

Zamani, Mohammad; Ghiasi, Mitra; Tahmasebi Limooni, Safiyeh (2024). Identifying Ambiguous Causal and Interfering Factors in Information Storage and Retrieval Systems. *Journal of Knowledge-Research Studies*, 3 (3): 35-51.

DOI: 10.22034/jkrs.2024.62883.1100

URL: https://jkrs.tabrizu.ac.ir/article_18574.html

©The Author(s)

Publisher: University of Tabriz

The paper is an open access and licensed under the Creative Commons CC BY NC license.



Identifying Ambiguous Causal and Interfering Factors in Information Storage and Retrieval Systems

Mohammad Zamani¹, Mitra Ghiasi^{2*}, Safiyeh Tahmasebi Limooni³

Received: August, 17, 2024; Revised: September, 29, 2024

Accepted: October, 1, 2024; Published: November, 21, 2024

Abstract

Purpose: Ambiguity significantly affects the efficiency and accuracy of information storage and retrieval systems. This research aims to identify the factors contributing to the creation and resolution of ambiguity.

Methodology: This qualitative study employs a grounded theory approach, involving 18 subject matter experts selected using a snowball sampling method. Data collection was conducted through semi-structured interviews, and analysis was performed using MAXQDA 20 software.

Findings: "Inherent ambiguity" emerged as the most significant causal factor (25%), while "intentional ambiguity" and "sentence structural ambiguity" were the least significant (8.4%). Among intervening factors, "intentional and unintentional ambiguity" accounted for the highest impact (41%), with "inauthentic sources" being the least (7%).

Conclusion: Ambiguous factors in information systems include written ambiguity, semantic ambiguity, structural ambiguity, and inherent ambiguity. This research addresses the gap in qualitative studies on ambiguity in such systems, offering insights to enhance data retrieval methods.

Value: Unlike prior quantitative studies focusing on technical aspects, this study provides a qualitative exploration of causal and intervening factors of ambiguity, contributing to the understanding of challenges in information retrieval.

Keywords: *Ambiguity, Causal Factors, Interfering Factors, Information Retrieval Systems*

1. PhD Student, Department of Knowledge and Information Science, Babol Branch, Islamic Azad University, Babol, Iran.

2. Associate Professor, Department of Knowledge Information Science and, Babol Branch, Islamic Azad University, Babol, Iran (Corresponding Author) mighiasi@gmail.com

3. Associate Professor, Department of Knowledge Information Science, Babol Branch, Islamic Azad University, Babol, Iran.

Extended Abstract

Introduction: Ambiguity is an inherent characteristic of natural language that poses significant challenges to information storage and retrieval systems. These systems often struggle with ambiguity, which can render retrieved information irrelevant or incomprehensible. In the domain of information science, ambiguity arises from terminological, syntactic, and semantic variations, which necessitate in-depth exploration and resolution strategies. Natural language, being complex and context-dependent, is prone to misinterpretation by machines due to its reliance on human cognitive abilities such as contextual understanding, semantic association, and syntactic organization (Tadin, 2011). For instance, the word "shoulder" illustrates semantic ambiguity, as it can mean a human body part, a roadside area for vehicles, or a tool used in hair arrangements. Humans intuitively disambiguate such words based on context, but machines lack this innate ability and require advanced algorithms to interpret meaning correctly. The inability to resolve ambiguity often results in search engines retrieving irrelevant pages, frustrating users and reducing system efficiency. Given these challenges, it becomes essential to identify the causal and intervening factors contributing to ambiguity in information retrieval systems. By addressing these factors, researchers and practitioners can develop strategies to enhance the functionality and accuracy of such systems. This study seeks to fill a gap in the existing literature by qualitatively exploring these factors, providing new insights that complement previous technical and quantitative studies.

Purpose: The primary objective of this study is to explore and identify the ambiguous causal and intervening factors in information storage and retrieval systems. The study aims to address the following key questions:

What are the main factors causing ambiguity in information retrieval systems?

What are the intervening factors that exacerbate the effects of ambiguity in these systems? The findings of this research are expected to benefit various fields, including linguistics, artificial intelligence, computer science, and information technology. By identifying and addressing ambiguity-related challenges, this research aims to enhance the user experience and efficiency of information retrieval systems.

Methodology: This study adopts a qualitative research approach, employing grounded theory as the primary methodological framework. Grounded theory was chosen for its ability to generate insights from qualitative data systematically, enabling the identification of causal and intervening factors that contribute to ambiguity. The research participants included 18 experts in fields relevant to information retrieval, selected through snowball sampling to ensure access to informed opinions. Data were collected using semi-structured interviews, which allowed participants to provide detailed and nuanced responses. These interviews were analyzed using MAXQDA 20 software, facilitating the coding and categorization of themes into causal and intervening factors.

The coding process consisted of three stages:

Open Coding: Concepts were identified and categorized based on interview data.

Axial Coding: Relationships between the categories were established to reveal patterns and hierarchies.



Journal of
Knowledge-Research Studies
(JKRS)

Vol 3

Issue 3

Serial Number 9

Selective Coding: Core categories were identified, and their significance was assessed based on the frequency and emphasis in the data.

Findings: The analysis revealed two primary categories of factors affecting ambiguity in information retrieval systems: causal factors and intervening factors.

1. Causal Factors of Ambiguity

Inherent Ambiguity (25%): This was the most prominent category, encompassing ambiguities that naturally arise due to the multifaceted nature of language.

Written Ambiguity (16.5%): Ambiguities that stem from inconsistencies or variations in written text.

Semantic Ambiguity of Words (12.5%): Instances where words have multiple meanings depending on context.

Spoken Ambiguity (12.5%): Ambiguities arising from verbal communication, often due to pronunciation or intonation variations.

Structural Ambiguity in Data (16.7%): Issues arising from the organization and presentation of data.

Intentional Ambiguity (4.8%): Ambiguities deliberately introduced, often for creative or strategic purposes.

Sentence Structural Ambiguity (4.8%): Ambiguities resulting from sentence construction or grammar.

2. Intervening Factors in Ambiguity

Intentional Ambiguity (41%): Deliberate ambiguities that impact the retrieval process.

Unintentional Ambiguity (41%): Ambiguities arising from errors or unintentional oversights in data or system design.

