

Disambiguation of Knowledge Domains by Emphasis on Medicine Fields: Systematic Review

Hamideh Asadi*

PhD in Library and Information Science; Information Retrieval; University of Tehran; Tehran, Iran Email: Asadi1366@gmail.com

Nader Naghshineh

PhD in Library and Information Science; Associated Professor; University of Tehran; Tehran, Iran Email: nnaghsh@ut.ac.ir

Molouk Sadat Hosseini Beheshti

PhD in Linguistics; Associated Professor; Terminology & Ontology Research Group; Iranian Research Institute for Information Science and Technology (IranDoc); Tehran, Iran; Email: beheshti@irandoc.ac.ir

Received: 06, Dec. 2023 Accepted: 15, May 2024

Abstract: The aim of this paper is to represent a comprehensive expression of disambiguation in knowledge domains based on published researches.

In order to facilitate the retrieval of studies in disambiguation field to conduct a systematic review, databases were searched by a systematic search. The obtained various aspects of the literature review of the retrieved researches became the basis for the representation of a comprehensive expression of the disambiguation in the synthesis. In addition, a qualitative analysis was performed for all retrieved studies using a critical appraisal checklist in the field of library and information science, and the validity of the researches to be included in this synthesis was investigated.

Systematic review presents a systematic framework based on the items of researches in the field of disambiguation. The items include the research methodology, data gathering tool and other elements. The hidden subjects in the literature review were extracted and classified in some class.

The investigation of disambiguation topic indicated that the disambiguation is done in different subjects, study disciplines and by variety goals. Choosing the proper methods of research and data collection and tools depends on the data type. Additionally, disambiguation is needed in various fields to obtain more evidences in it and to help researchers in special fields.

Keywords: Ambiguity, Classification, Information Science, Knowledge Management, Systematic Reviews

* Corresponding Author

Iranian Journal of
**Information
Processing and
Management**

Iranian Research Institute
for Information Science and Technology
(IranDoc)

ISSN 2251-8223

eISSN 2251-8231

Indexed by SCOPUS, ISC, & LISTA

Vol. 40 | No. 2 | pp. 413-438

Winter 2025

<https://doi.org/10.22034/ijpm.2024.713420>



ابهام‌زدایی از حوزه‌های دانشی با تأکید بر حوزه‌های پزشکی:

مرور نظام‌مند

حمیده اسدی

دکتری علم اطلاعات و دانش‌شناسی؛
بازیابی اطلاعات؛ دانشگاه تهران؛ تهران، ایران؛
Asadi1366@gmail.com

نادر نقشینه

دکتری علم اطلاعات و دانش‌شناسی؛ دانشیار؛
دانشگاه تهران؛ تهران، ایران؛
mnaghsh@ut.ac.ir

ملوک‌السادات حسینی بهشتی

دکتری زبان‌شناسی همگانی؛ دانشیار؛
پژوهشگاه علوم و فناوری اطلاعات (ایرانداک)؛
تهران، ایران؛
beheshti@irandoc.ac.ir



دریافت: ۱۴۰۲/۰۹/۱۵ | پذیرش: ۱۴۰۳/۰۲/۲۶ | مقاله برای اصلاح به مدت ۴ روز نزد پدیدآوران بوده است.

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

شاپا (چاپی) ۲۲۵۱-۸۲۲۳

شاپا (الکترونیکی) ۲۲۵۱-۸۲۲۳

نمایه در SCOPUS و ISI، LISTA

jipm.irandoc.ac.ir

دوره ۴۰ | شماره ۲ | صص ۴۱۳-۴۳۸

زمستان ۱۴۰۳

<https://doi.org/10.22034/jipm.2024.713420>



چکیده: هدف این مرور، ارائه صورت‌بندی جامعی از ابهام‌زدایی در حوزه‌های دانشی در پژوهش‌های منتشر شده است. برای انجام این مرور نظام‌مند، مطالعات مرتبط با حوزه ابهام‌زدایی در پایگاه‌های اطلاعاتی بازیابی شد. ابعاد گوناگونی که از مرور ادبیات پژوهش‌های بازیابی شده به دست آمد، مبنای ارائه صورت‌بندی جامعی از ابهام‌زدایی در این ترکیب قرار گرفت. افزون بر این، تحلیل کیفی برای کل مطالعات بازیابی شده با بهره‌گیری از بازینه ارزیابی انتقادی در حوزه علم اطلاعات و دانش‌شناسی انجام و اعتبار پژوهش‌ها برای حضور در این ترکیب بررسی شد.

این مرور نظام‌مند، یک چارچوب ترکیبی از مؤلفه‌های حوزه ابهام‌زدایی ارائه کرد. این مؤلفه‌ها شامل روش‌های انجام پژوهش (بر اساس منابع و نیز بر اساس یادگیری ماشینی) و ابزار گردآوری داده (جست‌وجوی اسناد، پرسشنامه، مصاحبه و ...)، منابع استفاده‌شده (منابع دانشی یا پیکره) برای ابهام‌زدایی و کاربردهای آن است. همچنین موضوعات پنهان (معنا و موضوع، نام، موجودیت‌های نامدار، احساس و نظر، صوت و تصویر و ...) در ادبیات پژوهش نیز استخراج و در چند گروه کلی طبقه‌بندی شد. بررسی موضوع ابهام‌زدایی نشان داد که ابهام‌زدایی در موضوعات و حوزه‌های مطالعاتی و با اهداف گوناگون انجام می‌شود و انتخاب روش

مناسب پژوهش و گردآوری داده و ابزارها، به نوع داده‌ها بستگی دارد. همچنین، ابهام‌زدایی در حوزه‌های گوناگون دانشی نیاز است تا ضمن کسب شواهد بیشتر برای آن، پژوهشگران حوزه‌های تخصصی را نیز یاری رساند.

کلیدواژه‌ها: ابهام، رده‌بندی، علم اطلاعات، مدیریت دانش، مرور سیستماتیک

۱. مقدمه

با گسترش فزاینده رسانه‌های دیجیتال و توسعه فناوری‌های رایانه‌ای، حجم انبوهی از متون و منابع اطلاعاتی برخط تولید می‌شود که برای بهره‌برداری از آن‌ها باید به پردازش و بررسی پرداخت و از میان آن‌ها اطلاعات درست را استخراج کرد (میردامادی، زارع بیدکی و رضائیان ۱۳۹۲).

استخراج و بازیابی اطلاعات و دانش از متن، یکی از اهداف متن‌کاوی در نظر گرفته شده و با افزایش توجه پژوهشگران به پردازش متون زبان طبیعی روش‌های بسیاری نیز برای آن به کار می‌رود (اسلامی و همکاران ۱۳۹۷؛ محدث دیلمی، صدر، و ترخان ۱۴۰۰).

پردازش زبان طبیعی را پردازش و تجزیه و تحلیل اطلاعات متنی با استفاده از رایانه تعریف کرده‌اند که پیشرفت‌های زیادی در حوزه محاسبات و هوش مصنوعی را سبب شده و از حوزه‌های مهم آن به‌شمار می‌رود (مهدی‌پور و رازقی یدک ۱۳۹۷؛ Li, and Yang 2018). زبان به‌عنوان یکی از رایج‌ترین و موثقت‌ترین راه‌های انتقال فکر و حس و بار این اطلاعات بر عهده واژگانی است که محتوای آن را تشکیل می‌دهد، اما به‌سبب پیچیدگی‌های زبان بشری انسانی هر واژه ممکن است ویژگی‌های (ساختار، ریخت، معنا و ...) بی‌شمار و متفاوت داشته باشد (محدث دیلمی، صدر، و ترخان ۱۴۰۰؛ Al-Saiagh et al. 2018). از این‌رو، زبان طبیعی دارای ابهام است که در سطوح مختلف از جمله ساختار درونی واژه، ریخت‌شناسی، معنایی و ... رخ می‌دهد. مهم‌ترین آن‌ها، مواجهه با واژه‌ای است که ابهام معنایی دارد (دشتی و بردسیری ۱۳۹۵؛ Menai 2014).

