۷

دریافت: ۱۴۰۰/۰۷/۰۸

**چکیده:** پژوهش حاضر با هدف بررسی الگوهای برچسب‌گذاری کاربران به مقالات مرتبط با حوزه علم اطلاعات و دانش‌شناسی در پایگاه «آکادمیا» انجام شد. روش پژوهش از لحاظ رویکرد، کمی و مبتنی بر پژوهش‌های تحلیل متن و از لحاظ نوع، کاربردی است. جامعه آماری پژوهش حاضر، شامل ۶۰۸۶ مقاله برگرفته از ۱۵۹ مجله انگلیسی‌زبان حوزه علم اطلاعات و دانش‌شناسی موجود در پایگاه استنادی «اسکوپوس» و جزء مجلات هسته در پایگاه «لیستا» است. مقالات مستخرج از مجلات پایگاه «اسکوپوس» (به تعداد ۱۹۴۳۳۷ مقاله) در پایگاه «آکادمیا» مورد جست‌وجو قرار گرفته و آن دسته از مقالاتی که توسط کاربران برچسب‌گذاری شده بود، انتخاب گردید. بررسی رابطه بین رشد انواع برچسب‌ها (یک-واژه‌ای، دو-واژه‌ای، سه-واژه‌ای، و چهار-واژه‌ای و بیشتر) و افزایش مدارک حاکی از همبستگی

نشریه علمی | رتبه بین‌المللی  
پژوهشگاه علوم و فناوری اطلاعات ایران  
(ایرانداک)

شاپا (چاپی) ۲۲۵۱-۸۲۲۳

شاپا (الکترونیکی) ۲۲۵۱-۸۲۳۱

نماینده در SCOPUS، و LISTA، ISC،

ijpm.irandoc.ac.ir

دوره ۳۸ | شماره ۱ | صص ۱۰۹-۱۳۸

پاییز ۱۴۰۱

<https://doi.org/10.35050/IJPM010.2022.015>



خطی بین آن‌ها بود. در بین گروه‌های مختلف برچسب‌ها بیشترین ضریب رشد مربوط به برچسب‌های دو-واژه‌ای (۰/۶۰۹) و کمترین ضریب رشد مربوط به برچسب‌های چهار-واژه‌ای و بیشتر (۰/۱۴۳) بود. ضریب رشد کل برچسب‌ها (جدید و تکراری) نیز ۵/۵۲ بود (یعنی به ازای هر مدرک ۵/۵۲ برچسب). همچنین، مشخص شد که برچسب‌های دو-واژه‌ای بیشترین انطباق (۵۴/۹۲ درصد) و برچسب‌های چهار-واژه‌ای و بیشتر کمترین انطباق (۱/۷۶ درصد) را بخش‌های مختلف مقالات (عنوان، چکیده، و کلیدواژه‌های نویسندگان) داشته است. کل برچسب‌ها نیز به میزان ۷/۵ درصد با عنوان، ۷۶/۶۱ درصد با چکیده، و ۱۵/۸۹ درصد با کلیدواژه‌های نویسندگان منطبق بود. در مورد استفاده مجدد از برچسب‌ها نیز آشکار شد که به‌طور کلی، ۳۸/۸ درصد برچسب‌ها مورد استفاده مجدد قرار گرفته است. از سوی دیگر، برچسب‌های دو-واژه‌ای با ۵۷/۵۹ درصد بیشترین و برچسب‌های چهار-واژه‌ای و بیشتر با ۷/۵۴ درصد کمترین استفاده مجدد را داشتند. نکته دیگر این که ۱۶ درصد از برچسب‌ها در سال اول و بیش از ۵۰ درصد برچسب‌ها تقریباً در ۳ سال اول مورد استفاده مجدد قرار گرفته‌اند. در پایان می‌توان گفت که وجود اجماع قابل توجه کاربران در مورد اصطلاحاتی معین بیانگر آن است که الگوهای جدید از برچسب‌های کاربران حداقل تا حدی با مفاهیم نمایه‌سازی حرفه‌ای در مورد محتوای مدارک سازگار است و با تمرکز بر پرکاربردترین برچسب‌ها و توزیع پایدار آن‌ها می‌توان به فرمولی برای تعیین وزن و حتی طرح‌های رده‌بندی و طبقه‌بندی دست یافت. همچنین، از فعالیت‌های کاربران در شبکه‌های اجتماعی به‌منظور افزایش کیفیت پیشنهادات در نظام‌های برچسب‌گذاری جمعی می‌توان بهره برد. نکته دیگر اینکه، بین نمایه‌سازی حرفه‌ای و برچسب‌گذاری کاربر پیوستگی وجود دارد و این دو نسبت به هم بیگانه نیستند. این پیوستگی می‌تواند پایه و اساسی برای یک سیستم مکمل دسترسی موضوعی که موجب غنی شدن نمایه‌سازی حرفه‌ای است، باشد.

**کلیدواژه‌ها:** برچسب‌گذاری اجتماعی، نمایه‌سازی، سازماندهی اطلاعات، رده‌بندی مردمی، شبکه‌های اجتماعی علمی، علم اطلاعات و دانش‌شناسی، آکادمیا

## ۱. مقدمه

توجه به روش نمایه‌سازی دموکراتیک در قرن ۲۱ و در بستر وب ۲ بیش از پیش افزایش یافته و در قالب برچسب‌گذاری اجتماعی<sup>۱</sup> و رده‌بندی‌های مردمی<sup>۲</sup> (که می‌توان به آن نمایه‌های مردمی هم گفت) بروز و ظهور یافته است. در سال‌های اخیر، رده‌بندی‌های مردمی به منزله نسل جدیدی از ابزارهای ذخیره و بازیابی، به‌کارگیری، بازنمون، و تولید اطلاعات در بستر وب ۲ مطرح شده است. در واقع، آن‌ها مکمل فنون سنتی سازماندهی دانش به شیوه‌های متفاوت هستند. ویژگی رده‌بندی‌های مردمی بُعد اجتماعی و توسعه از

1. social tagging

2. folksonomies

طریق کاربران است (Alexiev et al. 2005). به عبارتی، تولید فراداده‌ها در بستر رده‌بندی‌های مردمی از طریق کاربران صورت گرفته و این نوع نظام‌های سازماندهی دانش روشی ساختاریافته در جهت نمایه‌سازی مدارک به‌شمار می‌رود. در حالی که در بیشتر موارد، نمایه‌سازان حرفه‌ای و متخصص مسئول نمایه‌سازی مدارک هستند، رده‌بندی‌های مردمی این امکان را به تولیدکنندگان و یا کاربران محتوا داده که خود مسئولیت نمایه‌سازی را بر عهده بگیرند (Raferty 2018).

امروزه، با رشد فزاینده شبکه‌های اجتماعی علمی، برجسب‌گذاری اجتماعی از طریق رده‌بندی‌های مردمی نه تنها سازماندهی اطلاعات از سوی کاربران را تحت تأثیر قرار داده، بلکه سازماندهی اطلاعات در حوزه علم اطلاعات و دانش‌شناسی را نیز مورد توجه قرار داده است. به گفته «چان» نمایه‌سازی موضوعی و برجسب‌گذاری اجتماعی دو رویکرد متفاوت برای ذخیره و بازیابی اطلاعات هستند (Chan 2009). این پژوهشگر بر پتانسیل برجسب‌های اجتماعی جهت تقویت واژگان کنترل‌شده و دسترسی موضوعی تصریح کرده و در ادامه می‌افزاید دلیل این امر آن است که برجسب‌های اجتماعی فهمی از دیدگاه و رفتار جست‌وجوی کاربران و همچنین، منبعی جهت افزودن اصطلاحات به واژگان کنترل‌شده فراهم می‌آورد.

بنابراین، می‌توان گفت که نمایه‌سازی منابع به زبان طبیعی کاربران از نکات بارز رده‌بندی‌های مردمی است؛ یعنی همان زبانی که هنگام جست‌وجوی اطلاعات توسط کاربران مورد استفاده قرار می‌گیرد. در واقع، شکل‌گیری رده‌بندی‌های مردمی بر اساس ذهن و زبان کاربر، نقطه قوت آن‌هاست، در حالی که نمایه‌سازی حرفه‌ای که امروزه در حوزه علم اطلاعات و دانش‌شناسی بحث می‌شود، با درک کمتری نسبت به نیازهای اطلاعاتی کاربران و آنچه در ذهن آن‌ها می‌گذرد، انجام می‌شود. همان‌طور که یافته‌های پژوهش‌های «محمدپور، مومنی، و علیپور حافظی» (۱۳۹۵)؛ Lee & Schleyer (2010, 2012)؛ Kipp (2006a, 2007b, 2011)؛ Rorissa (2010)؛ Lyer & Bungo (2011)؛ «قنواتی» و همکاران (۱۳۹۷)، «خادمیان، کوبی، و فریده عصاره» (۱۳۹۶)، و «موحدیان» و همکاران (۱۳۹۹) در این حوزه نشان داده است، برجسب‌های اجتماعی کاربران با اصطلاحات نمایه‌ای نمایه‌سازان و یا سرعنوان‌های موضوعی دارای همپوشانی کمی بوده و میان این دو گروه فهم بسیار متفاوتی نسبت به مدارک وجود دارد.

بر این اساس، سؤالی که ذهن پژوهشگران را به خود معطوف داشته، این است که

آیا می‌توان با تکیه بر نقطه قوت رده‌بندی‌های مردمی و برچسب‌های اجتماعی و در واقع، با الگویی از این نظام‌ها در جهت بهبود سازماندهی اطلاعات حرکت کرد؛ به گونه‌ای که هنگام نمایه‌سازی مدارک توجه بیشتری به ذهن و زبان کاربر صورت گیرد؟

باید به این نکته توجه داشت که تخصیص برچسب به مدارک علمی توسط خوانندگان و کاربران یکی از روش‌های جدیدی است که با توجه به گسترش فناوری اطلاعات به صورت گسترده ممکن شده است. این عمل، که می‌توان نام برچسب‌گذاری و یا نمایه‌سازی مردمی به آن منتسب کرد، حاصل همکاری متخصصان موضوعی و علاقه‌مندان حوزه‌های موضوعی به صورت گروهی است. مسائلی که در حوزه ادبیات نمایه‌سازی موضوعی و به‌ویژه ماشینی مطرح است، می‌تواند قابل تعمیم به حوزه برچسب‌گذاری باشد، زیرا هر دو با مبحث تعیین اصطلاحات موضوعی مدارک در ارتباط هستند. از جمله موضوعات مطرح در حوزه نمایه‌سازی، تعیین اصطلاحات تک و چندواژه‌ای، اهمیت بخش‌های مختلف یک متن برای تشخیص واژگان مهم و حاوی محتوا، تحلیل‌های مرتبط با سری زمانی در خصوص زمان حضور و خروج اصطلاحات موضوعی از یک حوزه علمی یا فنی، و مباحث آماری در خصوص فراوانی حضور اصطلاحات در یک متن یا مجموعه‌ای از مدارک است. پرداختن به این موضوعات می‌تواند واقعیت‌های پنهانی را نمایان کند که در تشخیص یا انتساب موضوعات به مدارک (به صورت دستی یا ماشینی) وجود دارد. مسئله‌ای که انگیزه اصلی انجام این پژوهش گردید، پنهان بودن چنین واقعیت‌ها و الگوهایی است که در پس‌زمینه تخصیص برچسب‌ها توسط کاربران شبکه‌های اجتماعی علمی به مدارک تخصصی وجود دارد.