Information Retrieval Algorithms (12%): Limitations in algorithmic capabilities to process and resolve ambiguous inputs.

Inauthentic Sources (6%): Unreliable or invalid sources of information that introduce inconsistencies.

The data indicate that "inherent ambiguity" plays the most critical role in causing ambiguity, while "intentional and unintentional ambiguity" dominate as intervening factors.

The findings underscore the complexity of ambiguity in information retrieval systems. "Inherent ambiguity," being a fundamental characteristic of natural language, requires comprehensive strategies for effective resolution. Similarly, intervening factors such as intentional and unintentional ambiguities highlight the need for better design and validation of algorithms and data sources.

The research also emphasizes the role of contextual understanding in disambiguation. Human cognition excels in interpreting ambiguous inputs by leveraging context, prior knowledge, and situational cues. Translating these cognitive processes into machine-readable algorithms remains a significant challenge.



Journal of
Knowledge-Research Studies
(JKRS)

Vol 3

Issue 3

Serial Number 9

Table 1. Data extracted from ambiguous causal factors in information retrieval systems

Percent	Abundance	Category	Class
16/5	4	Written ambiguity	Causal factors of ambiguity
12/5	3	The semantic ambiguity of words	
4/8	2	Intentional ambiguity	
25	6	Inherent ambiguity	
12/5	3	Speech ambiguity	
4/8	2	Sentence structural ambiguity	
16/7	4	Structural ambiguity in the data	
%100	24		Total

Table 2. Ambiguous intervening factors in information retrieval systems

Percent	Abundance	Category	Class
41	7	Intentional ambiguity	Intervening factors in ambiguity
41	7	Unintentional ambiguity	
12	2	Information retrieval algorithms	
6	1	Inconvenient sources	
%100	17		Total

Conclusion: This study provides a qualitative exploration of the ambiguous causal and intervening factors in information retrieval systems. The results identify key categories of ambiguity, offering a framework for addressing these challenges through improved system design and algorithmic development.

The findings have practical implications for professionals in linguistics, artificial intelligence, and information technology. By addressing ambiguity at both the causal and intervening levels, these systems can be optimized to deliver more accurate and relevant results.

Value: Unlike previous studies, which predominantly focus on quantitative analyses and technical aspects of ambiguity, this research employs a grounded theory approach to explore the underlying causes and effects qualitatively. This novel perspective contributes to a deeper understanding of ambiguity in information retrieval systems, addressing a critical gap in the literature. The insights gained from this study can inform the development of more robust algorithms and methodologies for disambiguation, ultimately enhancing the efficiency and user experience of information retrieval systems.

Keywords: ambiguity, causal factors, interfering factors, information retrieval systems

References

- Abbasi, M. (2017). *Disambiguation of Persian words using deep learning algorithm*. [Master's thesis, Mohaghegh Ardabili University]. Irandoc [In Persian]
- Abdollahzadeh, A., Moghimi, S. M., & Mazari, E. (2024). Conceptualizing Self Sacrificial Leadership in Public Organizations: A Grounded Theory Approach. *Journal of*



Journal of
Knowledge-Research
Studies
(JKRS)

Vol 3

Issue 3

Serial Number 9

- Endowment & Charity Studies*, 2(1), 203-226. doi: 10.22108/ecs.2024.140244.1084 [In Persian]
- Abeyesiriwardana, M., & Sumanathilaka, D. (2024). A Survey on Lexical Ambiguity Detection and Word Sense Disambiguation. *arXiv preprint arXiv:2403.16129*. <https://doi.org/10.48550/arXiv.2403.16129>
- AlMousa, M., Benlamri, R., & Khoury, R. (2022). A novel word sense disambiguation approach using WordNet knowledge graph. *Computer Speech & Language*, 74, 101337. <https://doi.org/10.1016/j.csl.2021.101337>
- Bar-Hillel, Y. (1960). The present status of automatic translation of languages. *Advances in computers*, 1, 91-163. <https://doi.org/10.1016/j.csl.2021.101337>
- Berendsen, R., Kovachev, B., Nastou, E. P., De Rijke, M., & Weerkamp, W. (2012). Result disambiguation in web people search. In *Advances in Information Retrieval: 34th European Conference on IR Research, ECIR 2012, Barcelona, Spain, April 1-5, 2012. Proceedings 34* (pp. 146-157). Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-28997-2_13.
- Butenko, I. (2023, October). Disambiguation in information retrieval in scientific and technical texts. In *AIP Conference Proceedings* (Vol. 2833, No. 1). AIP Publishing. <https://doi.org/10.1063/5.0151708>.
- Charmaz, Kathy. (2006). Constructing a basic theory, a practical guide for qualitative analysis M. (Kazemimohved, H. Irani, 2017, trans.). Rasa. [In Persian]
- Chen, P., Bowes, C., Ding, W., & Choly, M. (2012). Word sense disambiguation with automatically acquired knowledge. *IEEE Intelligent Systems*, 27(04), 46-55. <https://doi.org/10.1109/MIS.2010.134>.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods Approaches* (3rd ed.). Sage Publications, Inc.
- Empson, W. (1930). *Seven Types of Ambiguity*. Chatto.
- Fakhrahmad, S. M., Rezapour, A. R., ZolghadriJahromi, M., & Sadreddini, M. (2011). A new word sense disambiguation system based on deduction. In *Proceedings of the World Congress on Engineering* (Vol. 2, pp. 1276-1281). Faust, M. (Ed.). (2011). *The handbook of the neuropsychology of language*. <https://onlinelibrary.wiley.com/doi/book/10.1002/9781118432501>.
- Gholami, F., & Ghaderzadeh, O. (2015). An Analysis of Women,s Experiences of the Augury (A Qualitative Study of Reasons and Consequences of Augury). *Journal of Applied Sociology*, 26(2), 107-124. doi: 10.22108/ecs.2024.140244.1084 [In Persian].
- Glaser, B. G., & Holton, J. (2005). Staying open: The use of theoretical codes in grounded theory. *The Grounded Theory Review*, 5(1), 1-20.
- Hristea, F., & Colhon, M. (2020). The long road from performing word sense disambiguation to successfully using it in information retrieval: An overview of the unsupervised approach. *Computational Intelligence*, 36(3), 1026-1062. <https://doi.org/10.1111/coin.12303>
- Imani, M. (2012). *Disambiguation and evaluation of information extracted from natural language text*. (Unpublished doctoral dissertation). Iran University of Science and Technology. [In Persian]
- Khodabakhshian, V. (2016). Improving the resolution of the semantic ambiguity of words in Wordnet by using the graph of the connection of the sense of words. [Master's thesis, Shahrood Islamic Azad University]. Irandoc. [In Persian]
- Lyons, J. (1970). *Semantics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511620614>
- Mansourian, Y. (2013). *Research method in information science and knowledge studies*. Samt. [In Persian]
- Mozafari, N. (2021). A Genetic-based Approach for Author Name Disambiguation Problem. *Iranian Journal of Information Processing and Management*, 36(3), 791-816. doi: 10.52547/jipm.36.3.791 [In Persian]