وجود ابهام مانع از بازیابی و حتی درک اطلاعات به‌ویژه اطلاعات تخصصی است و همچنین بر عملکرد برنامه‌های کاربردی از جمله ترجمه ماشینی، بازیابی اطلاعات، دسته‌بندی متون و خلاصه‌سازی خودکار تأثیر می‌گذارد و چالش‌هایی را به‌وجود می‌آورد (رضایی دینانی، کربلاآقا کامران و میرزائیان ۱۴۰۰؛ Martinez, de Lacalle & Agirre (2008)؛ Lu et al. 2019). هدف ابهام‌زدایی انتخاب خودکار بهترین معنا از میان معانی موجود است که برای درک و فهم متن ضروری است. ابهام‌زدایی معنایی یا رفع ابهام معنایی اصطلاحی است که برای

استخراج و انتخاب صحیح‌ترین و مناسب‌ترین معنا از واژگان چندمعنایی بر اساس متن / بافت به کار می‌رود و در حوزه‌های گوناگون از جمله بازیابی اطلاعات، متن‌کاوی، استخراج اطلاعات، وب معنایی، ترجمه ماشینی و ... مورد توجه پژوهشگران قرار گرفته است. طی فرایند ابهام‌زدایی، معنای جهان‌شمول برای یک واژه یا مفهوم یا متن خاص شناسایی می‌شود. ممکن است معنای به‌دست آمده از ابهام‌زدایی با آنچه که در فرهنگ‌ها و اصطلاحنامه‌ها آمده، تطابق نداشته باشد. بنابراین، توصیف و تبیین معنای صحیح آن‌ها ضروری است. (دالوند، جودکی و نعمتی پویا ۱۳۹۸؛ خیرمند پاریزی و نورمندی‌پور ۱۳۹۵؛ Al-Saiagh؛ Al-Mubaid and Gungu 2012؛ Park, Shin, and Lee 2022؛ Hogenboom, Brojba-Micu & Frasinca 2021؛ et al. 2018؛ Keshavarzian Nodehi and Moghadam Charkari 2022). از این‌رو، در این پژوهش تلاش شده تا ضمن ارائه صورت‌بندی جامعی از ابهام‌زدایی در حوزه‌های دانشی گوناگون، به‌طور ویژه ابهام‌زدایی از حوزه‌های پزشکی مورد بررسی قرار گیرد.

۲. روش پژوهش

هدف پژوهش حاضر ارائه صورت‌بندی جامعی از ابهام‌زدایی در حوزه‌های دانشی است. بر این اساس از مرور نظام‌مند به‌عنوان روش پژوهش مناسب با این هدف استفاده شده است. در این پژوهش، مرور نظام‌مند با استفاده از دستورالعمل چند-مرحله‌ای برگرفته از کتاب Jesson, Matheson & Lacey (2011) انجام شده که منابع مرتبط با موضوع ابهام‌زدایی را بررسی و تحلیل می‌کند.



نمودار ۱. مراحل انجام مرور نظام‌مند

«مرور نظام‌مند» را فرایند جست‌وجو، ارزیابی، ترکیب و تفسیر پژوهش‌های کمی و کیفی درباره یک موضوع خاص می‌دانند (Catalano 2013)، اما این اصطلاح یا مفهوم به دو معنای روش‌شناسی تجویزی (ابزار-روش) یا گزارش‌برونداد خودبسنده (گزارش) به کار می‌رود. در معنای نخست، ابزار یا روش نظام‌مند و شفاف برای گردآوری، ترکیب و ارزیابی یافته‌های پژوهشی درباره یک موضوع یا پرسش خاص است که به دنبال به حداقل رساندن سوگیری‌های ناشی از مطالعات منفرد و غیرنظام‌مند است. در معنای دوم، یک مقاله پژوهشی است که مطالعات مرتبط با موضوع را شناسایی می‌کند، کیفیت آن‌ها را ارزیابی می‌کند و نتایج آن را با استفاده از روش علمی به‌طور خلاصه بیان می‌دارد (Jesson, Matheson & Lacey 2011).

هر کدام از این دو معنا را که در نظر بگیریم، مرور نظام‌مند بیشتر راهبردهای جست‌وجوی ساختاریافته درباره یک موضوع خاص را دربرمی‌گیرد که هدف آن، آگاهی از مطالعات پژوهشی منتشرشده در آن حوزه است که به شکل سنتی، امکان‌پذیر نبوده و منابع مفید را در اختیار پژوهشگران، فعالان و سیاست‌گذاران علمی قرار می‌دهد. همچنین، به‌عنوان مدل جست‌وجوی پژوهشگر برای مقالات آن حوزه به کار می‌رود؛ جست‌وجویی که در آینده نیز قابل تکرار است (Catalano 2013).

مرور نظام‌مند بیشتر برای بررسی پژوهش‌های حوزه پزشکی و سلامت به کار رفته است، اما در حوزه‌های دیگر و از جمله علم اطلاعات و دانش‌شناسی نیز نتایج مفیدی به دست داده است (Jesson, Matheson & Lacey 2011).

مرور نظام‌مند را در ۳، ۵ و ۸ مرحله تعریف کرده‌اند. در این پژوهش از روش هشت-مرحله‌ای شامل دامنه و نقشه، طرح و دستورالعمل، مستندات، معیارهای شمول و عدم شمول، جست‌وجو و غربالگری، ارزیابی کیفی، استخراج داده‌ها و ترکیب استفاده شده است (Jesson, Matheson & Lacey 2011) تا مرتبط‌ترین پژوهش‌های انجام‌شده با موضوع ابهام‌زدایی شناسایی شود. در ادامه، این مراحل تشریح می‌شوند.

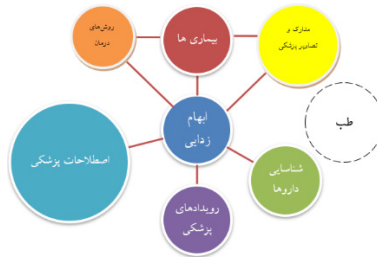
۱. دامنه و نقشه

در نقشه ساده‌ای که از دامنه مرور ادبیات پژوهش در موضوع مورد نظر ترسیم می‌شود، هر یک از حوزه‌های مربوط به این موضوع با دایره نشان داده می‌شود. روابط میان این دایره‌ها نشان می‌دهد که در کدام حوزه‌های موضوع کار پژوهشی انجام شده است.

چنانچه حلقه‌ای جدا افتاده و ارتباطی وجود نداشته باشد، یعنی در آن حوزه، پژوهشی وجود ندارد. بدین ترتیب، خلأهای موضوعی شناسایی می‌شوند. این نقشه همچنین می‌تواند حجم کارهای انجام‌شده در موضوع مورد نظر را نیز تبیین کند.

از آنجا که تأکید این پژوهش بر حوزه‌های پزشکی است، دامنه و نقشه پژوهش در این حوزه ترسیم شده است تا اطلاعات بیشتری درباره آن به دست آید.

از پژوهش‌های گردآمده برای ترکیب حاضر، تعداد ۱۵ پژوهش به ابهام‌زدایی در حوزه پزشکی و زیست‌پزشکی پرداخته و موضوعات گوناگون این حوزه را بررسی کرده‌اند. پژوهشگران در این حوزه با بررسی علائم مشابه به روشی برای تشخیص بیماری‌ها دست یافتند و برخی از روش‌های درمانی را تصریح کردند. همچنین آن‌ها با بررسی تصاویر پزشکی و مدارک درمانی برای دسته‌بندی موضوعی و تشخیص بیماری‌ها تلاش کردند و با شناسایی محتوا و مواد تشکیل‌دهنده داروها، به ابهام‌زدایی درباره داروها پرداختند. آن‌ها رویدادهای پزشکی را بر اساس مشخصات آن شناسایی کرده و بیشتر آن‌ها روش‌هایی را برای ابهام‌زدایی از واژگان و اصطلاحات این حوزه آزمودند. شکل زیر، نقشه ابهام‌زدایی در حوزه پزشکی و زیست‌پزشکی را به تصویر می‌کشد.



نمودار ۲. دامنه و نقشه پژوهش در حوزه پزشکی

۲. طرح و پروتکل

تدوین طرح یا دستورالعملی از پژوهش‌های مورد نظر، چشم‌انداز نظری و عملی و مفهومی یک مرور را محقق می‌سازد، هدف یا اهداف پژوهش را تصریح می‌کند و پرسش یا پرسش‌های پژوهش را بیان می‌دارد. این طرح درجه عینیت پژوهش را تعیین می‌کند؛ زیرا که مسئله و مراحل انجام پژوهش را مشخص می‌سازد.