از آنجا که در علم اطلاعات و دانش‌شناسی تأکید بر تحلیل حوزه و انجام پژوهش‌ها (به‌ویژه سازماندهی اطلاعات) به صورت حوزه‌ای است، جهت انجام پژوهش حاضر، حوزه علم اطلاعات و دانش‌شناسی که حوزه مطالعاتی پژوهشگران پژوهش حاضر نیز هست، انتخاب گردید.

از جمله بسترهای علمی که شرایط بررسی برچسب‌های اجتماعی را دارد، پایگاه اجتماعی «آکادمیا» است. این شبکه اجتماعی که اکنون در وبگاه «الکسا» رتبه‌ای بین ۳۰۰ تا ۴۰۰ جهان را دارد (Alexa.com 2021) و پژوهشگران تمام حوزه‌ها را از سرتاسر جهان به

خود جذب کرده، بستر مناسبی جهت تحلیل برجسب‌های تخصیص‌یافته از سوی آن‌ها به مقالات فراهم نموده است (Academia.edu 2021).

بر این اساس، سؤال اصلی پژوهش حاضر این است که برجسب‌هایی که کاربران به مقالات حوزه علم اطلاعات و دانش‌شناسی در شبکه اجتماعی «آکادمیا» تخصیص داده‌اند، چه خصوصیت‌هایی دارند؟ در واقع، با طرح چنین سؤالی پژوهش حاضر در صدد است به سؤالات زیر پاسخ دهد:

۱. الگوی رشد گروه‌های مختلف برجسب‌های تخصیص‌یافته به مقالات مرتبط با حوزه علم اطلاعات و دانش‌شناسی از نظر تعداد واژه‌های تشکیل‌دهنده نسبت به افزایش مدارک در پایگاه «آکادمیا» چگونه است؟
۲. رابطه بین تعداد واژه‌های تشکیل‌دهنده برجسب‌های تخصیص‌یافته توسط کاربران با بخش‌های مختلف مقالات (عنوان، چکیده، کلیدواژه‌های نویسندگان) مرتبط با حوزه علم اطلاعات و دانش‌شناسی در پایگاه «آکادمیا» از لحاظ انطباق چگونه است؟
۳. الگوی زمانی بین اولین زمان تخصیص یک برجسب به مقالات مرتبط با حوزه علم اطلاعات و دانش‌شناسی تا استفاده مجدد از آن در واحدهای مختلف زمانی در پایگاه «آکادمیا» چگونه است؟

## ۲. پیشینه پژوهش

بررسی پژوهش‌های داخلی و خارجی نشان داد که در راستای پژوهش حاضر، تاکنون دو رویکرد وجود دارد:

۱. پژوهش‌هایی که برجسب‌های اجتماعی را با واژگان موجود در منابع اطلاعاتی (از جمله صفحات وب و مقالات) و نیز با اصطلاحات نمایه‌ای از نظر انطباق آن‌ها با یکدیگر مقایسه نموده‌اند؛
۲. پژوهش‌هایی که بر اساس ویژگی‌های برجسب‌ها، به کشف خصوصیات، الگوهای رشد، و استفاده مجدد از برجسب‌ها روی آورده‌اند.

دسته اول: انطباق بین برجسب‌های اجتماعی و واژگان موجود در بخش‌های مختلف منابع اطلاعاتی. از جمله پژوهش‌هایی که در این دسته قرار می‌گیرند، به شرح زیر است:  
با مقایسه برجسب‌های اجتماعی پایگاه del.icio.us با عنوان‌ها و متن صفحات وب، «همان، کوتریکا و گارسیا-مولینا» دریافتند که ۱۶ درصد از برجسب‌ها در عناوین صفحات

وب ظاهر شدند. نتیجه گیری کلی آن‌ها این بود که برچسب‌گذاری اجتماعی قادر است داده‌های اضافی را به‌منظور جست‌وجوی بهتر منابع فراهم نماید (Heymann, Koutrika & Garcia-Molina 2008).

با انجام پژوهشی در شبکه اجتماعی «کانوتی»<sup>۱</sup> در مورد مقالات حوزه فناوری رایانه و با بررسی برچسب‌های این پایگاه، «هکنر، مولباخر و ولف» دریافتند که ۴۹ درصد از برچسب‌ها از کلیدواژه‌های عنوان، و ۹ درصد از چکیده مقالات انتخاب شده بود. آن‌ها همچنین نشان دادند که تنها ۳۰ درصد از برچسب‌ها نسبت به کلیدواژه‌های تخصیص یافته به مدارک از سوی نویسندگان دارای همپوشانی است و کاربران در مقایسه با نویسندگان تمایل بیشتری نسبت به اختصاص اصطلاحات عام‌تر و ساده‌تر به مدارک از خود نشان دادند (Heckner, Mühlbacher & Wolff 2008).

با تحلیل کلیدواژه‌های عنوان، چکیده، سرعنوان‌های موضوعی، و اصطلاحات نمایه‌ای ۴۵ مجله علم فیزیک و برچسب‌های مربوطه توسط «هستین و پیترز» در پایگاه‌های «سایت یولایک»<sup>۲</sup>، «کانوتی» و «بیسونومی»<sup>۳</sup> مشخص شد که بیشترین انطباق مربوط به چکیده و عنوان بود. در واقع، ۷۷/۶ درصد از برچسب‌های کاربران با چکیده و ۶۶ درصد با واژگان عنوان مطابقت داشتند. همچنین، ۲۹/۳ درصد برچسب‌ها با کلیدواژه‌های نویسندگان انطباق داشت (Haustein & Peters 2012).

«محمدپور، مومنی، و علیپور حافظی» با هدف تعیین میزان همخوانی میان برچسب‌های اختصاص یافته به مقالات علم اطلاعات و دانش‌شناسی در پایگاه «سایت یولایک» و توصیف‌گرهای اختصاص یافته به همان مقالات در پایگاه «لیستا»<sup>۴</sup> دریافتند که کمتر از ۲ درصد از برچسب‌های اختصاص یافته به مقالات منتخب حوزه مذکور در پایگاه «سایت یولایک» با آنچه به‌عنوان توصیف‌گر به همان مقالات در پایگاه «لیستا» تخصیص یافته، همخوانی دارند (۱۳۹۵).

«خادمیان» و همکاران با هدف میزان مطابقت و امکان جایگزینی یا تکمیل سرعنوان‌های موضوعی «کتابخانه کنگره» با برچسب‌های اجتماعی «لایبری‌تینگ»<sup>۵</sup> در کتاب‌های حوزه‌های علوم انسانی، علوم اجتماعی، و علوم طبیعی دریافتند که در ۴۳ درصد موارد مطابقت دقیق و تقریبی و در ۲۳ درصد موارد مطابقت نسبی با سرعنوان‌های

1. Connotea

2. CiteUlike

3. Bibsonomy

4. LISTA

5. LibraryThing



موضوعی وجود داشت. محاسبه ضریب شباهت «ژاکار»<sup>۱</sup> نیز نشان داد که جایگزینی در سه حوزه تأیید نمی‌شود (۱۳۹۶).

«قناتی» و همکاران در پژوهش خود با هدف شناسایی میزان تطابق زبان نمایه‌سازان، نویسندگان و برجسب‌گذاران در پایگاه‌های اطلاعاتی «اریک» و «مندلی» دریافتند که میزان تطابق کلیدواژه‌های تخصیص داده‌شده توسط نویسندگان مدارک با برجسب‌های تخصیص داده‌شده توسط برجسب‌گذاران به همان مدارک در وبگاه «مندلی» ۱۵ درصد و میزان تطابق توصیفگرهای تخصیص داده‌شده توسط نویسندگان به مدارک در پایگاه «اریک» با برجسب‌های تخصیص داده‌شده توسط برجسب‌گذاران به همان مدارک در وبگاه «مندلی» ۳ درصد و در نهایت، میزان تطابق هر سه زبان مورد مطالعه ۱/۱ درصد بود (۱۳۹۷).

«موحدیان» و همکاران با هدف بررسی وضعیت برجسب‌های تخصیص داده‌شده به کتاب‌های انگلیسی با موضوع اسلام در پایگاه «لایبری ینگ» در مقایسه با سرعنوان‌های موضوعی «کنگره» دریافتند که میانگین فراوانی سرعنوان‌های موضوعی بیش از ۳/۵ درصد برای هر کتاب بود. بیش از ۸۳ درصد از برجسب‌ها با سرعنوان‌ها انطباق معنایی و یا لفظی نداشتند، اما حدود ۶۸ درصد از برجسب‌های نامنطبق با سرعنوان‌ها از متن کتاب‌ها قابل استخراج بودند (۱۳۹۹).

مقایسه برجسب‌های تخصیص‌یافته به مقالات رشته علوم دریایی در پایگاه «سایت یولایک» با کلیدواژه‌های نویسندگان و واژه‌های عناوین همان مقالات توسط «وایدیا و هارینارایانا» نشان داد که میزان همپوشانی بین برجسب‌های کاربران با کلیدواژه‌های نویسندگان ۴۴/۴۷ درصد و میزان همپوشانی برجسب‌ها با واژه‌های عنوان مقالات ۳۶/۲۹ درصد است (Vaidya & Harinarayana 2019).

دسته دوم: پژوهش‌هایی که بر اساس ویژگی‌های برجسب‌ها به کشف خصوصیات، الگوهای رشد، و استفاده مجدد از برجسب‌ها روی آورده‌اند. برخی از این پژوهش‌ها به قرار زیر است:

مطالعه رشد مجموعه برجسب‌ها در گذر زمان با استفاده از داده‌های پایگاه «فلیکر»<sup>۲</sup> توسط «مارلو» و همکاران نشان داد که افزایش برجسب‌های جدید به صورت خطی با افزایش مدارک جدید (عکس‌ها) در این پایگاه مرتبط است و به ازای هر عکس ۲

1. Jaccard similarity coefficient

2. Flickr

برچسب توسط کاربران اضافه می‌گردد (Marlow et al. 2006).

«توکل‌ی زاده راوری» به بررسی الگوی کمی رشد موضوعات پزشکی نسبت به افزایش مدارک پرداخته است. این پژوهشگر برچسب‌ها را مورد مطالعه قرار نداده، بلکه رشد موضوعات را نسبت به افزایش مدارک بررسی نموده است که با سؤال اول پژوهش حاضر در ارتباط است. یافته‌های وی نشان داد که امروزه، در ازای انتشار بیش از ۲۵۰ مدرک در موضوعات گوناگون علوم پزشکی به صورت خطی، یک موضوع جدید فراهم و به «مش»<sup>۱</sup> اضافه می‌گردد (۱۳۸۶).

ارزیابی رفتار برچسب‌گذاری کاربران در شبکه اجتماعی «سایت یولایک» توسط «فاروق» و همکاران و محاسبه نرخ استفاده مجدد از برچسب‌ها نشان داد که متوسط رخداد استفاده مجدد از هر برچسب ۳/۹ درصد است (Farooq et al. 2007).

با بررسی جنبه‌های زمانی مربوط به علایق کاربران و نیز پدیده تغییر گرایش‌های موضوعی در سه نظام برچسب‌گذاری (del.icio.us، بیسونومی<sup>۲</sup> و فلیکر) توسط «یین» و همکاران، این یافته حاصل شد که احتمال این که کاربران از برچسب‌های جدید استفاده کنند، ۱۳/۹ درصد است (Yin et al. 2011). چنین نکته‌ای بیانگر آن است که در ۸۶/۱ درصد موارد کاربران با توجه به برچسب‌هایی که پیش‌تر در نظام‌های رده‌بندی مردمی استفاده شده، انتخاب خود را انجام می‌دهند.