Journal of
Knowledge-Research Studies
(JKRS)

Vol 3

Issue 3

Serial Number 9

- Mozaffari, Z., Taki, G., Sabbagh Jaffari, M., & Yusefian, P. (2018). Preposition sense disambiguation in Persian using semantic frames. *Language Research*, 9(1), 99-117. doi: 10.22059/jolr.2018.66652 [In Persian]
- Nasri, M., Nosratzadeh, H., & Modarres Khiabani, S. (2023). Collocations and Lexical Ambiguities and Their Related Problems (Failures) in Farsi Equivalent Selection of Banking Texts by Application of Structural Semantics View. *Journal of Foreign Language Research*, 13(2), 177-192. DOI: 10.22059/jflr.2022.349034.989 [In Persian]
- Rezaei Dinani, M., Karbala Aghaei Kamran, M., & Mirzaeian, V. (2022). Effectiveness of Semantic Tagging in Sense Disambiguation of Specialized Homographs from the Perspective of False Drop in Retrieving Scientific Texts. *Librarianship and Information Organization Studies*, 33(1), 107-124. doi:10.30484/nastinfo.2021.2914.2063 [In Persian]
- Sadeghi, V., & Mohammadi, N. (2022). Lexical retrieval in read speech: The effect of speech prosody. *ZABANPAZHUI (Journal of Language Research)*, 14(44), 59-85. doi:10.22051/jlr.2022.38936.2127 [In Persian]
- Shiri, G. (2011). The importance and types of ambiguity in the researches. *Literary Arts*, 3(2), 15-36. dor:20.1001.1.20088027.1390.3.2.4.5 [In Persian]
- Sotoudeh, H. & Honarjouyan, Z. (2011). "Relationship in Information Science", National Library Studies and Information Organization, 64, 130-132. [In persian].
- Tadayyon, M. A. (2011). *Studying the current problems of solving ambiguity in machine translation based on a collection of texts and providing new solutions*. [Master's thesis, Shiraz University]. Irandoc [In Persian]
- Zamani, M. (1403). "Discovery of disambiguation and disambiguation studies in information retrieval systems, with an emphasis on Persian language". PhD thesis, Azad University - Babol branch. [In persian].
- Zamani, M., Ghiasi, M., & Tahmasabi Limoni, S. (2024). Ambiguity and Disambiguation in Information Retrieval Systems (Systematic Review). *Library and Information Sciences*, 26(4), 57-81. doi: 10.30481/lis.2024.429511.2125 [In Persian]
- Zhang, D., Zhang, H., Tang, J., Hua, X. S., & Sun, Q. (2020). Causal intervention for weakly-supervised semantic segmentation. *Advances in Neural Information Processing Systems*, 33, 655-666. <https://doi.org/10.48550/arXiv.2009.12547>
- Zuheros, C., Tabik, S., Valdivia, A., Martínez-Cámara, E., & Herrera, F. (2019). Deep recurrent neural network for geographical entities disambiguation on social media data. *Knowledge-Based Systems*, 173, 117-127. <https://doi.org/10.1016/j.knosys.2019.02.030>



Journal of
Knowledge-Research Studies
(JKRS)

Vol 3

Issue 3

Serial Number 9



زمانی، محمد؛ قیاسی، میترا؛ طهماسبی لیمونی، صفیه (۱۴۰۳). شناسایی عوامل علی و مداخله گر ابهام‌زاد در سیستم‌های ذخیره و بازیابی اطلاعات. نشریه مطالعات دانش پژوهی، ۳ (۳): ۳۵-۵۱.

DOI: 10.22034/jkrs.2024.62883.1100

URL: https://jkrs.tabrizu.ac.ir/article_18574.html



ناشر: دانشگاه تبریز

© نویسندگان

این مقاله به صورت دسترسی باز و با لایسنس CC BY NC کپی‌رایت‌کامانز قابل استفاده است.

شناسایی عوامل علی و مداخله گر ابهام‌زا در سیستم‌های ذخیره و بازیابی اطلاعات

محمد زمانی^۱، میترا قیاسی^۲، صفیه طهماسبی لیمونی^۳

۱. دانشجوی دکتری، گروه علم اطلاعات و دانش‌شناسی، واحد بابل، دانشگاه آزاد اسلامی، بابل، ایران.

۲. دانشیار، گروه علم اطلاعات و دانش‌شناسی، واحد بابل، دانشگاه آزاد اسلامی، بابل، ایران (نویسنده مسئول) mighiasi93@gmail.com

۳. دانشیار، گروه علم اطلاعات و دانش‌شناسی، واحد بابل، دانشگاه آزاد اسلامی، بابل، ایران.

تاریخ بازنگری: ۸ مهر ۱۴۰۳

تاریخ دریافت: ۲۷ مرداد ۱۴۰۳

تاریخ انتشار: ۱ آذر ۱۴۰۳

تاریخ پذیرش: ۱۰ مهر ۱۴۰۳

چکیده

هدف: ابهام یکی از چالش‌هایی است که ایجاد و رفع آن تأثیر زیادی بر کارایی و دقت سیستم‌های ذخیره و بازیابی اطلاعات دارد. این مطالعه در پی شناسایی عوامل علی و مداخله گر ابهام‌زا در این سیستم‌هاست.