هدف این ترکیب، بررسی و ارائه صورت‌بندی جامعی از ابهام‌زدایی در حوزه‌های دانشی با تأکید بر حوزه پزشکی است. برای دستیابی به این هدف، پرسش‌های زیر

مطرح شده و در این مرور به آن‌ها پاسخ داده می‌شود. این پرسش‌ها عبارت‌اند از:

- ◇ روش‌های به کاررفته برای ابهام‌زدایی کدام است؟
- ◇ کدام ابزارها برای گردآوری داده‌ها به کار می‌رود؟
- ◇ چه داده‌هایی جامعه و نمونه پژوهش را تشکیل داده است؟
- ◇ انواع منابع به کاررفته برای استخراج داده برای ابهام‌زدایی کدام است؟
- ◇ کدام حوزه‌های دانشی به ابهام‌زدایی پرداخته‌اند؟
- ◇ ابهام‌زدایی در چه موضوعاتی انجام شده است؟
- ◇ کاربردهای ابهام‌زدایی در حوزه‌ها و موضوعات گوناگون چیست؟

۳. مستندات

در یک مرور نظام‌مند باید تصمیم‌های پژوهشگر مستند شود تا فرایند پژوهش برای خواننده شفاف شود تا سایر پژوهشگران نیز بتوانند آن را بازطراحی کنند. عناوین پایگاه‌های داده، سال پوشش، واژگان جست‌وجو، دیگر محدودیت‌های جست‌وجو و تعداد باز یافت‌ها از جمله اطلاعاتی است که باید ثبت شود. این مستندات بسیار مهم هستند؛ زیرا به واسطه هر نوع جست‌وجو، مجموعه نتایج متفاوتی به دست می‌آید که در تحلیل و تفسیر هم تفاوت ایجاد می‌کند. افزون بر این، سایر پژوهشگران را درباره تکرار و چگونگی این پژوهش کمک می‌کند و خلاصه‌ای از فرایند پژوهش به شمار می‌رود.

برای انجام این ترکیب و گردآوری مقالات منتشر شده در موضوع ابهام‌زدایی، از پایگاه‌های اطلاعاتی برخط و معتبر بهره‌گیری شده است. پایگاه‌های «مگ‌ایران»، «نور مگز»، «آی‌اس‌سی»^۱، «مرکز منطقه‌ای شیراز»، «اس‌آی‌دی»^۲ و «سیولیکا» برای یافتن مقالات فارسی جست‌وجو شد و مقالات انگلیسی نیز از پایگاه‌های «ام‌الد»، «جی‌استور»، «پروکوئست» و «الزویر» به دست آمد. با توجه به تعداد محدود مقالاتی که به موضوع ابهام‌زدایی پرداخته بودند، بازه زمانی مشخصی برای گردآوری مقالات تعریف نشده و تمام مقالات این حوزه ذخیره شد.

کلیدواژه‌های جست‌وجو برای این ترکیب در پایگاه‌های اطلاعاتی فارسی: «ابهام‌زدایی»، «رفع ابهام»، «تعیین معنا»، «ویژگی معنایی»، «دسته‌بندی کلمات»، «دسته‌بندی متون»، «خوشه‌بندی کلمات»، «خوشه‌بندی متون» بود و در پایگاه‌های اطلاعاتی غیرفارسی،

1. ISC

2. SID

کلیدواژه‌های: "word sense disambiguation"، "semantic features"، "word categorization"، "context classification"، "classification"، "word categorization"، "context classification"، "classification"، "categorization" جست‌وجو شد. شایان ذکر است که تمام جست‌وجوها در مقالاتی انجام شد که فایل تمام‌متن آن‌ها در پایگاه‌ها موجود و قابل ذخیره بود. بدین ترتیب، تعداد ۶۷۷ مقاله ذخیره شد.

۴. معیارهای شمول و عدم شمول

در این نوع پژوهش، مقالاتی نیاز است که بتوانند پرسش‌های پژوهش را پاسخ دهند. از این رو، معیارهای شمول و عدم شمول بایستی شفاف بیان شوند و با خروجی مطابقت داشته باشند تا بتوانند دانش آن حوزه را به تصویر کشند. با کمک این معیارها آثاری با بیشترین ارتباط ظاهری با موضوع مورد نظر به دست می‌آید؛ زیرا کلیدواژه‌ها تنها در عنوان و چکیده جست‌وجو می‌شوند. پس از غربالگری، کل مقالات، مرور و اطلاعات لازم برای استخراج داده‌ها گردآوری می‌شود. در این مرحله، معیارهای کیفی به کار می‌رود تا مقالات در این ترکیب قرار گیرند یا حذف شوند. این معیارها سبب می‌شود که انتخاب مقالات برای قرار گرفتن در ترکیب از ذهنیت و قضاوت پژوهشگر به دور باشد.

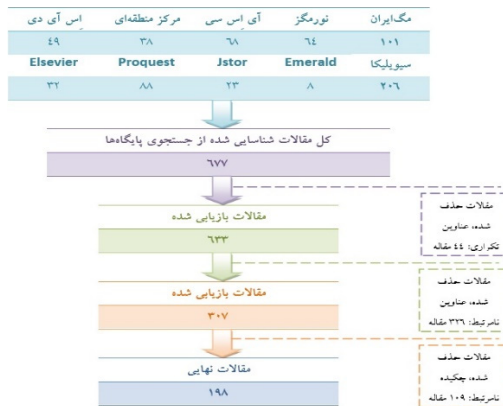
چنانکه پیشتر هم اشاره شد، جست‌وجوی منابع، هم در پایگاه‌های فارسی و هم در پایگاه‌های غیرفارسی انجام و بدین ترتیب، مقالاتی که کلیدواژه‌های مزبور را در عنوان یا چکیده خود داشتند، شناسایی و ذخیره شد. منابع لاتین تنها به زبان انگلیسی محدود و مقالات سایر زبان‌ها در جست‌وجو کنار گذاشته شد. از نظر نوع منابع، به مقاله‌های مجلات و همایش‌ها محدود و مقالاتی که از نظر محتوا، مروری و توصیفی نامیده می‌شود، حذف و تنها مقالات کیفی که با استفاده از روش‌های پژوهش انجام شده، در بررسی و ترکیب قرار گرفت. همچنین، همه مقالاتی که برای این ترکیب استفاده شده، در مجلات و منابع معتبر منتشر شده و به چاپ رسیده است.

۵. جست‌وجو و غربالگری

همه جست‌وجوها در فاصله زمانی یک ماهه (اول آبان تا دوم آذر ۱۴۰۱ معادل ۲۳ اکتبر تا ۲۳ نوامبر ۲۰۲۲) انجام و مقالات موجود در پایگاه‌های علمی معتبر شناسایی و ذخیره شد. برای این کار، ابتدا، عناوین تکراری حذف شد. سپس، باقی مقالات مرور و عناوین نامرتبط با موضوع پژوهش حاضر از مجموعه مقالات خارج شد. در مرحله بعد، چکیده

مقالات باقی‌مانده مطالعه شد تا مقالات نامرتب از نظر محتوا حذف شود. در انتها، ۱۹۸ مقاله واجد شرایط در این ترکیب قرار گرفت.

نام پایگاه‌ها و تعداد منابع یافت‌شده، تعداد عناوین تکراری، عناوین و چکیده‌های نامرتب در جدول زیر آمده است.



نمودار ۳. جست‌وجو و غربالگری منابع مرور نظام‌مند

۶. ارزیابی کیفی

منابع منتخب در مرور نظام‌مند برای ترکیب، ارزیابی می‌شود. مهم‌ترین جنبه ارزیابی، بررسی روش مطالعات است. ابزارها و استانداردهای گوناگونی برای ارزیابی آثار در اینترنت وجود دارد که بنابر ویژگی‌های هر حوزه متفاوت است و بایستی بر اساس استانداردهای کیفی قابل اجرا باشد. از آنجا که بیشتر پژوهش‌ها بر مبنای روش پیمایش است، ارزیابی کیفی باید بر اساس ویژگی‌ها و جزئیاتی باشد که نویسنده در طراحی پژوهش خود در مقاله ارائه کرده است.

از جمله ابزارهای ارزیابی کیفی، بازبینی‌هایی^۱ است که به تنوع قابل دسترس بوده، محدودیت‌های سایر ابزارهای ارزیابی را نیز ندارد و به بهبود نتیجه ارزیابی کمک می‌کند. اگرچه بازبینی‌های بی‌شماری برای ارزیابی انتقادی در مرور نظام‌مند و به رایگان در دسترس است، اما بازبینی انتقادی «ای‌بی‌ال گلین»^۲ به حوزه علم اطلاعات و دانش‌شناسی اختصاص دارد و عمومیت آن به حدی است که برای ارزیابی مطالعات کمی

1. checklist

2. Glynn's Evidence Based Librarianship (EBL) Critical Appraisal Checklist

و کیفی مناسب است (Catalano 2013). از این‌رو، کل مقالات در این تحلیل با استفاده از بازبینه انتقادی «ای‌بی‌ال‌گیلن» ارزیابی شده است.