با مطالعه نظام‌های برچسب‌زنی اجتماعی (del.icio.us، گودریلدز<sup>۳</sup> و لایبری‌تینگ) از سوی «زوبیاگا، مارتینز و برنسو» به منظور فهم توزیع برچسب‌ها در رده‌بندی‌های مردمی آشکار شد که رشد مدارک و برچسب‌های تخصیص یافته از سوی کاربران از یک رشد لگاریتمی پیروی کرده است. بر این اساس در پایگاه del.icio.us به ازای هر مدرک وب ۳۳/۳۵ برچسب، در «لایبری‌تینگ» به ازای هر کتاب ۱۴/۵۳ برچسب، و در «گودریلدز» به ازای هر کتاب ۱۳/۳۳ برچسب از سوی کاربران تخصیص یافته بود (Zubiaga, Mart'inez & Fresno 2011b).

با هدف کشف این نکته که کاربران کدام ویژگی‌های مدارک را بیشتر مورد توجه قرار می‌دهند، «چوی و سین» به بررسی برچسب‌های کاربران پایگاه «ناینز»<sup>۴</sup> در حوزه

1. MeSH

2. Bibsonomy

3. Goodreads

4. NINES

تاریخ نسبت به منابع متنی، غیرمتنی، و کتابشناختی پرداختند. یافته‌های پژوهش آن‌ها آشکار ساخت که از ۱۵۴۰ برجسب منحصربه‌فرد نزدیک به ۷۶ درصد تک‌واژه‌ای و ۲۴ درصد چندواژه‌ای بودند. همچنین، در بین برجسب‌های منحصربه‌فرد، ۵۳/۴۴ درصد فقط یک بار استفاده شده، ۱۳/۱۲ درصد برجسب‌ها دو بار و ۶/۴۹ درصد نیز سه بار مورد استفاده قرار گرفتند. از این میان، فقط ۵۲/۵۲ درصد بیش از ۱۰۰ بار استفاده شدند (Choi & Syn 2016).

از جمع‌بندی پژوهش‌های انجام‌شده نکات زیر به دست آمد:

مقایسه پژوهش‌های گروه اول و دوم نشان‌دهنده این مطلب است که گروه اول بیشتر جنبه توصیفی داشته و گروه دوم جنبه تحلیلی دارد؛ به طوری که، گروه اول پژوهش‌ها بیشتر به ارائه آمار توصیفی در مورد میزان برجسب‌های برگرفته از بخش‌های مختلف منابع اطلاعاتی و یا اصطلاحات نمایه‌ای و نیز انطباق برجسب‌های اجتماعی با اصطلاحات نمایه‌ای نمایه‌سازان و یا سرعنوان‌های موضوعی تأکید داشته، در حالی که گروه دوم بر کشف قوانینی در مورد فعالیت کاربران، مطالعه الگوهای رشد برجسب‌ها و استفاده مجدد از برجسب‌ها تأکید دارد.

تمرکز بیشتر پژوهش حاضر بر برجسب‌های تک‌واژه‌ای و چندواژه‌ای و کاربرد آن‌ها در نمایه‌سازی است. افزون بر این، آنچه مورد تأکید قرار گرفته، بررسی همزمان ویژگی‌های انواع برجسب‌ها از نظر الگوی رشد تعداد واژه‌های تشکیل‌دهنده نسبت به افزایش مدارک، انطباق برجسب‌های تخصیص‌یافته توسط کاربران با بخش‌های مختلف مقالات (عنوان، چکیده، کلیدواژه‌های نویسندگان)، و الگوهای زمانی استفاده مجدد از برجسب‌های موجود (بقای برجسب‌ها) است که می‌تواند تکمیل‌کننده پژوهش‌های پیشین قلمداد گردد.

افزون بر این، پژوهش حاضر با گزینش بیش از ۶۰۰۰ مقاله در تمام زمینه‌های مهم علم اطلاعات و دانش‌شناسی که برگرفته از ۱۵۹ مجله معتبر این حوزه در پایگاه «اسکوپوس» است، به جامعیت و وسعت کار توجه داشته و ارائه الگوها و رویکردهای برجسب‌گذاری کاربران به صورت جامع و وسیع از ویژگی‌های این پژوهش است. از سوی دیگر، در این پژوهش برجسب‌های اجتماعی در پایگاه «آکادمیا» مورد بررسی قرار گرفته است که امروزه، یکی از پرمراجعه‌ترین شبکه‌های اجتماعی علمی دنیاست. در مورد پژوهش حاضر آنچه که حائز اهمیت است، پوشش اغلب تخصص‌های

علم اطلاعات و دانش‌شناسی با تعداد زیاد مقالات و برجسب‌های مربوط به آن‌ها در این شبکه اجتماعی علمی است که مورد بررسی قرار می‌گیرد. نکته دیگر این که در پژوهش‌های قبلی موردی که به بررسی پایگاه مذکور پرداخته باشد، مشاهده نشد.

### ۳. روش پژوهش

پژوهش حاضر از لحاظ رویکرد، کمی و مبتنی بر پژوهش‌های تحلیل متن و از لحاظ نوع، کاربردی است. متن کاوی یا تحلیل متن شامل بازبایی اطلاعات به منظور مطالعه توزیع فراوانی واژگان، شناسایی الگوها، برجسب‌گذاری، استخراج اطلاعات، فنون داده کاوی، و پیش‌بینی است. هدف از تحلیل متن، تبدیل متن به داده‌ها به منظور تجزیه و تحلیل آن از طریق به کارگیری انواع روش‌های تحلیلی است (Miner et. al. 2012).

جامعه آماری پژوهش حاضر، شامل ۶۰۸۶ رکورد مقاله برگرفته از ۱۵۹ مجله انگلیسی‌زبان در حوزه علم اطلاعات و دانش‌شناسی موجود در پایگاه استنادی «اسکوپوس» و جزء مجلات هسته در پایگاه «لیستا» است. رکوردهای مقالات مستخرج از مجلات پایگاه «اسکوپوس» (به تعداد ۱۹۴۳۳۷ رکورد) در پایگاه «آکادمیا» مورد جست‌وجو قرار گرفته و آن دسته از مقالاتی که توسط کاربران برجسب‌گذاری شده بود، انتخاب گردید. بنابراین، هم در مرحله انتخاب مجلات (به دلیل این که تمام مجلات انگلیسی‌زبان انتخاب شدند) و هم در مرحله انتخاب مقالات (به دلیل این که تمام مقالات برجسب‌گذاری شده انتخاب شدند) از روش سرشماری برای مطالعه بهره‌جویی شد. در روش سرشماری کار نمونه‌گیری انجام نمی‌شود، بلکه کل جامعه برای مطالعه انتخاب می‌گردد (حبیبی ۱۳۹۲). به منظور گردآوری داده‌ها از پایگاه «آکادمیا» از سه نرم‌افزار (ربات) در مراحل مختلف استفاده شد که هر سه نرم‌افزار بر اساس زبان C# هنگام گردآوری داده‌ها بر اساس شرایط پژوهش و داده‌های مورد نیاز تهیه گردید.

بر این اساس، ابتدا تعداد ۸۳۴۰ مقاله استخراج گردید. سپس، مراحل زیر انجام شد:

۱. ورود داده‌ها به برنامه و تشکیل پیکره متن؛
۲. پیش‌پردازش داده‌ها: در این مرحله داده‌هایی که حاوی اطلاعات مبهم و ناخوانا (مانند حروف و علائم زائد) و نیز دارای برجسب‌هایی نامرتب و تکراری (مانند برجسب Refugee Studies که در برجسب ۱۶۰۰ رکورد با عنوان Book review تکرار شده) بود، حذف گردید. همچنین، علائم سجاوندی (شامل نقطه، دونقطه، ویرگول، پرانتز،

کروشه، آکولاد، گیومه، علامت سؤال، علامت تعجب) حذف شد. سپس، به‌منظور یکدستی در متن، تمام واژه‌های جمع (چه با قاعده و چه بی‌قاعده) به واژه‌های مفرد تبدیل گردید. این کار پیش‌پردازش به‌منظور تسهیل در تشکیل ماتریس مدرک-برجسب صورت گرفت. سرانجام، تعداد ۶۰۸۶ رکورد جهت تجزیه و تحلیل باقی ماند. جدول ۱، جامعه آماری پژوهش را نشان می‌دهد:

جدول ۱. جامعه آماری پژوهش

تعداد رکوردهای مقالات	برجسب یک-واژه‌ای		برجسب دو-واژه‌ای		برجسب سه-واژه‌ای		برجسب چهار-واژه‌ای و ...		کل برجسب‌ها	
	بدون تکرار	با تکرار	بدون تکرار	با تکرار	بدون تکرار	با تکرار	بدون تکرار	با تکرار	بدون تکرار	با تکرار
۶۰۸۶	۱۳۵۴	۷۱۱۹	۴۰۴۶	۱۸۳۹۶	۱۳۷۵	۳۵۹۴	۹۵۸	۵۰۸۷	۷۷۳۳	۳۴۱۹۶

۳. تقسیم برجسب‌های تخصیص‌یافته به مدارک بر اساس تعداد واژه‌های تشکیل‌دهنده آن‌ها به یک-واژه‌ای، دو-واژه‌ای، سه-واژه‌ای و چهار-واژه‌ای و بیشتر؛

۴. تشکیل ماتریس مدرک-برجسب: به‌منظور مدل‌سازی در سطح واژگان، ماتریس مدرک-برجسب تشکیل شد. در این ماتریس تعداد تکرار هر برجسب در هر مدرک مشخص گردید.

به‌منظور تجزیه و تحلیل داده‌ها از نرم‌افزار «اکسل ۲۰۱۶» و «اس‌پی‌اس اس ۲۶» استفاده شد. همچنین، به‌منظور بررسی این که برجسب‌ها در کدام بخش از مقالات عیناً تکرار شده‌اند، با کمک زبان C# برنامه رایانه‌ای تهیه گردید و به کمک آن انطباق بین برجسب‌ها با بخش‌های مختلف مقالات (عنوان، چکیده، و کلیدواژه‌های نویسنده) انجام شد.

سرانجام، به‌منظور پاسخگویی به سؤال ۱، از همبستگی خطی «پیرسون» بهره‌جویی شد. در همبستگی خطی ارتباط بین دو یا چند متغیر تعیین می‌گردد. در صورتی که تغییرات یک متغیر بر حسب مقادیر مختلف متغیر دیگر تغییر نکنند، دو متغیر مستقل از یکدیگرند. در غیر این صورت، بین دو متغیر همبستگی وجود دارد. ضریب همبستگی مقداری بین ۱- و ۱+ است. ضریب منفی نشان‌دهنده رابطه منفی است که در آن با افزایش یک متغیر، متغیر دیگر کاهش می‌یابد. ضریب مثبت رابطه مثبت را نشان می‌دهد که در این نوع رابطه با افزایش یا کاهش یک متغیر، متغیر دیگر نیز افزایش یا کاهش می‌یابد. ضریب صفر یا

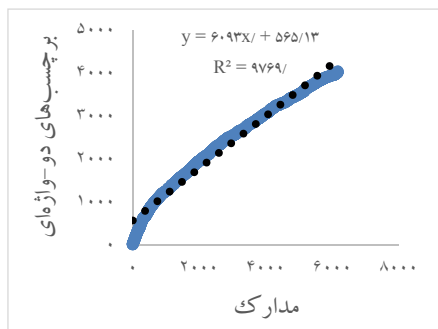
نزدیک صفر نشان‌دهنده نبود رابطه یا وجود رابطه اندک بین متغیرهاست (پاول ۱۹۹۱).  
به منظور پاسخگویی به سؤال ۲، از آمار توصیفی از جمله فراوانی، درصد، و نمودار خطی  
و دایره‌ای استفاده شد.