روش‌شناسی: پژوهش حاضر از نظر هدف، کاربردی و از نظر رویکرد، کیفی و با روش داده‌بنیاد انجام گرفت. جامعه آماری ۱۸ متخصص موضوعی بود که با روش نمونه‌گیری گلوله برفی انتخاب شدند. ابزار گردآوری داده‌ها، مصاحبه نیمه ساختاریافته بود. برای تحلیل و کدگذاری داده‌ها، از نرم‌افزار مکس کیودا ۲۰ استفاده شد.

یافته‌ها: بیشترین مقوله اصلی در ابهام‌زایی، «ابهام‌زایی ذاتی» (۶ خرده مقوله، ۲۵٪) و کمترین «ابهام عمدی و ابهام ساختاری جمله» (۲ خرده مقوله، ۸/۴٪) هستند. همچنین بیشترین مقوله مداخله گر در ابهام‌زایی، «ابهام عمدی و غیر عمدی» (۷ خرده مقوله، ۴۱٪) و کمترین مقوله «منابع نامعتبر» (۱ خرده مقوله، ۷٪) بود.

نتایج: نتایج نشان می‌دهد که از دیدگاه متخصصان موضوعی عوامل علی ابهام‌زا در سیستم‌های ذخیره و بازیابی اطلاعات شامل ابهام نوشتاری، ابهام عمدی، ابهام گفتاری، ابهام‌زایی ذاتی، ابهام معنایی کلمات، ابهام ساختاری جمله، ابهام ساختاری در داده‌ها و عوامل مداخله گر ابهام‌زا شامل ابهام عمدی، ابهام غیر عمدی، الگوریتم‌های بازیابی اطلاعات، منابع نامعتبر هستند.

اصالت و ارزش: تحقیقاتی که تاکنون در زمینه عوامل ابهام‌زایی توسط پژوهشگران ایرانی و غیر ایرانی انجام گرفته است، اغلب از نوع کمی بوده و بیشتر جنبه فنی این موضوع را بررسی کرده‌اند. رویکرد داده بنیاد در پژوهش حاضر امکان جمع‌بندی نظرات و کشف عوامل علی و مداخله گر را فراهم آورده است که قبلاً بررسی نشده است.

کلیدواژه‌ها: ابهام‌زایی، عوامل علی ابهام‌زا، عوامل مداخله گر ابهام‌زا، سیستم‌های ذخیره و بازیابی اطلاعات

با پیشرفت فناوری اطلاعات، بحثی که از فهم ماشینی مطرح می‌شود، با بحث مطرح شده در حوزه علوم شناختی^۱ متفاوت است و منظور از فهم توسط ماشین در این حوزه، همان شکل ساخت یافته اطلاعات، یعنی روابط میان موجودیت‌هاست. به این ترتیب چیزی که به عنوان فهم از ماشین توقع می‌رود، پایین تر از فهم انسانی است. برای حل این مشکل، سعی می‌شود از نقاط قوت ماشین برای پوشاندن نقاط ضعف آن استفاده نمود. در مسئله مطالعه ماشینی، اگرچه انسان می‌تواند با دقت بسیار بالایی روابط میان موجودیت‌ها را استخراج نماید، اما از پردازش انبوه اطلاعات عاجز خواهد بود. در واقع آنچه که در استخراج روابط برای انسان به صورت دقت بسیار بالا و سرعت پایین وجود دارد، در ماشین می‌تواند به طور عکس آن واقعیت پیدا کند (ایمانی، ۱۳۹۲).

ابهام از جمله ویژگی‌های زبانی است که چالش‌های فراوانی را در سیستم‌های ذخیره و بازیابی اطلاعات ایجاد کرده و باعث نامرتب شدن و یا نامفهوم شدن آن می‌شود. در حوزه ابهام‌زدایی، به مطالعه در گستره واژه‌شناسی، نحو و معناشناسی نیازمند است. زبان طبیعی از بعد معنایی، مملو از ابهام است (تدین، ۱۳۹۱). درک و بازشناسی واژگان یکی از مهم‌ترین مهارت‌های شناختی زبان است. از آنجائی که واژگان انواع مختلفی از اطلاعات زبانی را شامل می‌شوند، درک و بازشناسی واژه، مستلزم اعمال فرایندهای متنوع مانند تجزیه و تحلیل ویژگی‌های فیزیکی زنجیره آوایی گفتار، بازیابی معنا، مقوله صرفی - نحوی واژگان و معنای گفتمانی و موقعیتی متناسب با بافت متن است (فاوست^۲، ۲۰۱۲). به این ترتیب، شنونده با دسترسی به الگوهای واج‌شناختی، صرفی، نحوی، معنایی، کاربردشناختی مرز واژگان را در گفتار پیوسته با سطح اطمینان بالایی شناسایی می‌کند (صادقی و محمدی نافچی، ۱۴۰۱).

مثال بارزی که می‌توان در مورد ابهام بیان کرد، کلمه «شانه» است که چندین معنی متفاوت در آن نهفته است (شانه به عنوان استخوان کتف انسان، شانه به عنوان توقفگاه خودرو در کنار جاده و شانه به عنوان ابزار مرتب کردن مو). ذهن انسان یک ماشین پیچیده است که می‌تواند کلمات و عبارات مبهم را به وسیله فهم موضوع انتخاب کند و معنای جمله را نیز متوجه شود. همچنین ذهن انسان به صورت خودکار نه تنها یک واژه بلکه گروهی از کلمات را در نظر می‌گیرد تا به معنا برسد؛ در حالی که در مورد ماشین برای برطرف کردن ابهام معنای کلمه، نیاز به حجم بالایی از درونداد و برون‌داد دارد (فخر احمد، رضاپور، ذوالفقاری جهرمی و صدرالدینی، ۲۰۱۱).

در رابطه با عوامل اگر از منظر مداخله نویسنده به بروز ابهام توجه شود، ابهام به دودسته ابهام عمدی و ابهام طبیعی تقسیم می‌شود. ناگفته پیداست که در ابهام عمدی، نویسنده یا به دلیل درگیر بودن در محدودیت‌های مختلف فرهنگی، اجتماعی و سیاسی نمی‌تواند سخن خود را شفاف بگوید و یا آن که با نوعی پیچیدگی خودساخته که اغلب نیز ریشه در پیچیدگی ذهن او دارد، می‌خواهد مخاطبان جستجوگر را در هر زمان و مکان به جانب نوشته‌های خود جذب کند و ذهنیت آن‌ها را به چالش بیشتری بکشد (شیری، ۱۳۹۰).