بازبینه «بی‌ال‌گیلن» شامل ۴ بخش می‌شود:

- ◇ جامعه (نمونه، معیارهای شمول و عدم شمول و ...)
- ◇ گردآوری داده (روش گردآوری، اعتبار ابزار، سوگیری، شواهد و ...)
- ◇ طراحی پژوهش (اعتبار، اخلاق، خروجی و ...)
- ◇ نتایج (نتایج موجه، کاربردهای عملی و ...).

هر بخش شامل چندین پرسش است که با گزینه‌های «بله»، «خیر» و «نامشخص» پاسخ گفته می‌شود. نمره ارزیابی بخش‌ها بایستی بیش از ۷۵ باشد تا پژوهش معتبر به حساب آید، اما ممکن است نمره ارزیابی بخشی کمتر از ۷۵ باشد. از این‌رو، اعتبار کلی مورد محاسبه و مبنای ارزیابی قرار می‌گیرد. نمره ارزیابی کمتر از ۷۵ برای بخش‌های بازبینه ممکن است به سبب استفاده از آمار استنباطی باشد یا در آن پژوهش برخی معیارهای مهم نادیده گرفته شده یا ذکر نشده و یا حذف شده باشد.

ارزیابی پژوهش‌های این ترکیب با بازبینه «گیلن» نشان داد که تنها ۴۸ درصد (۹۵ اثر) از پژوهش‌ها معتبر هستند و این‌ها پژوهش‌هایی هستند که جزئیات انجام پژوهش را آشکارا بیان کرده‌اند. اما ۵۲ درصد (۱۰۳ اثر) پژوهش‌ها در تبیین ابزار گردآوری داده، روش نمونه‌گیری و سایر جنبه‌های طراحی پژوهش به صراحت عمل نکرده یا عملکرد ضعیفی داشته‌اند.

به‌منظور کسب اطمینان از اعتبار این بازبینه برای این ترکیب، برخی پژوهش‌ها به شکل تصادفی انتخاب و با «ابزار ارزیابی انتقادی کرو»^۱ نیز ارزیابی شده‌اند. «ابزار ارزیابی انتقادی کرو» برای ارزیابی انواع متفاوت پژوهش مفید است و برای اعتبارسنجی استفاده می‌شود (Catalano 2013).

نتایج این ارزیابی در جدول «طرح پژوهشی» (پیوست ۱) آمده است.

۷. استخراج داده‌ها

پس از انجام فرایندهای جست‌وجو و ارزیابی در مرور نظام‌مند، می‌توان به توصیف و تشریح مقالات پرداخت و اطلاعات آن‌ها را در جداول مربوط ثبت کرد.

1. Crow's Critical Appraisal Tool

اطلاعاتی که پس از بررسی پژوهش‌ها در جداول ثبت می‌شود، مبنای پاسخگویی به پرسش‌های مرور نظام‌مند قرار می‌گیرد.

داده‌های پژوهش بر اساس دو جدول «اطلاعات کتابشناختی» (پیوست ۲) و «طرح پژوهشی» به‌دست آمده است. جدول «طرح پژوهشی» اطلاعاتی همچون روش گردآوری داده، حجم نمونه، روش نمونه‌گیری، ابزار گردآوری داده، نتایج ارزیابی انتقادی و نظرات را شامل می‌شود. جدول «اطلاعات کتابشناختی» نیز اطلاعات هر مقاله از جمله نویسنده، کشور، سال نشر، حوزه یا رشته موضوعی هدف / کاربرد مطالعه را دربرمی‌گیرد.

۸. ترکیب

ترکیب و تحلیل، آخرین و مهم‌ترین مرحله مرور نظام‌مند است که به‌عنوان ارتباط میان بخش‌های گوناگون پژوهش تعریف شده و هدف از آن، تطبیق و ارائه داده‌های استخراج‌شده از مطالعات است که ویژگی‌ها و نتایج پژوهش‌ها را به روشی نظام‌مند خلاصه می‌کند. از کنار هم قرار گرفتن اطلاعات به‌دست آمده، مسئله پژوهش یا خلأ دانشی تبیین می‌شود.

بر اساس نوع مرور و موضوع پژوهش می‌توان ترکیب و تحلیل‌های گوناگونی ارائه کرد. تاکنون دو نوع ترکیب شناسایی شده است: روایتی یا جدولی. روایت برای پژوهش‌های توصیفی یا به تنهایی کیفی، و جدول برای پژوهش‌های کمی و دارای فرایندهای آماری به کار می‌رود.

برای ترکیب جنبه‌های گوناگون مرور، یادداشت‌هایی درباره تمام ابعاد مطالعات و پژوهش‌ها تهیه و اطلاعات مقالات با استفاده از رویکرد ترکیب تفسیری در جدول‌هایی طبقه‌بندی می‌شود. جدول‌ها کمک می‌کند تا ویژگی‌های مقالات در یک مجموعه گردآوری و برای تحلیل در اختیار پژوهشگر قرار گیرد. بدین ترتیب، امکان بهتری برای استخراج داده‌ها و پاسخگویی به پرسش‌های پژوهش وجود دارد.

گاهی ترکیب دشوار بوده و این از آن جهت است که پژوهش به سؤالات مشابه، اما در اساس متفاوت پاسخ می‌دهد. مزیت استفاده از رویکردهایی مانند مرور نظام‌مند یا فراترکیب این است که شواهد گوناگون باید تفسیر شوند تا بتوان به‌واسطه آن پدیده‌ای را تشریح کرد.

تجزیه و تحلیل یافته‌ها

برای ترکیب و تحلیل ۱۹۸ مطالعه گردآمده برای پژوهش حاضر از جداول «طرح پژوهشی» و «اطلاعات کتابشناختی» استفاده می‌شود. ترکیب تفسیری با پاسخ به پرسش‌های پژوهش صورت پذیرفته و به روشن شدن ابعاد موضوع ابهام‌زدایی کمک می‌کند. پرسش‌های ترکیب در ادامه پاسخ داده می‌شود:

♦ روش‌های به‌کاررفته برای ابهام‌زدایی کدام است؟

در تمام ۱۹۸ پژوهشی که در موضوع ابهام‌زدایی شناسایی و در نهایت برای ترکیب انتخاب شده، پیمایش به‌عنوان روش پژوهش انتخاب شده است. پژوهش‌های این ترکیب در پیمایش خود برای ابهام‌زدایی یا تعیین مناسب‌ترین معنا، موضوع، احساس، شناسایی موجودیت‌های نامدار و ... بر اساس منابع از روش‌های مبتنی بر قاعده یا منابع دانش، روش‌های مبتنی بر پیکره یا آمار و نیز روش‌های ترکیبی و بر اساس یادگیری ماشینی از روش‌های نظارتی، بی‌نظارت و نیمه‌نظارتی بهره برده‌اند.

روش‌های یادگیری عمیق و یادگیری انتقالی نیز از دیگر روش‌های یادگیری ماشینی است که دارای قابلیت‌های بیشتر بوده، و برای بررسی حجم زیادتری از داده‌ها کاربرد دارد و در برخی از پژوهش‌های این ترکیب به‌کار رفته، و همان‌سه روش نظارتی، بی‌نظارت و نیمه‌نظارتی برای آن‌ها هم اعمال می‌شود.

تنوع این روش‌ها امکان استفاده ترکیبی از آن‌ها را فراهم آورده و با بررسی و اعمال مدل‌های گوناگون سعی در پاسخگویی به پرسش‌های پژوهش داشته‌اند و بیشتر آن‌ها مدعی هستند که روش یا مدل پیشنهادی آن‌ها ابهام‌زدایی و تعیین معنا، موضوع، احساس و ... را بهبود بخشیده است و نسبت به دیگر روش‌ها و مدل‌ها برتری‌هایی دارد که دلیل آن را اجرایی بودن روش‌ها و الگوریتم‌ها برای حجم زیادی از داده‌ها می‌دانند. با این حال، در برخی پژوهش‌ها به صراحت اشاره نشده که از کدام‌یک از روش‌های نامبرده بهره‌گیری شده است.