پاسخگویی به سؤال ۳، با استفاده از تحلیل بقا<sup>۱</sup> و جداول عمر<sup>۲</sup> صورت گرفت. تحلیل  
بقا مجموعه‌ای از روش‌های آماری است که به زمان وقوع یک رخداد می‌پردازد (Guo  
2010). به عبارت دیگر، تحلیل بقا مجموعه‌ای از راهکارهای آماری برای تجزیه و تحلیل  
داده‌هاست که در آن متغیر حاصل، زمان وقوع رخدادهاست. از جمله این رخدادها می‌توان  
به مرگ، بیماری، بهبود، و یا پایان یک واقعه اشاره کرد (Kleinbaum & Klein 2012).

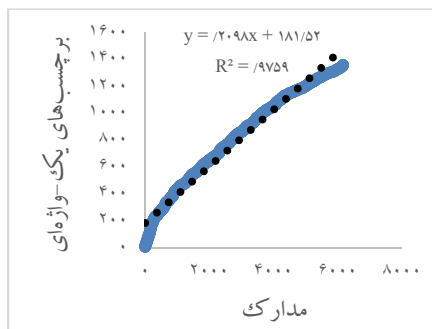
#### ۴. یافته‌های پژوهش

سؤال ۱. الگوی رشد گروه‌های مختلف برچسب‌های تخصیص یافته به مقالات مرتبط با  
حوزه علم اطلاعات و دانش‌شناسی از نظر تعداد واژه‌های تشکیل‌دهنده نسبت به افزایش  
مدارک در پایگاه «آکادمیا» چگونه است؟

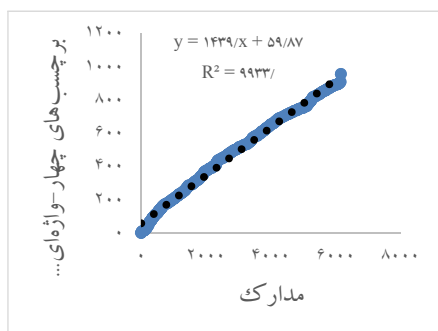
رشد برچسب‌های یک-واژه‌ای، دو-واژه‌ای، سه-واژه‌ای، و چهار-واژه‌ای و بیشتر، و  
نیز کل برچسب‌ها نسبت به افزایش مدارک به صورت ضریب همبستگی «پیرسون» در  
قالب شکل ۱، و نمودارهای شش‌گانه (الف) تا (و) و جدول ۲، مورد بررسی قرار گرفت:



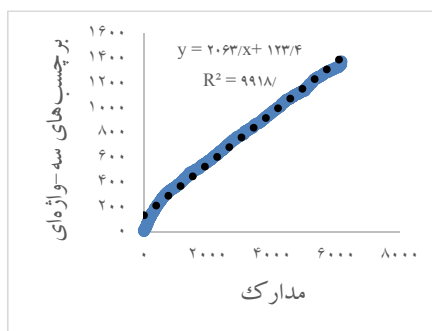
(ب)



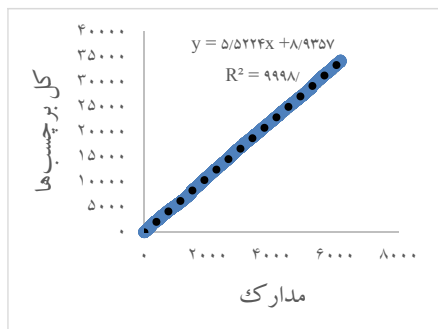
(الف)



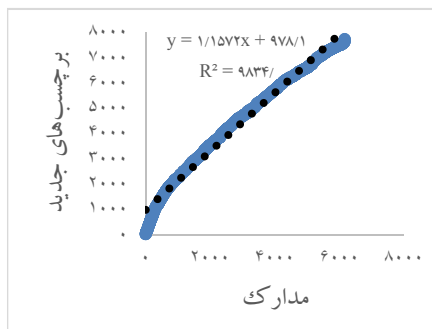
(د)



(ج)



(هـ)



(و)

شکل ۱. رابطه بین رشد برجسب‌ها و افزایش مدارک

نمودار (الف) در شکل ۱، نشان‌دهنده همبستگی خطی (با اطمینان ۹۷ درصد) بین رشد برجسب‌های یک واژه‌ای نسبت به افزایش مدارک در پایگاه «آکادمیا»ست. ضریب رشد برجسب‌های یک-واژه‌ای ۲۰۹/ بوده و به این معناست که به ازای بارگذاری هر مدرک

در پایگاه «آکادمیا» ۰/۲۰۹ برچسب یک-واژه‌ای از سوی کاربران به مدارک موجود در این پایگاه تخصیص یافته است (یعنی به‌ازای هر ۵ مدرک، ۱/۰۴ برچسب یک-واژه‌ای).

نمودار (ب) در شکل ۱، نشان‌دهنده همبستگی خطی (با اطمینان ۹۷ درصد) بین رشد برچسب‌های دو-واژه‌ای نسبت به افزایش مدارک در پایگاه «آکادمیا» است. ضریب رشد برچسب‌های دو-واژه‌ای ۰/۶۰۹ بوده و به این معناست که به‌ازای بارگذاری هر مدرک در پایگاه «آکادمیا» ۰/۶۰۹ برچسب دو-واژه‌ای از سوی کاربران به مدارک موجود در این پایگاه تخصیص یافته است (یعنی به‌ازای هر ۲ مدرک ۱/۲۱ برچسب دو-واژه‌ای)

نمودار (ج) در شکل ۱، نشان‌دهنده همبستگی خطی (با اطمینان ۹۹ درصد) بین رشد برچسب‌های سه-واژه‌ای نسبت به افزایش مدارک در پایگاه «آکادمیا» است. ضریب رشد برچسب‌های سه-واژه‌ای ۰/۲۰۶ بوده و به این معناست که به‌ازای بارگذاری هر مدرک در پایگاه «آکادمیا» ۰/۲۰۶ برچسب سه-واژه‌ای از سوی کاربران به مدارک موجود در این پایگاه تخصیص یافته است (یعنی به‌ازای هر ۵ مدرک، ۱/۰۳ برچسب سه-واژه‌ای).

نمودار (د) در شکل ۱، نشان‌دهنده همبستگی خطی (با اطمینان ۹۹ درصد) بین رشد برچسب‌های چهار-واژه‌ای و بیشتر نسبت به افزایش مدارک در پایگاه «آکادمیا» است. ضریب رشد برچسب‌های چهار-واژه‌ای و بیشتر ۱/۴۳ بوده و به این معناست که به‌ازای بارگذاری هر مدرک در پایگاه «آکادمیا» ۰/۱۴۳ برچسب چهار-واژه‌ای و بیشتر از سوی کاربران به مدارک موجود در این پایگاه تخصیص یافته است (یعنی به‌ازای هر ۷ مدرک، ۱/۰۰۱ برچسب چهار-واژه‌ای و بیشتر).

نمودار (ه) در شکل ۱، نشان‌دهنده همبستگی خطی (با اطمینان ۹۸ درصد) بین رشد کل برچسب‌های جدید نسبت به افزایش مدارک در پایگاه «آکادمیا» است. ضریب رشد کل برچسب‌ها ۱/۱۵۷ بوده و به این معناست که به‌ازای بارگذاری هر مدرک در این پایگاه ۱/۱۵۷ برچسب جدید از سوی کاربران به مدارک موجود در این پایگاه تخصیص یافته است.

نمودار (و) در شکل ۱، نشان‌دهنده همبستگی خطی (با اطمینان ۹۹ درصد) بین رشد کل برچسب‌ها (جدید و تکراری) نسبت به افزایش مدارک در پایگاه «آکادمیا» است. ضریب رشد کل برچسب‌ها ۵/۵۲ بوده و به این معناست که به‌طور کلی، به‌ازای بارگذاری هر



مدرک در این پایگاه ۵/۵۲ برجسب از سوی کاربران به مدارک موجود در این پایگاه تخصیص یافته است.

در جدول ۲، ضریب رشد انواع برجسب‌ها نسبت به افزایش مدارک آمده است:

جدول ۲. ضریب رشد انواع برجسب‌ها نسبت به افزایش مدارک

انواع برجسب‌ها	ضریب رشد نسبت به افزایش مدارک
یک-واژه‌ای	۰/۲۰۹
دو-واژه‌ای	۰/۶۰۹
سه-واژه‌ای	۰/۲۰۶
چهار-واژه‌ای و بیشتر	۰/۱۴۳
کل برجسب‌های جدید	۱/۱۵۷
کل برجسب‌های جدید و تکراری	۵/۵۲

همان‌طور که جدول ۲، نشان می‌دهد، در بین گروه‌های مختلف برجسب‌ها بیشترین ضریب رشد مربوط به برجسب‌های دو-واژه‌ای (۰/۶۰۹) و کمترین ضریب رشد مربوط به برجسب‌های چهار-واژه‌ای و بیشتر (۰/۱۴۳) است.

سؤال ۲. رابطه بین تعداد واژه‌های تشکیل‌دهنده برجسب‌های تخصیص‌یافته توسط کاربران با بخش‌های مختلف مقالات (عنوان، چکیده، و کلیدواژه‌های نویسندگان) مرتبط با حوزه علم اطلاعات و دانش‌شناسی از لحاظ انطباق در پایگاه «آکادِمیا» چگونه است؟

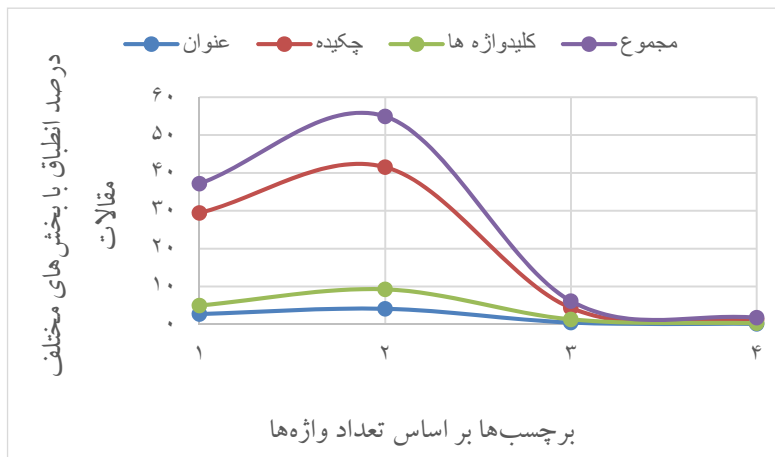
در سؤال دوم پژوهش وجود برجسب‌های یک-واژه‌ای، دو-واژه‌ای، سه-واژه‌ای، و چهار-واژه‌ای و بیشتر در بخش‌های مختلف مقالات (عنوان، چکیده، کلیدواژه‌های نویسندگان) مورد بررسی قرار گرفت که در قالب جدول ۳، و نمودارهای ۱ تا ۲ ارائه شده است:

جدول ۳. انطباق انواع برچسب‌ها با بخش‌های مختلف مقالات (عنوان، چکیده، و کلیدواژه‌های نویسندگان)