حال اگر احتیاج، مادر اختراع باشد، ابهام نیز مادر علم است؛ انگیزه‌هایی که انسان را به‌طور عملی به‌جانب اکتشافات علمی می‌کشاند. برای تکمیل خود پس از عرصه‌گشایی به حس احتیاج در درون، بلافاصله ابهام را به وجود می‌آورد تا چستی و چگونگی آن احتیاج و طریقه رفع آن را به پرسش‌گری بگذارد و آنگاه ذهن کنجکاو را به جستجو برای کشف راه‌های رسیدن به ابزارهای لازم برای برطرف کردن آن نیازها به کندوکاو و فعالیت‌های علمی وادارد و محصول نهایی این فرایند (احتیاج، ابهام و علم) تبدیل به همان چیزی می‌شود که نامش اختراع است؛ بنابراین احساس ندانستگی، آدم‌ها را به‌جانب کنجکاوی و پرسش‌گری می‌کشاند و پرسش‌گری، زمینه‌ذهنی آدم‌ها را برای دستیابی به پاسخ‌های مناسب به پویایی وامی‌دارد و محصول این پرسش و پاسخ، پیشرفت‌های علمی را به دنبال دارد. در این میان، بخش عمده‌ای از رسالت هنر و ادبیات، به دلیل سروکار داشتن با عاطفه، تخیل، تفکر و تولید پرسش‌گری در ذهن انسان است (شیری، ۱۳۹۰).

وجود کلمات چندمعنایی و ابهام‌زا^۱، باعث بازیابی حجم زیادی از مدارک نامرتبط یا کم‌ربط با نیاز اطلاعات کاربر می‌شود؛ در نتیجه بازیابی اطلاعات، اثربخشی لازم را نخواهد داشت. این مورد به‌ویژه برای کلمات چندمعنایی دوچندان می‌شود. در واقع کلمات ابهام‌زا یکی از مهم‌ترین چالش‌های پردازش زبان طبیعی به‌شمار می‌آیند. برای نمونه، در زبان فارسی به‌دلیل ساختار خاص، تعداد کلمات ابهام‌زا و مشکلات ناشی از آن‌ها در پردازش زبان فارسی نسبتاً زیاد است. در زبان انگلیسی نیز این مشکل وجود دارد (ستوده و هنرجویان، ۱۳۹۱).

سیستم‌های ذخیره و بازیابی اطلاعات^۲ برای رفع کلمات ابهام‌زا به‌صورت خودکار قادر به تشخیص عبارات، اصطلاحات و واژه‌های ابهام‌زا نیستند و به‌تنهایی توانایی ابهام‌زدایی^۳ موارد ابهام‌زا را ندارند. متخصصان برای رفع موارد ابهام‌زا از روش‌های مختلفی استفاده می‌کنند. به نظر می‌رسد این تجربه بیشتر کاربران اینترنتی در دنیای وب است که گاهی صفحات بازیابی شده توسط موتورهای کاوش، حاوی نیازهای اطلاعاتی موردنظر آن‌ها نیستند (برندسن^۴، ۲۰۱۲).

منظور از عوامل علی در این پژوهش، عواملی است که به‌طور مستقیم باعث ایجاد ابهام می‌شوند. این عوامل شامل عدم وضوح در تعریف مفاهیم، تنوع در داده‌ها و کمبود اطلاعات است و منظور از عوامل مداخله‌گر به عواملی اشاره دارند که می‌توانند بر رابطه بین عوامل علی و ابهام تأثیر بگذارند (زمانی، ۱۴۰۳) و پژوهشگران قصد دارند در این پژوهش با بهره‌گیری از رویکرد نظریه داده‌بنیاد عوامل فوق را مورد شناسایی و جمع‌بندی قرار دهند.

ازجمله فواید این پژوهش، می‌توان چنین بیان نمود که شناسایی عوامل علی و مداخله‌گر^۵ ابهام‌زایی در سیستم‌های ذخیره و بازیابی اطلاعات می‌تواند راهنمایی مناسب برای متخصصان رشته‌های دست‌اندرکار طراحی سیستم‌های مورد نظر (زبان‌شناسی، هوش مصنوعی، کامپیوتر، فناوری اطلاعات و علم اطلاعات)

1. Multi-meaning and Ambiguous words
2. Systems of information retrieval
3. Disambiguation
4. Berendsen
5. Causal and Interferer factors



در جهت رفع ابهام‌های موجود در سیستم‌های فعلی باشد و در نهایت طراحی سیستم‌هایی قابل اطمینان برای کاربران نهایی در مرحله بازیابی اطلاعات را به همراه آورد. به این منظور در این مقاله به دو پرسش زیر پاسخ داده خواهد شد.

- الف. عوامل علی ابهام‌زا در سیستم‌های ذخیره و بازیابی اطلاعات کدام‌اند؟
ب. عوامل مداخله‌گر ابهام‌زا در سیستم‌های ذخیره و بازیابی اطلاعات کدام‌اند؟

۲- پیشینه پژوهش

در این قسمت، نخست پیشینه پژوهش با تمرکز بر موضوع پژوهش معرفی و مورد بررسی قرار می‌گیرد. نتایج جستجو در موضوع ابهام و عوامل علی و مداخله‌گر ابهام‌زایی در پایگاه‌های مختلف داخلی و خارجی نشان داد که تا به امروز این حوزه مورد مطالعه اندکی قرار گرفته است که در این قسمت به چند پژوهش اشاره می‌شود. اغلب پژوهش‌هایی که در این حوزه انجام شده است، کاربردی بوده و موضوعاتی از قبیل ابهام و ابهام‌زدایی معنایی، راهکارهای شناخت ابهام، مبتنی بر رفع ابهام معنایی کلمات و همچنین پردازش زبان طبیعی، هوش مصنوعی و زبان‌شناسی انجام شده است.