از منظر روش‌شناختی

براساس یادگیری ماشینی

براساس منابع

بیمه نظارتی

نی نظارت

نظارتی

نظارتی
روش

تجزیه و تحلیل
مبتنی بر
روش

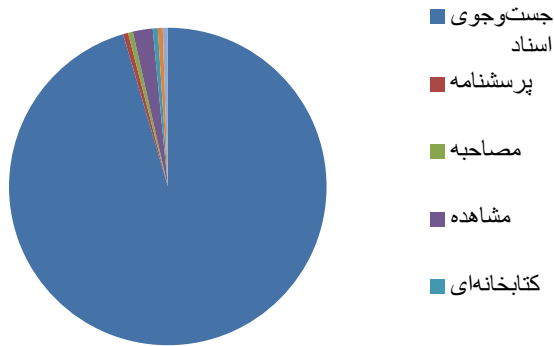
روش مبتنی بر
قاعده یا منابع
دانش

نمودار ۴. دسته‌بندی روش‌های ابهام‌زدایی

◆ کدام ابزارها برای گردآوری داده‌ها به کار می‌رود؟

با اینکه در تمام ۱۹۸ پژوهشی که در موضوع ابهام‌زدایی شناسایی شده از روش پیمایش بهره‌گیری شده، اما این پژوهش ابزارهای گوناگونی را برای گردآوری داده‌های خود به کار برده است. ۹۵ درصد از آن‌ها روش جست‌وجوی اسناد را برگزیده‌اند، یک پژوهش از پرسشنامه بهره‌جسته و یک اثر مصاحبه را انتخاب کرده، ۴ مطالعه دست به مشاهده زده و یک پژوهش داده‌های خود را به روش کتابخانه‌ای گرد آورده است. یک مطالعه تلفیقی از روش مشاهده و جست‌وجوی اسناد بهره‌گرفته و یک اثر درباره روش گردآوری داده خود توضیحی ارائه نکرده است (برای آگاهی از ابزارهای به‌کاررفته برای هر پژوهش، به پیوست ۱ نگاه کنید). استفاده حداکثری از جست‌وجوی اسناد به این دلیل است که بیشتر پژوهشگران به دنبال ارائه روش یا کاربست الگوریتمی برای ابهام‌زدایی هستند و این روش‌ها و الگوریتم‌ها بایستی در حجم قابل توجهی از اسناد اجرا و ارزیابی شوند. همچنین پژوهش‌ها از نرم‌افزارهای داده‌کاوی و تحلیل متن گوناگون برای پردازش داده‌ها بهره‌گرفته‌اند: «پروتژه»، «رپیدماینر»، ابزارها و کتابخانه‌های پایتون و ... از جمله آن‌ها هستند.

1. Protege
2. RapidMiner



نمودار ۵. دسته‌بندی ابزار گردآوری داده‌ها

◆ چه داده‌هایی جامعه و نمونه پژوهش را تشکیل داده است؟

جامعه بیشتر پژوهش‌ها متون و اسنادی هستند که بنا بر اهداف پژوهش گرد آمده و مورد بررسی قرار گرفته‌اند؛ هرچند برخی از آن‌ها به صراحت به آن اشاره نداشته‌اند.

از جمله منابع دانشی استفاده‌شده در پژوهش‌های این ترکیب می‌توان از سرعنوان‌ها، اصطلاحنامه‌ها، دایره‌المعارف‌ها و حتی پایگاه‌های اطلاعاتی و ... نام برد. پیکره‌های به کاررفته نیز منابعی از جمله مقالات، متون خبری و ... را دربردارند و در برخی موارد منابعی مانند تصاویر پزشکی، دست‌نوشته‌ها، تک‌نگاشت‌ها، فیلم‌ها، وسایل نقلیه، پروانه‌های ثبت اختراع را شامل می‌شوند که به فراخور اهداف پژوهش‌ها و از سوی پژوهشگران تهیه شده‌اند.

در برخی پژوهش‌ها تمام منابع دانشی یا اسناد پیکره بررسی شده و در برخی دیگر به نمونه‌ای یا زیرمجموعه‌ای از آن اکتفا شده تا با توجه به حجم زیاد داده‌ها، کار محاسبه و پردازش را تسهیل کنند. از بین ۱۹۸ پژوهش این ترکیب، آن تعداد که به بررسی نمونه پرداخته‌اند، تنها ۷ پژوهش نمونه تصادفی انتخاب کردند و باقی مطالعات اشاره‌ای به روش نمونه‌گیری و تعیین حجم نمونه نداشتند. از آنجا که حجم نمونه باید کافی باشد تا نتایج قابلیت تعمیم‌پذیری داشته باشند و به گروه خاصی نبایست تعلق داشته باشند، در برخی پژوهش‌ها نمی‌توان با اطمینان از تعمیم‌پذیری پژوهش سخن گفت.

در برخی پژوهش‌ها که نیاز به حجم زیادی از داده‌ها احساس می‌شود، چندین منبع دانشی یا پیکره به نام با یکدیگر تلفیق شدند تا پاسخگوی پرسش‌های پژوهش خود باشند. چند پژوهش هم ترکیبی از منابع دانشی و پیکره را برای پژوهش خود انتخاب کرده

و به بررسی آن‌ها پرداخته‌اند.

بیشتر منابع به کاررفته در پژوهش‌های این ترکیب، اسناد متنی بودند که در قالب منابع دانشی یا پیکره گرد آمده بودند. در مواردی که هدف پژوهش ابهام‌زدایی از صوت و تصویر بوده، از مجموعه اسناد صوتی، تصویری (عکس و فیلم)، استفاده کردند تا اهداف پژوهش تأمین شود.

◆ انواع منابع به کاررفته برای استخراج داده به منظور ابهام‌زدایی کدام است؟

پژوهش‌های انتخاب‌شده برای ترکیب حاضر از منابع / محمل‌های گوناگونی برای کسب شواهد یا ارزیابی روش‌های پیشنهادی خود استفاده کرده‌اند:

به جز تعدادی چند از پژوهش‌ها که از ابزار مشاهده، مصاحبه و پرسشنامه و ... برای پاسخگویی به مسئله پژوهش خود استفاده کرده‌اند، بقیه با جست‌وجوی اسناد، داده‌های خود را گردآوری کرده و برای این کار از اطلاعاتی در قالب منابع دانشی یا پیکره و تلفیقی از هر دو بهره برده‌اند. بیشتر این منابع / محمل‌ها استاندارد و شناخته شده هستند: SentiWord Net، Brown، MSH، NLM، Reuters، SEMEVAL، SENSEVAL، Word Net، TWA، COCA، Wikipedia، پیکره همشهری و برخی دیگر کمتر شناخته شده و تعدادی دیگر محقق ساخته هستند.

برخی از این منابع مانند Word Net در زبان‌های گوناگون تهیه شده‌اند؛ مانند Indo Word Net، Arabic Word Net، Hindi Word Net، Korea Word Net و Latvian Word Net. از پایگاه‌های اطلاعاتی که به عنوان منبع گردآوری داده استفاده شده، می‌توان به موارد زیر اشاره کرد: Pubmed، Medline و گنج.

در پژوهش‌هایی که ابهام‌زدایی از صوت و تصویر مبنای کار بود، از منابع غیرمتنی بهره گرفته شده؛ مانند: DCASE، MovieLens، RGBD، Kuang DB، FIDS BD، FG Net و Crorel Data Set. و برای داده‌های احساسی منابع ثبت نظر در پایگاه‌های تجاری به کار رفته است؛ مانند: آمازون، دیجی کالا، جملک، توئیتر و Senti Word Net.

◆ کدام حوزه‌های دانشی به ابهام‌زدایی پرداخته‌اند؟

هرچند ابهام‌زدایی هم‌زمان به حوزه‌های زبان‌شناسی و علوم رایانه‌ای مرتبط است، اما گاهی تسلط و نفوذ یکی از این حوزه‌ها بر دیگری غلبه دارد و این به هدف پژوهش وابسته است. ۳۷ درصد پژوهش‌های این ترکیب به حوزه زبان‌شناسی و با هدف پردازش

زبان طبیعی می‌پردازد و در ۳۱ درصد از پژوهش‌ها، فناوری اطلاعات موضوع غالب است. حتی در پژوهش‌هایی که به ابهام‌زدایی و تعیین معنا و احساس و ... در علوم گوناگون پرداخته می‌شود، زبان‌شناسی و علوم رایانه‌ای ثابت است.

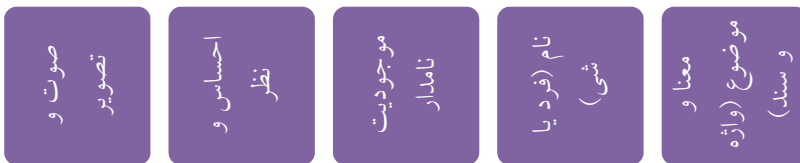
بررسی پژوهش‌های این ترکیب نشان داده که افزون بر این دو حوزه مطالعاتی، سایر حوزه‌ها هم به ابهام‌زدایی پرداخته و آن را در مباحث خود به کار برده‌اند. از جمله آن‌ها می‌توان به حوزه‌های پزشکی، زیست‌پزشکی، اقتصاد، علم اطلاعات و دانش‌شناسی و ... اشاره کرد.

◆ ابهام‌زدایی در چه موضوعاتی انجام شده است؟

با انجام مرور نظام‌مند مشخص می‌شود که در چه موضوعاتی ابهام وجود دارد و به تبع آن، لزوم ابهام‌زدایی از آن احساس می‌شود.