گروه برچسب‌ها	جامعه برچسب‌ها		عنوان		چکیده		کلیدواژه‌های نویسندگان		تمامی قسمت‌ها	
	درصد	فراوانی	درصد	فراوانی	درصد	فراوانی	درصد	فراوانی	درصد	فراوانی
یک-واژه‌ای	۱۳۵۴	۱۷/۵۱	۷۸۶	۲/۷۳	۸۴۸۷	۲۹/۴۴	۱۴۴۰	۵	۱۰۷۱۳	۳۷/۱۷
دو-واژه‌ای	۴۰۴۶	۵۲/۳۲	۱۱۹۱	۴/۱۳	۱۱۹۷۵	۴۱/۵۳	۲۶۷۰	۹/۲۶	۱۵۸۳۶	۵۴/۹۲
سه-واژه‌ای	۱۳۷۵	۱۷/۷۸	۱۴۰	/۴۸	۱۲۵۳	۴/۳۴	۳۸۲	۱/۳۳	۱۷۷۵	۶/۱۵
چهار-واژه‌ای و بیشتر	۹۵۸	۱۲/۳۹	۴۵	/۱۶	۳۷۴	۱/۳۰	۸۹	/۳	۵۰۸	۱/۷۶
کل برچسب‌ها	۷۷۳۳	۱۰۰	۲۱۶۲	۷/۵۰	۲۲۰۸۹	۷۶/۶۱	۴۵۸۱	۱۵/۸۹	۲۸۸۳۲	۱۰۰

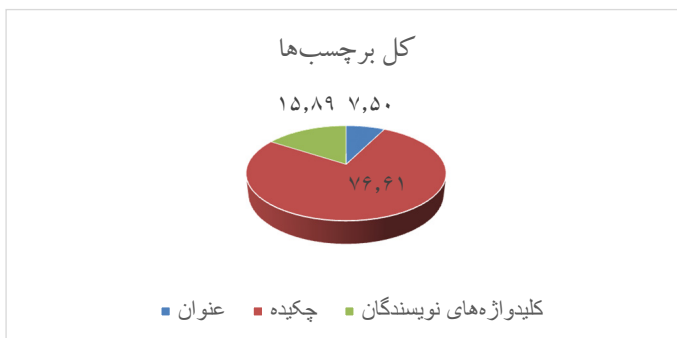
جدول ۳، نشان داد که برچسب‌های یک-واژه‌ای به میزان ۲/۷۳ درصد، دو-واژه‌ای ۴/۱۳ درصد، سه-واژه‌ای ۱۷/۷۸ درصد، و چهار-واژه‌ای و بیشتر به میزان ۱۲/۳۹ درصد با عنوان انطباق دارند. همچنین، برچسب‌های یک-واژه‌ای به میزان ۲۹/۴۴ درصد، دو-واژه‌ای ۴۱/۵۳ درصد، سه-واژه‌ای ۴/۳۴ درصد، و چهار-واژه‌ای و بیشتر به میزان ۱/۳۰ درصد با چکیده انطباق دارند. از سوی دیگر، برچسب‌های یک-واژه‌ای به میزان ۵ درصد، دو-واژه‌ای ۹/۲۶ درصد، سه-واژه‌ای ۱/۳۳ درصد، و چهار-واژه‌ای و بیشتر به میزان ۳ درصد با کلیدواژه‌های نویسندگان انطباق دارند.

نمودار ۱، انطباق گروه‌های مختلف برچسب‌ها را با بخش‌های مختلف مقالات به صورت مقایسه‌ای و بر اساس درصد نشان داده است:



نمودار ۱. مقایسه انطباق گروه‌های مختلف برچسب‌ها با بخش‌های مختلف مقالات

بر اساس نمودار ۱، می‌توان دریافت که برچسب‌های دو-واژه‌ای بیشترین انطباق (۵۴/۹۲ درصد) و برچسب‌های چهار-واژه‌ای و بیشتر کمترین انطباق (۱/۷۶ درصد) را با بخش‌های مختلف مقالات (عنوان، چکیده، و کلیدواژه‌های نویسندگان) داشته است. همچنین، کل برچسب‌ها به میزان ۷/۵ درصد با عنوان، ۷۶/۶۱ درصد با چکیده، و ۱۵/۸۹ درصد با کلیدواژه‌های نویسندگان منطبق بود (نمودار ۲).



نمودار ۲. انطباق کل برچسب‌ها با بخش‌های مختلف مقالات

سؤال ۳. الگوی زمانی بین اولین زمان تخصیص یک برچسب به مقالات مرتبط با حوزه علم اطلاعات و دانش‌شناسی تا استفاده مجدد از آن در واحدهای مختلف زمانی (بر حسب سال نشر) در پایگاه «آکادِمیا» چگونه است؟

در سؤال سوم پژوهش حاضر، استفاده مجدد از برچسب‌ها به‌طور کلی و نیز بر حسب گروه‌های مختلف برچسب (یک-واژه‌ای، دو-واژه‌ای، سه-واژه‌ای، و چهار-واژه‌ای و بیشتر) بر اساس زمان نشر مقالات مجلات (۲۰۲۰-۱۹۷۲) مورد بررسی قرار گرفت که در قالب جدول‌های ۴ تا ۷ و نمودار ۳ ارائه شده است:

جدول ۴. استفاده مجدد از برچسب‌ها به‌طور کلی

نوع برچسب	فراوانی	درصد
برچسب‌های استفاده مجدد	۲۹۹۷	۳۸/۸
برچسب‌های سانسور شده	۴۷۳۶	۶۱/۲
کل برچسب‌ها	۷۷۳۳	۱۰۰

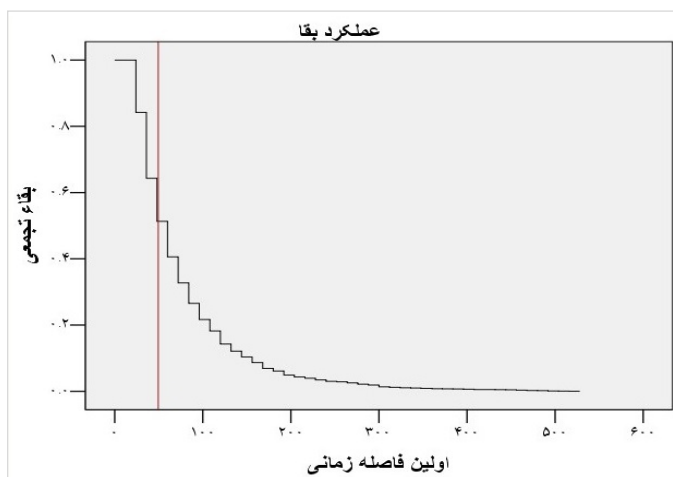
بر اساس جدول ۴، می‌توان دریافت که از ۷۷۳۳ برچسب جدید (غیر تکراری) تعداد ۲۹۹۷ برچسب (۳۸/۸ درصد) مورد استفاده مجدد قرار گرفته و ۴۷۳۶ برچسب (۶۱/۲ درصد) فقط یک بار استفاده شده است.

در جدول ۴، تعداد برچسب‌های سانسور شده نمایانگر برچسب‌هایی هستند که تنها یک بار مورد استفاده قرار گرفته و تاکنون مجدداً استفاده نشده‌اند. بنابراین، این برچسب‌ها از آزمون حذف شده و الگوی زمانی بین اولین زمان تخصیص تا استفاده مجدد با استفاده از برچسب‌هایی که حداقل یک بار مجدداً استفاده شده به کمک تحلیل بقا و جدول عمر در ۱۰ سال اول (جدول ۵) محاسبه شد:

جدول ۵. جدول عمر الگوی زمانی بین اولین زمان تخصیص تا احتمال استفاده مجدد از برچسب‌ها

فاصله‌های زمانی (سال) تا پایان هر فاصله زمانی	فراوانی برچسب‌های باقی‌مانده	فراوانی برچسب‌های مورد استفاده مجدد	شانس استفاده مجدد از برچسب‌های باقی‌مانده در هر فاصله زمانی	انحراف معیار
۱	۲۹۹۷	۴۷۴	۱۶	/۰۰۱
۲	۲۵۲۳	۵۹۵	۲۴	/۰۰۱
۳	۱۹۲۸	۳۹۰	۲۰	/۰۰۱
۴	۱۵۳۸	۳۲۲	۲۱	/۰۰۰
۵	۱۲۱۶	۲۳۵	۱۹	/۰۰۰
۶	۹۸۱	۱۸۵	۱۹	/۰۰۰
۷	۷۹۶	۱۴۷	۱۸	/۰۰۰
۸	۶۴۹	۱۰۳	۱۶	/۰۰۰
۹	۵۴۶	۱۱۸	۲۲	/۰۰۰
۱۰	۴۲۸	۶۵	۱۵	/۰۰۰

یافته‌های جدول ۵، نشان می‌دهد که ۱۶ درصد از برچسب‌ها در سال اول مورد استفاده مجدد قرار گرفته و ۲۴ درصد از برچسب‌هایی که در سال اول مجدداً استفاده نشده‌اند، تا پایان سال دوم مورد استفاده مجدد قرار گرفته‌اند و ...



نمودار ۳. احتمال استفاده مجدد از برچسب‌ها در یک بازه زمانی

در نمودار ۳، خط شاخص بیانگر این یافته است که نیمی از برچسب‌ها تا ۴۹/۴۷ ماه (یعنی حدود ۴ سال) پس از ایجاد برچسب مورد استفاده مجدد قرار گرفته‌اند. در مرحله بعد، استفاده مجدد از برچسب‌ها به تفکیک گروه‌های مختلف برچسب‌ها (یک-واژه‌ای، دو-واژه‌ای، سه-واژه‌ای، و چهار-واژه‌ای و بیشتر) با استفاده از برآورد «کاپلان مایر»<sup>۱</sup> مورد بررسی قرار گرفت که در جدول ۶، ارائه شده است:

جدول ۶. استفاده مجدد از برچسب‌ها بر حسب تعداد واژه‌های تشکیل دهنده آن‌ها

نوع برچسب	فراوانی اولیه	درصد فراوانی اولیه	فراوانی استفاده مجدد	درصد استفاده مجدد
یک-واژه‌ای	۱۳۵۴	۱۷/۵۱	۶۰۴	۲۰/۱۵
دو-واژه‌ای	۴۰۴۶	۵۲/۳۲	۱۷۲۶	۵۷/۵۹
سه-واژه‌ای	۱۳۷۵	۱۷/۷۸	۴۴۱	۱۴/۷۲
چهار-واژه‌ای و ...	۹۵۸	۱۲/۳۹	۲۲۶	۷/۵۴
جمع	۷۷۳۳	۱۰۰	۲۹۹۷	۱۰۰

بر اساس جدول ۶، برچسب‌های دو-واژه‌ای با ۵۷/۵۹ درصد بیشترین و برچسب‌های چهار-واژه‌ای و بیشتر با ۷/۵۴ درصد کمترین استفاده مجدد را داشته‌اند.