امپسون^۱ (۱۹۳۰) در کتاب جامع خود هفت نوع ابهام را مورد بررسی قرار داد. از اواخر دهه ۱۹۴۰، زیف^۲، نظریه قانون معانی^۳ را مطرح کرد. در سال ۱۹۵۰ کوپلان، قانونی ارائه داد که در یک متن خاص، در کلمه اطراف یک کلمه ابهام‌زا معادل تمام جمله ابهام‌زا هستند. در سال ۱۹۵۷ مسترمن^۴، تئوری مبنی بر یافتن معنای دقیق یک کلمه ابهام‌زا با استفاده از واژه‌نامه ارائه نمود (صادقی و محمدی نافچی، ۱۴۰۱). بارهیلل^۵ (۱۹۶۰)، در پژوهشی که در خصوص وضعیت ترجمه زبان‌های متفاوت در زمان خود انجام داد، به این نتیجه رسید که یافتن معنای کلمات، کار ساده‌ای نخواهد بود و با چالش‌های زیادی روبه‌رو خواهد بود.

خدابخشیان (۱۳۹۶) یک روش کاربردی را در جهت بهبود رفع ابهام معنایی کلمات در وردنت انجام داد که از ارتباط معنای کلمات استفاده کرد. مظفری و همکاران (۱۳۹۷) سامانه‌ای جهت رفع ابهام معنایی از حروف اضافه طراحی کردند. مظفری (۱۴۰۰) برای رفع ابهام نام نویسندگان مقالات الگوریتم خاصی را طراحی کرد. مظفری در این پژوهش، راهکاری را برای شناسایی رکوردهای متعدد که به یک نویسنده تعلق دارند، ارائه داد. رضایی دینانی و همکاران (۱۴۰۱) با برجسب‌گذاری هم‌نویسه‌های تخصصی، در کاهش ریزش کاذب متن کامل، تأثیر مستقیم ایجاد کرد. زمانی و همکاران (۱۴۰۲) موضوعی مرتبط با ابهام و ابهام‌زدایی در سیستم‌های بازیابی اطلاعات، در قالب مرور نظام‌مند انجام دادند. نتایج این پژوهش حاکی از این بود که موضوع ابهام معنایی کلمات و ابهام‌زدایی واژگان از دغدغه‌های اغلب پژوهشگران ایرانی بوده است. گونن و گولدبرگ^۶ (۲۰۱۶) پژوهشی در خصوص ابهام‌زدایی معنایی کلمه انجام دادند.



نشریه مطالعات دانش پژوهی

صفحه ۳۸

دوره ۳، شماره ۳

پیاپی ۹

1. Empson
2. Zifp
3. Law of Meaning
4. Masterman
5. Bar-Hillel
6. Gonen & Goldberg

نتایج پژوهش نشان داد که سامانه رفع ابهام معنایی در این پژوهش عملکرد خوبی داشته است. همچنین زوهروس و همکاران (۲۰۱۹) پژوهشی توسط شبکه عصبی بازگشتی عمیق برای موجودیت جغرافیایی ابهام‌زدایی پژوهشی را انجام دادند.

هریستیا و کلهن^۱ (۲۰۲۰) پژوهشی در حوزه ابهام‌زدایی معنایی کلمات انجام دادند که با پژوهش ژانگ و همکاران^۲ (۲۰۲۰) و الموسی و همکاران^۳ (۲۰۲۲) همسو و مشترک بودند. بوتنکو^۴ (۲۰۲۳) پژوهشی را با عنوان «ابهام‌زدایی در بازیابی اطلاعات در متون علمی و فنی» انجام داد. در این پژوهش انواع ابهام‌زدایی در بازیابی اطلاعات را در متون علمی و فنی توصیف می‌کند.

با بررسی پژوهش‌های انجام‌شده مشخص گردید که در موضوع ابهام‌زا در سیستم‌های ذخیره و بازیابی اطلاعات با راهکارهای متفاوتی صورت می‌گیرد. در این خصوص پژوهش خدابخشیان (۱۳۹۶) با الموسی، بن لامری و خوری (۲۰۲۲)، همسو و قابل مقایسه هستند. در خصوص رفع ابهام معنایی کلمات، پژوهش‌های فراوانی در ایران و خارج از ایران صورت گرفته است. از جمله این پژوهش‌ها گونن و گولدرگ (۲۰۱۶) و پژوهش زوهروس، تابیک، والدیویا، مارتینز و هررا^۵ (۲۰۱۹) هم‌راستا هستند و همه آن‌ها بر موضع مشترک «ابهام‌زدایی معنای کلمات» تأکید دارند. روش پژوهش ایسیری واردانا (۲۰۲۴)، از لحاظ روش پژوهش با روش پژوهش خدابخشیان (۱۳۹۶) همسو و مشترک است؛ ولی از لحاظ درک و شناخت ابهام با موضوع پژوهش حاضر همسو است.

اگرچه پژوهش‌های انجام‌شده در ایران و خارج از ایران در تمام حوزه‌های مربوط به رفع ابهام وارد شده و راه‌کارها و روش‌های مناسبی که انجام شده و به مرحله عملیاتی رسیده، موضوع مدیریت و کشف عوامل علی ابهام‌زایی و ابهام‌زدایی در سیستم‌های بازیابی اطلاعات با تأکید بر زبان فارسی از نظرها پنهان مانده است؛ اما آنچه در پژوهش‌های مربوط به ابهام‌زایی دیده می‌شود، خلأ پژوهش مربوط به شناسایی عوامل علی ابهام‌زایی و مداخله‌گر ابهام‌زایی در سیستم‌های ذخیره و بازیابی اطلاعات است.

۳- روش‌شناسی پژوهش

پژوهش کیفی از مجموعه‌ای از روش‌ها، فنون و ابزارها مثل مطالعه موردی، تجربه شخصی، درون‌نگری، شرح حال، روایت، مشاهده مشارکتی، مشاهده غیرمشارکتی، مصاحبه عمیق، انواع بشر ساخته‌ها (مصنوعات) فرآورده‌های فرهنگی و انواع متون مبتنی بر مشاهده، متون تاریخی، متون تعاملی و بصری استفاده می‌کند و گستره وسیعی از روش‌های تأویل و تحلیل‌های تفسیری را می‌پوشاند که یکی از این روش‌های تأویل و تحلیل در آن تئوری داده‌بنیاد است (گلیرر، ۲۰۰۲).