از میان پژوهش‌های این ترکیب، موضوعاتی که نیازمند ابهام‌زدایی هستند، شناسایی و سازماندهی شده‌اند. این حوزه‌ها شامل معنا و موضوع (واژه یا سند)، نام (فرد یا شیء)، موجودیت‌های نامدار، احساس و نظر، و صوت و تصویر است. ابهام‌زدایی هر یک از این موضوعات، با بهره‌گیری از داده‌های متنی، صوتی، تصویری و حتی شکل‌های منتسب به شبکه‌ها و رسانه‌های اجتماعی انجام شده است.

از منظر موضوعی



نمودار ۶. دسته‌بندی موضوعات ابهام‌زدایی

◆ کاربردهای ابهام‌زدایی در حوزه‌ها و موضوعات گوناگون چیست؟

تمام پژوهش‌های این ترکیب در موضوع ابهام‌زدایی انجام شده که اهداف یا کاربردهای بی‌شماری را دنبال می‌کنند. در ادامه، به این اهداف و کاربردها اشاره می‌شود:

در حوزه زبان‌شناسی:

- ◇ هدف بیشتر پژوهش‌ها بهره‌گیری از روش‌های محاسباتی و الگوریتم‌هاست تا به واسطه آن از برخی واژگان و اصطلاحات ابهام‌زدایی شود، معانی آن تعیین گردد و یا موضوع سند به دست آید؛
- ◇ برخی پژوهش‌ها به دنبال تعیین نقش تصریفی واژه و ابهام‌زدایی از آن در بافت زبانی است؛
- ◇ بررسی نقش نوایی و نحوی واژگان نیز از دیگر کاربردهای ابهام‌زدایی در این حوزه به‌شمار می‌رود.

در حوزه علوم رایانه‌ای:

- ◇ در بیشتر پژوهش‌ها ابهام‌زدایی برای آزمون روش‌های محاسباتی و الگوریتم‌های جدید به کار می‌رود؛
- ◇ در برخی پژوهش‌ها هدف ابهام‌زدایی، طراحی و ارزیابی نرم‌افزارها و بن‌سازه برای تقویت عملکرد یک حوزه یا بهبود در توسعه یک نظام اطلاعاتی و ارتباطی است؛
- ◇ هدف ابهام‌زدایی در تعدادی از پژوهش‌ها، تعیین مناسب‌ترین معنا و برابریابی بهتر برای ترجمه ماشینی و کاهش خطا در این نوع ترجمه است؛
- ◇ خلاصه‌سازی خودکار و انتخاب جمله‌هایی که به بهترین شکل محتوای سند/متن را منعکس می‌کند، از دیگر کاربردهای ابهام‌زدایی در پژوهش‌ها به‌شمار می‌رود؛
- ◇ در برخی پژوهش‌ها، هدف، تعیین مناسب‌ترین موضوع برای سند/متن در زبان‌های گوناگون و طبقه‌بندی منابع هم‌موضوع با کمک روش‌های خودکار و بدون نظارت است.

در حوزه علم اطلاعات و دانش‌شناسی - بازیابی اطلاعات:

- ◇ پژوهش‌های این حوزه به دنبال ابهام‌زدایی برای دستیابی به اطلاعات دقیق یا آزمون روش‌های بازیابی اطلاعات است؛
- ◇ در برخی پژوهش‌ها از ابهام‌زدایی در راستای بهبود جست‌وجو یا انتخاب بهترین منبع برای پرسش مورد جست‌وجو بهره‌گیری می‌شود تا بازخورد بهتری دریافت شود و کیفیت بازیابی اطلاعات افزایش یابد؛
- ◇ رفع ابهام از نویسندگان همنام یا نویسندگانی با چند املائی نمایه‌شده در پایگاه

اطلاعاتی از دیگر اهداف و کاربردهای پژوهش‌های این حوزه است؛
 ◇ کاربرد دیگر ابهام‌زدایی در این حوزه، سازماندهی اسناد وب به‌ویژه خبرگزاری‌ها
 برای دستیابی بهتر کاربران به داده‌های مورد نیاز است.

حوزه زیست‌پزشکی:

◇ عمده پژوهش‌های این حوزه با هدف ابهام‌زدایی از برخی مفاهیمی است که در این
 حوزه به معانی گوناگون به کار می‌رود و به عقیده پژوهشگران از جمله حوزه‌های
 موضوعی است که قابلیت بررسی انواع روش‌ها و الگوریتم‌های ابهام‌زدایی را به
 سبب ماهیت داده‌های آن داراست؛

حوزه پزشکی:

◇ در پژوهش‌های این حوزه ابهام‌زدایی برای شناسایی بیماری‌ها بر اساس ویژگی‌های
 مشترک افراد مبتلا یا بر اساس شباهت میان مدارک و تصاویر پزشکی آن‌هاست؛
 ◇ دیگر کاربرد ابهام‌زدایی در پژوهش، شناسایی روش‌های درمانی و داروها به‌ترتیب،
 بر اساس مشخصات و محتوای آن‌هاست؛
 ◇ شناسایی رویدادهای پزشکی نیز به‌عنوان یکی از کاربردها و اهداف پژوهش‌های
 ابهام‌زدایی در حوزه پزشکی مطرح است؛
 ◇ بیشترین سهم پژوهش‌های حوزه پزشکی در موضوع ابهام‌زدایی برای شناسایی و رفع
 ابهام از واژگان و اصطلاحات این حوزه است؛

اقتصاد و کسب‌وکار:

◇ در پژوهش‌های این حوزه، هدف برخی مطالعات، بررسی و تحلیل میزان خرید و فروش یک
 محصول است؛
 ◇ آگاهی از میزان رضایتمندی خریداران و استفاده‌کنندگان نیز هدف و کاربرد تعدادی از
 پژوهش‌های این حوزه است؛
 ◇ برخی پژوهش‌ها برای پیش‌بینی قیمت‌ها به ابهام‌زدایی پرداخته‌اند.

چندین اثر منفرد هم‌بازایی شده‌اند؛ مانند پژوهش‌های بازایی اطلاعات در حوزه هنر
 که شناسایی مشکلات و کمک به رفع آن‌ها و نیز بهبود کیفیت خدمات هدف آن‌ها به
 حساب می‌آید.



نمودار ۷. دسته‌بندی کاربردهای ابهام‌زدایی در حوزه‌ها و موضوعات گوناگون

در میان منابع فارسی، قدیمی‌ترین پژوهش در سال ۱۳۸۱ و جدیدترین پژوهش در سال ۱۴۰۱ (۲۰ سال فاصله) انجام شده است. برای پژوهش‌های غیرفارسی نیز قدیمی‌ترین پژوهش مربوط به سال ۲۰۰۰ و جدیدترین پژوهش سال ۲۰۲۲ (۲۲ سال فاصله) است. بازه زمانی پژوهش‌های انجام‌شده نشان می‌دهد که این حوزه به نسبت جدید بوده و به موازات پیشرفت در حوزه هوش مصنوعی و یادگیری ماشینی بروز کرده است. همچنین، در ایران همتای سایر کشورهای جهان، به مباحث این حوزه پرداخته شده و مطالعاتی را در زبان فارسی انجام داده‌اند.

بررسی و تحلیل پژوهش‌های این ترکیب، تنوع ملیتی پژوهشگرانی را که به بررسی ابهام‌زدایی در موضوعات گوناگون پرداخته‌اند، نشان می‌دهد. می‌توان چنین دریافت که ابهام و به تبع آن ابهام‌زدایی در بسیاری زبان‌ها وجود دارد و پژوهشگران هر کشور به فراخور نیاز خود از ابهام‌زدایی با اهداف گوناگون بهره می‌جویند تا به حل مشکلات بومی خود پردازند. کشورهای آسیایی و به‌ویژه جنوب شرق آسیا، بیشتر در این زمینه به پژوهش پرداخته‌اند. هرچند علت‌یابی این مسئله نیازمند پژوهشی جامعه‌شناسانه است، با این حال شاید بتوان علت آن را در نوع رسم‌الخط کشورهای این حوزه جست‌وجو کرد. الفبای فارسی، هندی، چینی، کره‌ای، عربی و ... نسبت به الفبای لاتین بسیار متفاوت بوده

و منابع دانشی یا پیکره خاص خود را نیاز دارد.

نکته دیگری که از بررسی پژوهش‌های ترکیب برمی‌آید، این است که بیشتر این پژوهش‌ها نیازمند حجم زیادی از داده‌هاست تا بتوان روش‌ها و مدل‌های پیشنهادی را در مورد آن‌ها اجرا و ارزیابی کرد.