1. Kaplan-Meier

نکته دیگری که از یافته‌ها حاصل شد، برچسب‌هایی بود که بیشترین استفاده را در بین بقیه برچسب‌ها داشت که ۱۰ مورد از پر استفاده‌ترین برچسب‌ها به ترتیب، در جدول ۷، آمده است:

جدول ۷. برچسب‌های دارای بیشترین استفاده در بین کاربران پایگاه «آکادِمیا»

ردیف	برچسب	فراوانی استفاده	اولین سال استفاده	آخرین سال استفاده
۱	LIBRARY AND INFORMATION STUDY	۲۷۰۵	۱۹۷۵	۲۰۱۹
۲	INFORMATION SYSTEM	۹۷۷	۱۹۷۵	۲۰۲۰
۳	SCIENTOMETRIC	۴۶۷	۱۹۸۰	۲۰۲۰
۴	INFORMATION RETRIEVAL	۳۳۲	۱۹۷۶	۲۰۲۰
۵	INFORMATION MANAGEMENT	۲۷۹	۱۹۹۱	۲۰۱۹
۶	DESIGN METHODOLOGY	۲۴۲	۱۹۸۱	۲۰۱۵
۷	LIBRARY AND INFORMATION SCIENCE	۲۴۰	۱۹۷۸	۲۰۱۹
۸	HUMAN	۲۲۴	۱۹۸۱	۲۰۱۷
۹	INFORMATION TECHNOLOGY	۲۱۶	۱۹۷۹	۲۰۱۷
۱۰	CASE STUDY	۲۰۷	۱۹۸۸	۲۰۱۵

نکته مهم در جدول ۷، میزان فراوانی و مدت‌زمان حضور هر اصطلاح در مجموعه مورد مطالعه است که در مورد اهمیت آن در بخش بحث و نتیجه‌گیری به تفصیل صحبت خواهد شد.

##### ۵. بحث و نتیجه‌گیری

در سؤال اول پژوهش، بررسی رابطه بین رشد انواع برچسب‌ها (یک-واژه‌ای، دو-واژه‌ای، سه-واژه‌ای و چهار-واژه‌ای و بیشتر) و افزایش مدارک حاکی از همبستگی خطی بین آن‌ها بود (شکل ۱). بیشتر تمرکز کاربران ابتدا بر اختصاص برچسب‌های دو-واژه‌ای و سپس، برچسب‌های دیگر بود. دلیل یافته مذکور را شاید بتوان در این نکته دانست که خاص‌تر بودن برچسب‌های چند-واژه‌ای نسبت به یک-واژه‌ای، باعث انتقال بهتر معنا شده و از ابهام معنایی جلوگیری می‌کند و کاربرد برچسب‌های یک-واژه‌ای به دلیل ابهام در معنا کاربرد کمتری توسط کاربران داشته است. نکته دیگری که شاید بتوان در مورد

آن صحبت کرد این که در حوزه علم اطلاعات و دانش‌شناسی که حوزه‌ای میان‌رشته‌ای است، اصطلاحات بیشتر به صورت دو-واژه‌ای بوده و شکل غالب اصطلاحات ترکیبی از دو واژه است. این مطلب را Choi & Syn (2016) با بررسی برجسب‌های کاربران پایگاه «ناینز» در حوزه تاریخ (که حوزه‌ای مستقل است) دریافتند که از ۱۵۴۰ برجسب، ۷۵/۹۰ درصد یک-واژه‌ای و ۲۴/۱ درصد چند-واژه‌ای بودند که با نتایج پژوهش حاضر همخوانی ندارد. البته، تأیید چنین نکته‌ای نیازمند پژوهش‌های بیشتر در این زمینه است. به هر حال، در صورت صحت این موضوع می‌توان از آن در ساخت نظام‌های سازماندهی دانش از جمله اصطلاحنامه‌ها، سرعنوان‌های موضوعی، و هستی‌شناسی‌ها و نیز مبحث نمایه‌سازی و به‌ویژه، نمایه‌سازی ماشینی که نیاز به برنامه‌نویسی برای ترکیب کلمات دارد، بهره جست؛ هر چند نظام‌های برجسب‌زنی اجتماعی و به‌ویژه پایگاه «آکادمیا» نیز که رشته علم اطلاعات و دانش‌شناسی را پوشش داده است، در قسمت پیشنهادات برجسب‌ها به کاربران می‌توانند از این نکته بهره‌جویی کنند.

همچنین، آشکار شد که بین رشد کل برجسب‌های تخصیص‌یافته به مقالات و افزایش مدارک رابطه خطی وجود دارد (شکل ۱). در این راستا «مارلو» و همکاران با مطالعه رشد مجموعه برجسب‌ها با استفاده از داده‌های پایگاه «فلیکر» (پایگاه بارگذاری عکس‌ها) در پژوهش خود نشان دادند که افزایش برجسب‌های جدید به صورت خطی با افزایش مدارک جدید (عکس‌ها) در این پایگاه مرتبط است و به ازای هر عکس، ۲ برجسب جدید توسط کاربران اضافه می‌گردد (Marlow et al. 2006). «هکنر، مولباخر و ولف» نیز با بررسی رشته فناوری رایانه در پایگاه «کانوتی» (مربوط به مدیریت منابع کتابشناختی) بیان داشتند که به ازای هر مدرک، کمی بیش از ۲ برجسب تخصیص یافته است (Heckner, Mühlbacher & Wolff 2008). همچنین، یافته‌های پژوهش «زوبیاگا، مارتینز و فرسنو» نشان داد که رشد مدارک و کل برجسب‌های تخصیص‌یافته از سوی کاربران از یک رشد لگاریتمی پیروی کرده است (Zubiaga, Mart'inez & Fresno 2011b). بر این اساس، در پایگاه del.icio.us به ازای هر مدرک وب ۳۳/۳۵ برجسب، در «لایبری تینگ» به ازای هر کتاب ۱۴/۵۳ برجسب، و در «گودریدز» به ازای هر کتاب ۱۳/۳۳ برجسب از سوی کاربران تخصیص یافته بود که با یافته‌های پژوهش حاضر همخوان نیست.

در واقع، تعداد برجسب‌های تخصیص‌یافته به مقالات در پایگاه «آکادمیا» به میزان ۲/۵ تا ۶ برابر کمتر از برجسب‌های تخصیص‌یافته به مدارک در سه پایگاه مورد بررسی

در پژوهش Zubiaga, Mart'inez, & Fresno (2011b) دلیل این نکته را شاید بتوان در وهله اول مربوط به عمر پایگاه‌ها دانست، چنانکه پایگاه del.icio.us در سال ۲۰۰۳، پایگاه «لایبرری ینگ» در سال ۲۰۰۵، و پایگاه «گودریذ» در ژانویه ۲۰۰۷ ایجاد گردیده است و نسبت به پایگاه «آکادمیا» که در اواخر سال ۲۰۰۸ راه‌اندازی شده، عمر بیشتری دارند؛ هرچند بافت محتوایی و مخاطب پایگاه‌های مورد مطالعه را نیز نمی‌توان نادیده گرفت. توضیح این که پایگاه del.icio.us مربوط به «بوک‌مارک»<sup>۱</sup> صفحات وب و پایگاه‌های «لایبرری ینگ» و «گودریذ» مربوط به اطلاعات کتابشناختی کتاب‌های عمدتاً عامه‌پسند و عمومی هستند، در حالی که پایگاه «آکادمیا» بیشتر در زمینه مقالات علمی تخصصی و دانشگاهی فعالیت می‌کند. در این زمینه، Santos-Neto et al. (2014) نیز در پژوهش خود بیان داشتند که مخاطب هدف و نوع محتوایی که برچسب دریافت کرده، نقش مهمی در رفتار برچسب‌گذاری کاربران ایفا می‌کند.

نکته دیگر اینکه رشد خطی با همبستگی مثبت نشان‌دهنده رشد پیوسته نظام برچسب‌گذاری به موازات تولید محتوا و بارگذاری مدارک در پایگاه «آکادمیا» بوده و حاکی از علاقه کاربران به برچسب‌گذاری و سازماندهی مقالات بارگذاری شده است. در واقع، چنین نکته‌ای بیانگر ارزشمندی سرمایه‌گذاری بر روی زیرساخت‌های نظام‌های برچسب‌گذاری اجتماعی از جنبه سازماندهی اطلاعات و دانش است.

یافته‌های سؤال دوم پژوهش نیز بار دیگر گرایش کاربران پایگاه مورد مطالعه و نیز نویسندگان مقالات را به استفاده از اصطلاحات و موضوعات دو-واژه‌ای نشان داد. همچنین، کل برچسب‌ها به میزان ۷۶/۶۱ درصد با چکیده منطبق بود (جدول ۳)، و این یافته حاکی از جایگاه برجسته چکیده مقالات در اخذ اصطلاحات نمایه‌ای هنگام نمایه‌سازی و نیز رتبه بالای چکیده در رتبه‌بندی و اولویت‌بخشی به قسمت‌های مختلف مقالات هنگام بازیابی اطلاعات است که می‌توان در الگوریتم‌های بازیابی اطلاعات به کار گرفت. این نکته همچنین نشان‌دهنده آن است که از آنجا که چکیده مقالات حاوی بخش عمده محتوای موضوعی آن‌ها و در واقع، آینه تمام‌نمای مدرک است، نه تنها مورد توجه نمایه‌سازان هنگام نمایه‌سازی قرار می‌گیرد، بلکه این ویژگی سبب شده که کاربران شبکه‌های اجتماعی علمی نیز هنگام برچسب‌گذاری مقالات توجه ویژه‌ای به چکیده داشته باشند. در عین حال، یافته‌های پژوهش حاضر در این خصوص نشان داد که میزان انطباق

1. bookmark



برجسب‌ها با عنوان کمترین میزان انطباق (۷/۵ درصد) در بین پژوهش‌های قبلی بوده است، چنانکه Heymann, Koutrika, & Garcia-Molina (2008) با مقایسه برجسب‌های اجتماعی پایگاه del.icio.us (بدون یکدست‌سازی داده‌ها) نشان دادند که ۱۶ درصد از برجسب‌ها در عناوین صفحات وب ظاهر شدند که دلیل اختلاف با یافته‌های پژوهش حاضر را هم می‌توان در نوع پایگاه و اهداف برجسب‌گذاری کاربران دانست و هم در عدم یکدست‌سازی داده‌ها. در این میان، Heckner, Mühlbacher & Wolff (2008) نیز با انجام پژوهشی در مورد مقالات حوزه فناوری رایانه و با بررسی برجسب‌های شبکه اجتماعی «کانوتی» (بدون یکدست‌سازی داده‌ها) دریافتند که ۴۹ درصد از برجسب‌ها از کلیدواژه‌های عنوان انتخاب شده بود و دلیل اختلاف با یافته‌های پژوهش حاضر را هم می‌توان در حوزه مورد بررسی و هم در عدم یکدست‌سازی داده‌ها دانست. همچنین، Haustein & Peters (2012) با تحلیل ۴۵ مجله علم فیزیک در قالب ۷۲۴ مقاله در بازه زمانی ۲۰۰۴-۲۰۰۸ و برجسب‌های مربوطه در پایگاه‌های «سایت یولایک»، «کانوتی» و «بیسونومی» (بعد از ترکیب و یکدست‌سازی داده‌ها مثل ریشه‌یابی، یکدست کردن انگلیسی بریتانیایی و آمریکایی، و حذف کاراکترهای خاص) نشان دادند که ۶۶ درصد از برجسب‌ها با واژگان عنوان مطابقت داشت که دلیل اختلاف با یافته‌های پژوهش حاضر را می‌توان در حوزه مورد بررسی و تعداد پایگاه‌های تحت مطالعه دانست. از طرفی، Vaidya & Harinarayana (2019) با مقایسه برجسب‌های تخصیص‌یافته به مقالات رشته علوم دریایی در پایگاه «سایت یولایک» (بعد از یکدست‌سازی مثل جمع و مفرد، و کاراکترهای خاص مانند پرانتز، اعداد، اختصارات، کلمات ترکیبی و ...) آشکار کردند که میزان همپوشانی بین برجسب‌های کاربران با واژه‌های عنوان مقالات ۳۶/۲۹ درصد است که دلیل اختلاف با یافته‌های پژوهش حاضر را شاید بتوان در حوزه و رشته مورد مطالعه دانست. در واقع، علوم پایه و مهندسی نسبت به علوم انسانی و علوم اجتماعی از این لحاظ متفاوت هستند، چنانکه Borko & Bernier (1978) نیز در کتاب خود با اشاره به نمایه‌سازی حرفه‌ای بیان داشتند که به‌طور کلی، ۳۴ تا ۸۶ درصد اصطلاحات نمایه‌ای که از سوی نمایه‌سازان به مدارک حوزه‌های علوم پایه و مهندسی تخصیص داده می‌شود، از واژگان عنوان انتخاب می‌شود و این مقدار در مورد علوم انسانی و اجتماعی کمتر است. این نکته خود بیانگر این است که عنوان مدارک در حوزه علوم پایه و مهندسی نسبت به علوم انسانی و اجتماعی محتوای موضوعی مدارک را بیشتر و بهتر انعکاس می‌دهد.