1. Hristea & Colhon
2. Zhang, Zhang, Tang, Hua & Sun
3. Almousa, Ben-Lameri & Khoori
4. Butenko
5. Zuheros, Tobik, Valdivia, Martinz & Herera
6. Glaser



رویکردهای اصلی در نظریه داده‌بنیاد عبارت‌اند از رویکرد نظام‌مند اشتراوس و کوربین^۱، رویکرد ظهور یابنده گلنر و رویکرد برساخت‌گرایی چارماز^۲ (کرسول^۳، ۲۰۰۹) بر اساس رویکرد نظام‌مند توسط اشتراوس و کوربین (۱۹۹۰) طرح‌ریزی شده و به‌منظور تحلیل داده‌ها و ارائه نظریه، سه مرحله کدگذاری پیشنهاد شده است که عبارت‌اند از کدگذاری باز، محوری و انتخابی. در ظهور یابنده گلنر (۲۰۰۲) مفروضات از پیش تعریف‌شده‌ای وجود ندارد و سه مرحله کدگذاری پیشنهادشده که عبارت‌اند از کدگذاری باز، انتخابی و نظری. در رویکرد برساخت‌گرایی چارماز (۲۰۰۶) عنوان می‌شود که دیگر طرح‌های ارائه‌شده در نظریه داده‌بنیاد اثبات‌گرایانه هستند و باید به‌جای استفاده از نمودارها، نقشه‌های مفهومی و واژه‌های پیچیده در ارائه نظریه، بیشتر بر دیدگاه‌ها، ارزش‌ها و باورهای افراد تأکید کرد (گلنر، ۲۰۰۵). در این پژوهش از دیدگاه عینیت‌گرای گلنری داده‌بنیاد استفاده شد.

پژوهش حاضر درصدد کشف عوامل علی و مداخله‌گر با توجه به داده‌های فارسی‌زبان است. روش داده‌بنیاد که امکان مطالعه تجربه زیسته و مکاشفات ذهنی و علمی متخصصان موضوعی را فراهم می‌آورد، برای انجام این پژوهش مناسب تشخیص داده شد. در این روش، نمونه‌گیری زمانی خاتمه می‌یابد که هر مقوله به اشباع نظری برسد؛ بنابراین حجم نمونه در ابتدای پژوهش قابل محاسبه نیست. ابزار پژوهش از نوع مصاحبه عمیق بر اساس پرسش‌های پژوهش بود. مصاحبه‌ها در محل کار مصاحبه‌شوندگان انجام شد. مجموع زمان صرف شده برای مصاحبه‌ها ۱۴۴۰ دقیقه بود. روش نمونه‌گیری برای انتخاب نفر اول از نوع هدفمند و نفرت‌بعدی گلوله برفی بود. با انجام مصاحبه شانزدهم، اشباع نظری حاصل شد و جهت اطمینان، دو مصاحبه دیگر نیز صورت گرفت، اما داده جدیدی به دست نیامد. درنهایت تحلیل داده‌ها بر اساس ۱۸ مصاحبه انجام شده، صورت پذیرفت.

جامعه آماری پژوهش حاضر که منطبق با معیارهای پژوهش بودند، در جدول زیر است:

جدول ۱، اطلاعات جمعیت‌شناختی مصاحبه‌شونده‌ها

ردیف	مکان خدمت	نسبت سازمانی	میزان تحصیلات	مرتبه شغلی	رشته	جنسیت	فراوانی	درصد
۱	دانشگاه سمنان	هیئت علمی	دکتری	دانشیار	هوش مصنوعی	مرد	۶	۳/۳۳٪
۲		هیئت علمی	دکتری	دانشیار	کامپیوتر	مرد		
۳		هیئت علمی	دکتری	استادیار	کامپیوتر	مرد		
۴		هیئت علمی	دکتری	استادیار	زبان‌شناسی	مرد		
۵		هیئت علمی	دکتری	استادیار	هوش مصنوعی	زن		
۶		هیئت علمی	دکتری	استادیار	زبان‌شناسی	مرد		
۷	دانشگاه پیام نور یزد	هیئت علمی	دکتری	استادیار	زبان‌شناسی	مرد	۱	۵/۵۵٪

1. Strauss and Corbin's Approach Systematic
 2. Charmaz
 3. Creswell



ردیف	مکان خدمت	نسبت سازمانی	میزان تحصیلات	مرتبه شغلی	رشته	جنسیت	فراوانی	درصد
۸	دانشگاه فردوسی مشهد	هیئت علمی	دکتری	استادیار	کامپیوتر	مرد	۱	۵,۵۵%
۹	دانشگاه آزاد- واحد سمنان	هیئت علمی	دکتری	استادیار	ادبیات فارسی	مرد	۱	۵,۵۵%
۱۰	مرکز منطقه‌ای علوم و فناوری شیراز	هیئت علمی	دکتری	استادیار	علم اطلاعات	زن	۳	۱۶/۶۵%
۱۱		هیئت علمی	دکتری	دانشیار	زبان شناسی	مرد		
۱۲		هیئت علمی	دکتری	استادیار	هوش مصنوعی	مرد		
۱۳	دانشگاه آزاد- واحد گیلان	هیئت علمی	دکتری	استادیار	فناوری اطلاعات	مرد	۲	۱۱/۱%
۱۴		هیئت علمی	کارشناس ارشد	مربی	فناوری اطلاعات	مرد		
۱۵	کتابخانه عمومی چهارم	هیئت علمی	دکتری	استادیار	علم اطلاعات	زن	۱	۵,۵۵%
۱۶	دانشگاه آزاد- واحد آباده	هیئت علمی	دکتری	دانشیار	زبان شناسی	زن	۱	۵,۵۵%
۱۷	پژوهشگاه علوم و فناوری اطلاعات ایران	هیئت علمی	دکتری	استاد	علم اطلاعات	مرد	۲	۱۱/۱%
۱۸		هیئت علمی	دکتری	استادیار	زبان شناسی	زن		
جمع	-	-	-	-	-	-	۱۸	۱۰۰%

با توجه به جدول ۱، بیشترین تعداد مصاحبه‌شوندگان دارای، مرتبه استادیار (۱۲ نفر) و کمترین تعداد (۱ نفر)، مرتبه استاد بودند. به لحاظ جنسیت، (۱۳ نفر، مرد) و (۵ نفر، زن) بودند و از لحاظ مدرک تحصیلی (۱۷ نفر مدرک تحصیلی دکتری) و (تعداد ۱ کارشناسی ارشد) داشتند.