نتیجه‌گیری

نتایج این ترکیب نشان داد که ابهام‌زدایی در موضوعات، حوزه‌های مطالعاتی و با اهداف گوناگون صورت می‌پذیرد. بیشتر پژوهش‌های این حوزه برای رفع ابهام از واژگان و اصطلاحات یا تعیین موضوع یک متن / سند انجام شده؛ اگرچه ابهام‌زدایی از سایر موضوعات مانند نام نویسنده، موجودیت‌های نامدار و ... نیز قابل توجه است. بیشترین پژوهش‌های ابهام‌زدایی سهم حوزه علوم رایانه است که با هدف طراحی و ارزیابی انواع مدل‌ها و روش‌ها، الگوریتم‌های گوناگون را برای رفع ابهام به کار می‌برند و به پژوهشگران سایر حوزه‌ها از جمله زبان‌شناسی، پزشکی، علم اطلاعات، اقتصاد و ... کمک می‌کنند.

این پژوهش نشان داد که انتخاب روش مناسب پژوهش و گردآوری داده و ابزارها، به نوع داده‌ها بستگی دارد. با این حال، بر اساس شواهد می‌توان گفت که انواع روش‌های مبتنی بر منابع دانشی یا پیکره و یا تلفیقی، نتایج قابل قبولی برای ابهام‌زدایی به دست می‌دهد. همچنین الگوریتم‌ها و مدل‌های شبکه عصبی برای ابهام‌زدایی و تعیین معنا و مفهوم نسبت به دیگر الگوریتم‌ها عملکرد رضایت‌بخش تری دارد. همچنین اگر داده‌ها ساختاریافته باشد، صرف زمان و هزینه کمتری را به همراه خواهد داشت. همچنین، برای داده‌های کلان روش‌های مبتنی بر یادگیری عمیق نتایج بهتری دارد و استفاده از آن‌ها بهبود کیفیت و افزایش دقت تحلیل را به ارمغان خواهد آورد.

تنوع منابع به کاررفته در پژوهش‌های این ترکیب، طیف گسترده و چشم‌انداز پژوهشگران را نشان می‌دهد، اما این نکته را نیز آشکار می‌سازد که با توجه به حجم زیاد داده‌ها برای چنین مطالعاتی، شاید بتوان از داده‌های پایگاه‌های اطلاعاتی در موضوعات گوناگون نیز بهره جست که بی‌شک بر نتایج پژوهش تأثیر به‌سزایی خواهد داشت. نتایج این ترکیب نشان داد که ابهام‌زدایی در حوزه‌های گوناگون دانشی نیاز است و لازم است انواع منابع مختلف در زبان‌های گوناگون و با کمک روش‌ها و الگوریتم‌های

متنوع بررسی و پردازش شوند تا ضمن ابهام‌زدایی از موضوعات گوناگون، صحت‌سنجی و اعتبارسنجی منابعی که برای ابهام‌زدایی استفاده می‌شود، عملکرد الگوریتم‌ها و روش‌ها نیز بررسی گردد. همچنین این ابهام‌زدایی در سایر حوزه‌های موضوعی نیز لازم است انجام شود تا فرصت گسترش و دستیابی به شواهد بیشتر فراهم گردد.

از مطالعه پژوهش‌های این ترکیب چنین برمی‌آید که اجزا و مکانیسم‌های موجود در پژوهش‌ها مانند نوع داده‌ها، تعیین حجم نمونه، انتخاب جامعه مناسب و ... از جمله چالش‌هایی است که این نوع پژوهش‌ها با آن روبه‌روست و فقدان مهارت‌های تخصصی به‌ویژه مهارت‌های رایانه‌ای و انتخاب مناسب‌ترین روش‌ها و الگوریتم‌ها برای انجام پژوهش‌های ابهام‌زدایی ممکن است بر انتخاب بهترین روش تأثیرگذار باشد. بنابراین، مهم است که پژوهشگران سایر حوزه‌های موضوعی نیز با مفاهیم، روش‌ها، الگوریتم‌ها و ... این حوزه آشنا باشند و از آن برای اهداف رشته و حوزه تخصصی بهره بگیرند. شاید به این واسطه موضوعات جدیدتری آشکار شود که نیاز به ابهام‌زدایی دارد و رفع ابهام از آن می‌تواند به آن رشته و پژوهشگران آن کمک کند.

یکی از حوزه‌هایی که به ابهام‌زدایی نیاز دارد، چنانکه در این پژوهش مشخص شد و تاکنون درباره رفع ابهام از آن و موضوعات مشمول، پژوهشی صورت نگرفته، طب سنتی است. به‌رغم وجود آثار بسیار در حوزه طب سنتی، آثار مکتوبی که به بررسی مفهوم و گستره صحیح حوزه طب سنتی پرداخته باشد، محدود است. این مسئله خلأهایی را در بحث‌های نظری این حوزه به‌وجود آورده و زمینه‌ساز اختلاف نظر و تشتت آرا در این حوزه شده است. از این‌رو، ابهام‌زدایی در این حوزه پیشنهاد می‌شود.

در خاتمه بایست عنوان کرد که بحث و نتیجه‌گیری برای این ترکیب بسیار سخت است؛ چرا که برای پرسش‌های مختلف، پاسخ‌های یکسان و مشابه وجود دارد. پس برای کسب اطلاعات بهتر پژوهشگران باید روش‌های مختلف را برای یک نوع داده خاص به کار ببرند یا یک روش را روی داده‌های گوناگون بررسی کنند. همچنین پژوهش‌هایی را درباره انواع منابع و حوزه‌های موضوعی گوناگون انجام دهند.

برای کسب نتایج بهتر درباره ابهام‌زدایی نیاز به پژوهشگرانی از حوزه علم رایانه است که روش‌ها و الگوریتم‌های جدید را آموزش دهند و در کنار متخصصان سایر حوزه‌های دانشی برای رفع ابهام در جنبه‌های مختلف آن رشته تلاش کنند تا ضمن کسب شواهد بهتر و بیشتر برای استفاده از ابداعات رایانه‌ای، به رفع مشکلات حوزه‌های تخصصی نیز

کمک کرده باشند. بدین ترتیب و با توجه به پیشرفت‌هایی که در حوزه علوم رایانه‌ای و هوش مصنوعی رخ داده، شاید بتوان پژوهش‌ها را به گونه‌ای طراحی کرد که نتایج حاصل از پردازش‌های ماشینی را هرچه بیشتر به تفکر انسانی نزدیک ساخت. پس لازم است این نوع مطالعه گسترش یابد.

فهرست منابع

- اسلامی، بهناز، مهدی حبیب‌زاده مطلق، مجید فولادیان، و زهرا رضایی. ۱۳۹۷. به کارگیری روش متن‌کاوی به منظور دسته‌بندی سلسله‌مراتبی متون با رویکرد یادگیری عمیق در داده‌های حجیم. مقاله ارائه‌شده در کنگره ملی توسعه پژوهش‌های نوین در مهندسی برق و کامپیوتر، تربت جام، ایران.
- خیرمند پاریزی، منیر، و رضا نورمندی‌پور. ۱۳۹۵. رفع ابهام معنایی کلمات فارسی با استفاده از رویکرد نظارت‌شده الگوریتم‌های IBL. علوم رایانشی ۱ (۲): ۶۳-۷۱.
- دالوند، زهرا، صبا جودکی، و سجاد نعمتی پویا. ۱۳۹۸. رفع ابهام کلمات در جمله با کلاس‌بندی معنایی با بهره‌گیری از SVM. مقاله ارائه‌شده در اولین کنفرانس بین‌المللی راهکارهای نوین در مهندسی، علوم اطلاعات و فناوری در قرن پیش رو، استانبول، ترکیه.
- دشتی، محمدصادق، و عمید خطیبی بردسیری. ۱۳۹۵. ارائه روشی جدید به منظور رفع ابهام معنایی کلمات با استفاده از معیارهای وزن‌دهی. مقاله ارائه‌شده در چهارمین همایش ملی کاربرد فناوری‌های نوین در علوم مهندسی. تربت حیدریه.
- رضایی دینانی، مینا، معصومه کربلاآقایی کامران، و وحیدرضا میرزاییان. ۱۴۰۱. واکاوی اثربخشی برچسب‌گذاری معنایی در رفع ابهام معنایی هم‌نویسه‌های تخصصی از نظر میزان دقت در بازیابی متون علمی. کتابداری و اطلاع‌رسانی ۲۴ (۴): ۳۱-۵۹.
- محدث دیلمی، فاطمه، حسین صدر، و مرتضی ترخان. ۱۴۰۰. ارائه یک مدل هوشمند به منظور تشخیص چندوجهی شخصیت کاربران با استفاده از روش‌های یادگیری ژرف. فناوری اطلاعات و ارتباطات ایران ۱۳ (۴۸-۴۷): ۷۲-۸۹.
- میردامادی، محمد مهدی، علی محمد زارع بیدکی، و مهدی رضائیان. ۱۳۹۲. قطعه‌بندی عبارات متون فارسی با استفاده از شبکه‌های عصبی. مهندسی برق و مهندسی کامپیوتر ایران ۱۱ (۲): ۷۶-۸۴.
- مهدی‌پور، الهام، و فرشته رازقی یدک. ۱۳۹۷. ارائه یک شبکه عصبی عمیق مبتنی بر کانولوشن برای طبقه‌بندی متون. مقاله ارائه‌شده در سیزدهمین سمپوزیوم بین‌المللی پیشرفت‌های علوم و تکنولوژی: سرزمین پایدار تازه‌های کامپیوتر و فناوری اطلاعات. ایران.