از سوی دیگر، یافته‌های پژوهش حاضر تا حدودی با پژوهش Hausteijn, & Peters (2012) که انطباق برچسب‌ها با چکیده‌های مقالات علم فیزیک را ۷۷/۶۶ درصد بیان کردند، منطبق است، اما با یافته‌های پژوهش Heckner, Mühlbacher & Wolff (2008) که این انطباق را با چکیده‌های مقالات حوزه فناوری رایانه ۹ درصد بیان کردند، همخوانی ندارد و دلیل این اختلاف را شاید بتوان هم در عدم یکدست‌سازی داده‌ها و هم در حوزه مورد بررسی دانست.

همچنین یافته‌های پژوهش حاضر، از انطباق ۱۵/۸۹ درصدی برچسب‌ها با کلیدواژه‌های نویسندگان حکایت داشت که تا حدودی با یافته‌های پژوهش «قنوتی» و همکاران (۱۳۹۷) که این مطابقت را در مورد پایگاه «مدلی» و در حوزه علوم تربیتی (با بررسی ۲۰ مجله، ۴۹۹ مقاله، ۱۶۷۱ برچسب، پس از یکپارچه‌سازی داده‌ها) ۱۵ درصد اعلام کردند، همخوانی دارد.

در عین حال، شاید نکته دیگری که بتوان در مورد اختلاف یافته‌های پژوهش‌های قبلی با پژوهش حاضر بیان داشت، این است که نمی‌توان نسخه و دستورالعمل واحدی را برای تمام حوزه‌ها و رشته‌های علمی و نیز برای تمام نظام‌های برچسب‌گذاری صادر کرد. در واقع، انطباق برچسب‌ها با بخش‌های مختلف مقالات به عواملی از جمله حوزه مورد مطالعه، نوع نظام برچسب‌گذاری (عمومی، تخصصی و ...)، روش پژوهش، نوع مدرک (مقاله، صفحه وب، و ...) و غیره بستگی داشته و نیازمند پژوهش‌های بیشتر در این زمینه است.

در سؤال سوم پژوهش، مشخص شد که به‌طور کلی، ۳۸/۸ درصد برچسب‌ها مورد استفاده مجدد قرار گرفته است. در این راستا، Farooq et al. (2007) با محاسبه نرخ استفاده مجدد از برچسب‌ها در پایگاه «سایت یولایک» (۲ سال و ۴ ماه) دریافتند که متوسط رخداد استفاده مجدد از هر برچسب ۳/۹ درصد است که همخوان با پژوهش حاضر نیست و مدت زمان مورد مطالعه می‌تواند در کمیت استفاده مجدد از برچسب‌ها تأثیرگذار باشد. افزون بر این، Yin et al. (2011) با بررسی سه نظام برچسب‌گذاری (del.icio.us، بیسونومی، و فلیکر) دریافتند که از بین ۶۶۷ برچسب غیر تکراری در ۸۶/۱ درصد موارد کاربران برچسب‌های قبلی را مورد استفاده مجدد قرار می‌دهند که نظام‌های برچسب‌گذاری با اهداف متفاوت و نیز نمونه مورد بررسی دلیل اختلاف با یافته‌های پژوهش حاضر است. همچنین، Choi & Syn (2016) با بررسی برچسب‌های کاربران پایگاه «ناینز» در حوزه تاریخ

(۸۳۱۰ برچسب، ۱۵۴۰ برچسب غیر تکراری) دریافتند که ۴۶/۵۶ درصد از برچسب‌ها بیش از یک‌بار استفاده شده‌اند و این با یافته‌های پژوهش حاضر همخوانی ندارد و دلیل آن را می‌توان در حوزه و نمونه مورد مطالعه دانست.

از سوی دیگر، هرچند در پژوهش‌های پیشین میزان استفاده مجدد از برچسب‌ها به تفکیک تعداد واژه‌های تشکیل‌دهنده آن‌ها مورد بررسی قرار نگرفته بود؛ اما در پژوهش حاضر نقش برجسته برچسب‌های دو-واژه‌ای در نمایه‌سازی اجتماعی توسط کاربران در حوزه علم اطلاعات و دانش‌شناسی آشکار گردید و خود دلیل دیگری است بر این که در این حوزه، ماهیت و شکل اصطلاحات به سمت دو-واژه‌ای گرایش داشته و ترجیح کاربران و در واقع متخصصان علم اطلاعات و دانش‌شناسی بر اصطلاحات دو-واژه‌ای بوده است. همچنین، در سؤال سوم پژوهش با کمک تحلیل بقا و جدول عمر (جدول ۵) آشکار شد که ۱) با کمک روش تحلیل بقا و جدول عمر می‌توان بقای عمر موضوعات را در حوزه‌ها و رشته‌های گوناگون علمی مشخص کرد و زمان ظهور و افول آن‌ها را تعیین نمود؛ و ۲) در حوزه علم اطلاعات و دانش‌شناسی ۱۶ درصد از موضوعات که در سال اول مورد استفاده مجدد واقع شده (موضوعات داغ)، می‌توان انتظار داشت که دامنه آن‌ها به سال بعد نیز کشیده شود و بدین ترتیب، می‌توان سیر استفاده از موضوعات علمی را در آینده مشخص کرد و اینکه پژوهش‌های حوزه علم اطلاعات و دانش‌شناسی در حال حاضر به چه سمت و سویی در جریان است.

در بخش نهایی سؤال سوم، جدولی (جدول ۷) حاصل شد که در آن انواع برچسب‌های مورد استفاده کاربران به همراه فراوانی و مدت‌زمان حضورشان وجود داشت. بر اساس جدول مذکور می‌توان با توجه به فراوانی استفاده و مدت‌زمان حضور هر برچسب در مجموعه مورد بررسی، به فرمولی برای تعیین وزن برچسب‌ها جهت استفاده در بازیابی اطلاعات در حوزه مذکور دست یافت و از آن در الگوریتم‌های بازیابی پایگاه‌های اطلاعاتی بهره جست. همچنین، بر اساس جدول مذکور می‌توان انتظار داشت موضوعاتی که پایه و اساس یک حوزه یا مجموعه هستند، پیوسته در متون آن حوزه حضور داشته باشند.

در پایان می‌توان گفت برچسب‌ها اغلب دربرگیرنده دیدگاه و بافت ذهنی برچسب‌گذاران است و از این جهت برچسب‌های اجتماعی را می‌توان غنی‌تر از اصطلاحات نمایه‌ای نمایه‌سازان دانست. همچنین، وجود اجماع قابل توجه کاربران در مورد اصطلاحاتی معین بیانگر آن است که الگوهای جدید از برچسب‌های کاربران حداقل

تا حدی با مفاهیم نمایه‌سازی حرفه‌ای در مورد محتوای مدارک سازگار است و با تمرکز بر پرکاربردترین برجسب‌ها و توزیع پایدار آن‌ها می‌توان به فرمولی برای تعیین وزن و حتی طرح‌های رده‌بندی و طبقه‌بندی دست یافت. افزون بر این، از فعالیت‌های کاربران در شبکه‌های اجتماعی به‌منظور افزایش کیفیت پیشنهادات در نظام‌های برجسب‌گذاری جمعی می‌توان بهره برد. نکتهٔ دیگر این‌که، بین نمایه‌سازی حرفه‌ای و برجسب‌گذاری کاربر پیوستگی وجود دارد و این دو نسبت به هم بیگانه نیستند. این پیوستگی می‌تواند پایه و اساسی برای یک سیستم مکمل دسترسی موضوعی که موجب غنی شدن نمایه‌سازی حرفه‌ای است، باشد. تفاوت‌های موجود بین برجسب‌ها و اصطلاحات نمایه‌ای بیانگر آن است که برجسب‌گذاری کاربران از اهداف سنتی دسترسی موضوعی فراتر رفته و بیانگر رابطه‌ای پویا بین مدرک و کاربر است که می‌تواند منجر به شیوه‌های جدید مدل‌سازی برای دسترسی موضوعی گردد.

در این میان آنچه که بیشتر ضروری می‌نماید، انجام پژوهش‌های بیشتر و هماهنگ‌تر برای دستیابی به الگوهای مطلوب و اطمینان‌بخش در این زمینه است. تنها در این صورت است که الگوهای منطبق با رفتار اطلاع‌یابی کاربران پایگاه‌های اطلاعاتی به‌منظور بهبود حوزهٔ سازماندهی دانش و به‌ویژه نمایه‌سازی، نمایه‌سازی ماشینی، و نظام‌های سازماندهی دانش حاصل خواهد شد.

### پیشنهاد‌های پژوهشی

۱. انجام پژوهش مشابه برای رشته‌ها و حوزه‌های علمی دیگر به‌صورت موردی و مقایسه‌ای در پایگاه «آکادِمیا»؛
۲. انجام پژوهش مشابه بر روی شبکه‌های اجتماعی علمی دیگر مانند «مندلی»، «بیسونومی» و ... به‌صورت موردی و مقایسه‌ای؛
۳. انجام پژوهش مشابه در مورد رشد و توسعهٔ موضوعات گوناگون در پایگاه‌های مختلف علمی؛
۴. انجام پژوهش مشابه با توجه به زمان واقعی برجسب‌گذاری کاربران در پایگاه‌های اجتماعی علمی؛
۵. مقایسهٔ پایگاه اجتماعی مقالات و پایگاه اجتماعی کتاب‌ها؛
۶. بررسی چرایی گرایش به سمت اصطلاحات دو-واژه‌ای در علم اطلاعات و دانش‌شناسی.

## فهرست منابع

- پاول، رونالد. ۱۹۹۱. *روش‌های اساسی پژوهش برای کتابداران*. ترجمه نجلا حریری. ۱۳۸۵. تهران: دانشگاه آزاد اسلامی، واحد علوم و تحقیقات.
- توکلی‌زاده راوری، محمد. ۱۳۸۶. توزیع آماری رشد موضوعات علوم پزشکی در طول زمان: تحلیل همبستگی بین توسعه MeSH و رشد Medline. *مدیریت اطلاعات سلامت*، ۴ (۲): ۱۹۲-۱۸۵. (دسترسی در ۱۵ آذر ۱۴۰۰) <https://www.sid.ir/fa/journal/ViewPaper.aspx?ID=81621>
- حبیبی، آرش. ۱۳۹۲. روش‌های نمونه‌گیری آماری. <https://parsmodir.com/db/research/sampling.php> «دسترسی در ۸ تیر ۱۴۰۰».
- خادمیان، مهدی، مرتضی کوبی، و فریده عصاره. ۱۳۹۶. امکان جایگزینی یا تکمیل سرعنوان‌های موضوعی کتابخانه کنگره با برجسب‌های اجتماعی لایبرری ۳ینگ (نمونه پژوهی حوزه‌های علوم انسانی، علوم اجتماعی و علوم طبیعی). *مطالعات ملی کتابداری و سازماندهی اطلاعات* ۲۸ (۲): ۴۳-۲۹. (دسترسی در ۲۳ آذر ۱۴۰۰) [http://nastinfo.nlai.ir/article\\_1720.html](http://nastinfo.nlai.ir/article_1720.html)
- قنواتی، مریم، علیرضا نوروزی، مریم ناخدا، و اشکان خطیر. ۱۳۹۷. بررسی میزان تطابق زبان نمایه‌سازان، نویسندگان، و برجسب‌گذاران در پایگاه اطلاعاتی اریک و مندلی. *پژوهشنامه پردازش و مدیریت اطلاعات* ۳۳ (۴): ۱۷۵۰-۱۷۲۹.
- محمدپور، معصومه، عصمت مومنی، و مهدی علیپور حافظی. ۱۳۹۵. بررسی میزان همپوشانی میان برجسب‌های مقالات کتابداری در پایگاه مردمی سایت یولایک و توصیفگرهای اختصاص یافته به همان مقالات در پایگاه لیستا. *فصلنامه بازیابی دانش و نظام‌های معنایی* ۳ (۸): ۱۴۰-۱۲۱. (دسترسی در ۱۵ آذر ۱۴۰۰) [https://jks.atu.ac.ir/article\\_7247.html](https://jks.atu.ac.ir/article_7247.html)
- موحدیان، قاسم، احمد شعبانی، مظفر چشمه‌سهرابی، و عاصفه عاصمی. ۱۳۹۹. دیگری‌سازی در نظام‌های سازماندهی منابع کتابخانه‌ای در تقابل با برجسب‌گذاری اجتماعی. *فصلنامه مطالعات ملی کتابداری و سازماندهی اطلاعات* ۳۱ (۳): ۷۱-۵۸. (دسترسی در ۲۳ آذر ۱۴۰۰) [http://nastinfo.nlai.ir/article\\_2419.html](http://nastinfo.nlai.ir/article_2419.html)

## References

- About Academia.edu. 2021. [www.academia.edu/about](http://www.academia.edu/about) (accessed Sept. 19, 2021).
- Academia.edu Traffic Statistics. 2021. <https://www.alexa.com/siteinfo/academia.edu> (accessed Sept. 05, 2021).
- Alexa.com. 2021. Academia.edu Traffic Statistics. <https://www.alexa.com/siteinfo/academia.edu> (accessed September 05, 2021).
- Alexiev, V., M. Breu, J. de Bruijn, D. Fensel, R. Lara, & H. Lausen (Eds.). 2005. *Information integration with ontologies: Experiences from an industrial showcase*. Chichester, England: Wiley & Sons.
- Borko, H. & C. L. Bernier. 1978. *Indexing Concepts and Methods*. New York: Academic Press.
- Chan, L. M. 2009. Social bookmarking and subject indexing. Paper presented at the IFLA Satellite Pre-Conference of the Classification and Indexing Section, *Looking at the Past and Preparing for the Future*. Florence, Italy. <http://www.webcitation.org/60msdmuFN> (accessed March 26, 2020).

- Choi, Y., & S. Y. Syn. 2016. Characteristics of tagging behavior in digitized humanities online collections. *Journal of the Association for Information Science and Technology* 67 (5): 1089–1104.
- Farooq, U., T. G. Kannampallil, Y. Song, C. H. Ganoë, J. M. Carroll, & C. L. Gilles. 2007. Evaluating tagging behavior in social bookmarking systems: Metrics and design heuristics. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work*. NewYork: ACM Press: 351-360.
- Guo, S. 2010. *Survival analysis*. NewYork: Oxford University Press.
- Haustein, S., & I. Peters. 2012. Using social bookmarks and tags as alternative indicators of journal content description. *First Monday* 17 (11). <http://firstmonday.org/ojs/index.php/fm/article/view/4110/3357> (accessed Nov. 14 2020).
- Heckner, M., S. Mühlbacher, & C. Wolff. 2008. Tagging Tagging. Analyzising User Keywords in Scientific Bibliography Management Systems. *Journal of Digital Information* 9 (2): 1–19.
- Heymann, P., G. Koutrika, & H. Garcia-Molina. 2008. Can social bookmarking improve web search? *Proceedings of the Second International Conference on Web Search and Web Dat Mining*. NewYork: ACM Press: 195-206.
- Kipp, M. E. 2006. Complementary or Discrete Contexts in Online Indexing: A Comparison of User, Creator and Intermediary words. *Canadian Journal of Information and Library Science* 30 (4). <http://dlist.sir.arizona.edu/1533/01/mkipp-caispaper.pdf>. (accessed Dec, 23, 2020).
- \_\_\_\_\_. 2007. Tagging Practices on Research Oriented Social Bookmarking Sites. *Proceedings of the Annual Conference of CAIS/ Actes du congrès annuel de l ACSI* 30. <http://dlist.sir.arizona.edu/2027/01/kipp%5F2007.pdf> (accessed Dec. 23, 2020).
- Kipp, M. E. 2011. Controlled vocabularies and tags: An analysis of research methods. In Smiraglia, Richard P., ed. *Proceedings from North American Symposium on Knowledge Organization*, 3: 23-32. <https://journals.lib.washington.edu/index.php/nasko/article/view/12787/11268> (accessed Sept. 20, 2020).
- Kleinbaum, D. G. & M. Klein. 2012. *Survival Analysis: A Self-Learning Text*, 3rd Edition. New York: Springer.
- Lee, D. H., & T. Schleyer. 2010. A comparison of MeSH terms and CiteULike social tags as metadata for the same items. In *Proceedings of the First ACM International Conference on Health Informatics*: 445–448. New York: ACM Press.
- \_\_\_\_\_. 2012. Social tagging is no substitute for controlled indexing: A comparison of medical subject headings and CiteULike tags assigned to 231,388 papers. *Journal of the American Society for Information Science and Technology* 63 (9): 1747- 1757. DOI: <http://10.1002/asi2265310.1002/asi22653>
- Lyer, H. & L. Bungo. (2011). An examination of semantic relationships between professionally assigned metadata and user-generated tags for popular literature in complementary and alternative medicine. *Information Research* 16 (3) paper 482. Retrieved from <http://InformationR.net/ir/16-3/paper482.html> (accessed Sept. 22, 2022).
- Marlow, C., M. Naaman, D. Boyd, & M. Davis. 2006. HT06, tagging paper, taxonomy, Flickr, academic article, to read. In *Proceedings of the 17th ACM Conference on Hypertext and Hypermedia*. [www.danah.org/papers/Hypertext2006.pdf](http://www.danah.org/papers/Hypertext2006.pdf). (accessed Sept. 20, 2020).
- Miner, G., D. Delen, J. Elder, A. Fast, T. Hill, & R.A. Nisbet. 201. *Practical text mining and statistical analysis for non-structured text data applications*. Oxford: Academic Press.
- Rafferty, P. 2018. Tagging. *Knowledge organization* 45 (6): 500-516. Doi: 10.5771/0943-7444-2018-6-500

- Rorissa, A. 2010. A comparative study of flickr tags and index terms in a general image collection. *Journal of the American Society for Information Science and Technology* 61 (11): 2230–2242. doi:10.1002/asi.21401.
- Santos-Neto, E., D. Condon, N. Andrade, A. Iamnitchi, & M. Ripeanu. 2014. Reuse, temporal dynamics, interest sharing, and collaboration in social tagging systems. *First Monday* 19 (7). <https://firstmonday.org/ojs/index.php/fm/article/view/4994/410> (accessed Sept. 4, 2020).
- Vaidya, P. & S. N. Harinarayana. 2019. Comparison of User-generated Tags with Subject Descriptors, Author Keywords, and Title Terms of Scholarly Journal Articles: A Case Study of Marine Science. *Journal of Information Science Theory and Practice* 7 (1): 29-38. DOI: 10.1633/JISTaP.2019.7.1.3.
- Yin, D., L. Hong, Z. Xue, & B. D. Davison. 2011. Temporal dynamics of user interests in tagging systems. In: *Proceedings of the twenty fifth AAAI conference on artificial intelligence*. <http://www.cse.lehigh.edu/~brian/pubs/2011/AAAI/temporal-dynamics.pdf> (accessed Sept. 10, 2020)
- Zubiaga, A., R. Mart´inez, V. Fresno 2011b. Analyzing tag distributions in folksonomies for resource classification. In *Proceedings of the 5th international conference on knowledge science, engineering and management, (KSEM'11)*: 91–102: Irvine: Springer-Verlag.

#### رسول سعادت

متولد ۱۳۶۱، دانشجوی دکتری علم اطلاعات و دانش‌شناسی دانشگاه اصفهان است. ایشان هم‌اکنون مسئول بخش مجموعه‌سازی و سازماندهی اطلاعات در اداره کتابخانه‌ها و اسناد پژوهشگاه علوم و فرهنگ اسلامی است. سازماندهی اطلاعات و علم‌سنجی از جمله علایق پژوهشی وی است.



#### احمد شعبانی

متولد ۱۳۳۵، دارای مدرک دکتری در رشته علم اطلاعات و دانش‌شناسی است. ایشان هم‌اکنون استاد گروه علم اطلاعات و دانش‌شناسی دانشگاه اصفهان است. مدیریت دانش و روش تحقیق از جمله علایق پژوهشی وی است.



#### عاصفه عاصمی

متولد سال ۱۳۴۹، دارای مدرک تحصیلی دکتری در رشته بی‌زینس انفورماتیک از دانشگاه کورونوس بوداپست است. ایشان هم‌اکنون عضو هیئت علمی و محقق دانشگاه کورونوس است. سیستم‌های اطلاعاتی، تجارت هوشمند، سیستم‌های سرمایه‌گذاری، کریپتوکارنسی و سیستم‌های پیشنهاددهنده سرمایه‌گذاری از جمله علایق وی است.



### مهرداد چشمه‌سهرایی

متولد ۱۳۵۳، دارای مدرک تحصیلی علوم اطلاعات و ارتباطات از دانشگاه استاندال فرانسه است. ایشان هم‌اکنون دانشیار گروه علم اطلاعات و دانش‌شناسی دانشگاه اصفهان است. وب معنایی، ذخیره و بازیابی اطلاعات، داده‌کاوی، سنجش و ارزیابی علم و پژوهش، علم‌شناسی و اخلاق علمی از جمله علایق پژوهشی وی است.



### محمد توکلی‌زاده راوری

متولد ۱۳۴۷، دارای مدرک تحصیلی دکتری در رشته علم‌سنجی از دانشگاه همبولت برلین است. ایشان هم‌اکنون دانشیار دانشگاه یزد است. علم‌سنجی، فناوری‌سنجی، تحلیل محتوا، نمایه‌سازی خودکار، برنامه‌نویسی رایانه در حوزه علم اطلاعات و دانش‌شناسی از جمله علایق پژوهشی وی است.