References

- Al-Mubaid, Hisham, and Sandeep Gungu. 2012. A Learning-Based Approach for Biomedical Word Sense Disambiguation. *The Scientific World Journal* 949247 :2012
- AL-Saiagh, Wafaa, Sabrina Tiun, Ahmed AL-Saffar, Suryanti Awang, & A. S. Al-Khaleefa. 2018. Word Sense Disambiguation Using Hybrid Swarm Intelligence Approach. *PLoS ONE* 13 (12): e0208695.
- Catalano, Amy. 2013. Pattern of Graduate Students' Information Seeking Behavior: A Meta-synthesis of the Literature. *Journal of Documentation* 69 (2): 243-247.
- Hogenboom, Alexander, Alex Brojba-Micu, & Flavius Frasinca. 2021. The Impact of Word Sense Disambiguation on Stock Price Prediction. *Expert Systems Applications* 184: 115568.
- Jesson, Jill K., Lydia Matheson, & Fiona M. Lacey. 2011. *Doing Your Literature Review: Traditional and Systematic Techniques*. Los Angeles, London, New Delhi, Singapore and Washington DC: Sage.
- Keshavarzian Nodehi, Azim, and Nasrollah Moghadam Charkari. 2022. A Metaheuristic with a Neural Surrogate Function for Word Sense Disambiguation. *Machine Learning with Application* 9: 100369.
- Li, Yang, and Tao Yang. 2018. Word Embedding for Understanding Natural Language: A Survey. In *Guide to Big Data Applications*, S. Srinivasan (Ed.), P. 83-104. [New York]: Springer.
- Lu, Wenpeng, Fanqing Meng, Shoujin Wang, Guoqiang Zhang, Xu Zhang, Antai Ouyang, & Xiaodong Zhang. 2019. Graph-Based Chinese Word Sense Disambiguation with Multi-Knowledge Integration. *Computers, Materials & Continua* 61 (1): 197-212.
- Martinez, David, Oier Lopez de Lacalle, & Eneko Agirre. 2008. On the Use of Automatically Acquired Examples for All-Nouns Word Sense Disambiguation. *Journal of Artificial Intelligence Research* 33: 79-107.
- Menai, Mohamed El Bachir. 2014. Word Sense Disambiguation using an Evolutionary Approach. *Informatica* 38: 155-169.
- Park, Jeong Yeon, Hyeong Jin Shin, and Jae Sung Lee. 2022. Word Sense Disambiguation Using Clustered Sense Labels. *Applied Sciences* 12: 1857.

پیوست ۱. نمونه‌ای از جدول «طرح پژوهشی»

نویسنده / کشور مطالعه	حجم نمونه	روش نمونه‌گیری	روش پژوهش	ابزار گردآوری داده	موضوع مطالعه	نمره ارزیابی انتقادی ای‌بی‌آل	نظرات
اسدی حقی و محمودی (ایران)	-	-	◇ پیمایش / منابع دانش ◇ یادگیری ماشینی	جست‌وجوی اسناد (آنتولوژی DBLP)	تعیین موضوع	< ۷۵	توضیح ناکافی درباره جامعه و نمونه؛ طراحی پژوهش خوب > ۷۵
اسلامی، حبیب‌زاده مطلق، فولادیان و رضایی (ایران)	۱۹۶۰۱۰	-	◇ پیمایش / پیکره ◇ یادگیری عمیق	جست‌وجوی اسناد (پیکره متون خبری CNN)	تعیین موضوع	< ۷۵	
اسماعیلی شایان، نظری و اسماعیلی شایان (ایران)	۱۰۱۰۰۴۶	-	◇ پیمایش / پیکره ◇ یادگیری عمیق - بی نظارت	جست‌وجوی اسناد (مقالات خبرگزاری فارس)	تعیین موضوع	> ۷۵	< ۷۵ الف
امامی و قائمی (ایران)	۳۵۰۰ متن	-	◇ پیمایش / پیکره ◇ یادگیری انتقالی	جست‌وجوی اسناد (پیکره محقق ساخته)	تعیین موضوع	< ۷۵	
بارزگانی و الماس گنج (ایران)	۸۰۰۰۰۰۰ کلمه	-	◇ پیمایش / پیکره ◇ یادگیری ماشینی - بی نظارت	جست‌وجوی اسناد (پیکره محقق ساخته)	تعیین معنا	< ۷۵	طراحی پژوهش خوب > ۷۵
باقرلو، طاوولی، کوزه‌گر و سام دلیری (ایران)	بیش از ۳۱۸ مقاله	-	◇ پیمایش / پیکره ◇ یادگیری ماشینی - نظارتی	جست‌وجوی اسناد (پیکره همشهری)	ابهام‌زدایی از موضوع	> ۷۵	

پیوست ۲. نمونه‌ای از جدول «اطلاعات کتابشناختی»

نویسنده/ کشور مطالعه	تاریخ نشر	حوزه مطالعاتی	هدف/ کاربرد	موضوع مطالعه
اسدی حقی و محمودی (ایران)	۱۳۹۳	علوم رایانه‌ای	تعیین مناسب‌ترین موضوع برای هر سند بر اساس شبکه‌های معنایی	تعیین موضوع
اسلامی، حبیب‌زاده مطلق، فولادیان و رضایی (ایران)	۱۳۹۷	علوم رایانه‌ای	تعیین مناسب‌ترین موضوع برای هر سند	تعیین موضوع
اسماعیلی شایان، نظری و اسماعیلی شایان (ایران)	۱۳۹۸	زبان‌شناسی	تعیین مناسب‌ترین موضوع برای هر سند	تعیین موضوع
امامی و قائمی (ایران)	۱۳۹۴	علوم رایانه‌ای	تعیین مناسب‌ترین موضوع برای هر سند بر اساس روش‌های بیزی	تعیین موضوع
بارزگانی و الماس گنج (ایران)	۱۳۸۶	علوم رایانه‌ای	شناسایی شباهت معنایی میان کلمه و قرار دادن آنها در یک گروه	تعیین معنا
باقرلو، طاوولی، کوزه‌گر و سام دلیری (ایران)	۱۳۹۳	زبان‌شناسی (NLP)	ارائه روشی برای افزایش دقت در شناسایی موضوع سند	ابهام‌زدایی از موضوع

حمیده اسدی

دارای مدرک تحصیلی دکتری در رشته علم اطلاعات و دانش‌شناسی با گرایش بازیابی اطلاعات از دانشگاه تهران است. حوزه‌های روش‌شناسی پژوهش، بازیابی اطلاعات و علم‌سنجی از جمله علایق پژوهشی وی است.



نادر نقشینه

دارای مدرک تحصیلی دکتری در رشته علم اطلاعات و دانش‌شناسی از دانشگاه تهران است. ایشان هم‌اکنون دانشیار گروه علم اطلاعات و دانش‌شناسی دانشگاه تهران است. حوزه‌های فناوری اطلاعات، داده‌کاوی، سبیرنتیک از جمله علایق پژوهشی وی است.



ملوک‌السادات حسینی بهشتی

دارای مدرک تحصیلی دکتری در رشته زبان‌شناسی با گرایش همگانی از دانشگاه تهران است. ایشان دانشیار پژوهشکده علوم اطلاعات، گروه اصطلاح‌شناسی و هستان‌شناسی پژوهشگاه علوم و فناوری اطلاعات (ایرانداک) است. مطالعه اصطلاح‌شناسی، مدیریت دانش، مدیریت اطلاعات، سازماندهی اطلاعات و پردازش زبان طبیعی از جمله علایق پژوهشی وی است.



پژوهش نامه
پردازش و
مدیریت
اطلاعات