مصاحبه‌ها به صورت حضوری یا تلفنی انجام شد و در مسیر مصاحبه به فراخور پاسخ مصاحبه‌کنندگان در صورت لزوم پرسش‌های تکمیلی از آنان پرسیده شد. مصاحبه‌ها با به کارگیری از تجهیزات ضبط صوت^۱ ضبط شده و سپس به صورت فایل ورد^۲ پیاده‌سازی گردید. آنگاه اظهارات مصاحبه‌شوندگان به سه شیوه کدگذاری (کدگذاری باز، کدگذاری محوری و کدگذاری انتخابی) انجام شد. نرم‌افزار مورد استفاده در این بخش مکس کیودا ۲۰ بود.

۴- یافته‌ها

با توجه به نتایج به دست آمده از مصاحبه‌ها و جداول مربوط به کدگذاری باز و تولید مفاهیم و کدگذاری انتخابی و محوری، به پرسش‌های پژوهش پاسخ داده شد.

جدول ۲، نمونه‌ای از کدگذاری باز و تولید مفاهیم از گزاره‌ها

ردیف	مفاهیم	گزاره‌ها
۱	ارتباط ابهام واژگانی با سیستم نوشتاری	ابهام اگر در سطح واژه باشد معمولاً به سیستم نوشتاری مربوط می‌شود.
۲	هم‌نویسه با ابهام سروکار دارد	هم‌آوا و هم‌نویسه، ما با ابهام در زبان سروکار داریم.
۳	تلاش برای ورود ابهام کمتر در نوشتار متن	در حوزه‌های تحقیقاتی ما سعی می‌کنیم در نوشتار متن علمی ابهام کمتری وارد کنیم.
۴	ابهام در واژه‌های هم‌نگاره یا هم‌نویسه	درواقع ما زبان نوشتاری در نظر می‌گیریم یعنی می‌رویم سراغ واژه‌هایی که یک‌شکل نوشته می‌شوند. به آن‌ها واژه‌های هم‌نگاره یا هم‌نویسه می‌گویند. یک‌جور نوشته می‌شوند.
۵	ابهام هم‌نویسه	کلمات که دارای ابهام هم‌آوا و هم‌نویسه هستند می‌توانند موجب ابهام واژگانی باشند.
۶	ابهام در شباهت نوشتار کلمات	مثلاً می‌گوییم شانام شکسته شما نمی‌دانید شانه تخم‌مرغ شکسته شانه سرم شکست یا کتف مفید است.
۷	ابهام در سطح جمله (چندمعنایی)	یک جمله بیش از یک معنا داشته باشد مثل «دوست ۲۰ ساله من». ممکنه دوستی که ۲۰ سالشه و یا این‌که دوستی که ۲۰ سال باهاش رفاقت داره.
۸	چندمعنایی کلمات	ابهام‌زایی دقیقاً به جملات و کلمات دارای ابهام در متن هستند.
۹	اشکال وجود ابهام معنایی در سیستم پردازش زبان طبیعی	یکی از اولین مشکلاتی که در هر سیستم پردازش زبان طبیعی وجود دارد، مسئله ابهام معنایی و ساختاری کلمات است.
۱۰	متن ^۱ هم به‌عنوان ابهام‌زا و هم به‌عنوان ابهام‌زدا	رفع ابهام خودش مستلزم قرار گرفتن آن در بستر است بسترها می‌توانند در متن باشند، و یا به‌صورت فیزیکی باشند مثلاً بگوییم من آن مرد را با تلسکوپ می‌بینم و در اینجا ابهام در سر جای خودش است ولی اگر اشاره کنی به خود آن مرد و بگوییم من آن آقا را با تلسکوپ می‌بینم و اشاره کند به آن مرد و آن بستر فیزیکی کمک کند تا بفهمد که موضوع اصلی چیست.



نشریه مطالعات دانش پژوهی

صفحه ۴۲

دوره ۳، شماره ۳

پیاپی ۹

جدول ۲، نشان‌دهنده نمونه‌ای از اولین مرحله کدگذاری از متن مصاحبه‌ها، تعداد ۲۴ کد مفهومی به‌عنوان خرده مقوله استخراج شد که مفاهیم به‌دست‌آمده، زیربنایی برای مراحل کدگذاری بعدی قرار گرفتند. همان‌طور که در جدول ۲ مشاهده می‌شود در مرحله کدگذاری باز نمونه‌ای از گزاره‌ها در مصاحبه‌ها بازبینی و به‌دقت و به‌طور متعدد به آن‌ها گوش داده شد و تمام نکات کلیدی و تأکیدی مصاحبه‌شوندگان مورد توجه قرار گرفت. سپس متن مصاحبه‌ها به‌صورت گزاره‌هایی ارائه گردید و بعد اقدام به کدگذاری باز شد. در مرحله اول کدگذاری، گزاره‌ها بر اساس نکاتی کلیدی و مهم به کدهایی با عنوان «مفاهیم» تبدیل شدند.

در مرحله بعدی کدهای مفاهیم مستخرج بر اساس مضامین و ویژگی‌هایشان دسته‌بندی و در دسته‌هایی به نام «خرده مقولات» جای گرفتند. در مرحله بعدی خرده مقولات مبنایی برای استخراج کدهای «مقولات اصلی» قرار گرفتند که کدهای مقولات اصلی از بطن خرده مقولات به وجود آمدند. در جدول ۳، کدگذاری اولیه (مفاهیم) و محوری (خرده مقولات) آورده شده است:

جدول ۳، نمونه‌ای از کدگذاری انتخابی و محوری (از گزاره‌های سؤالات مصاحبه)

مقوله‌ها	خرده مقوله
۱. ابهام نوشتاری	ارتباط ابهام واژگانی با سیستم نوشتاری
	هم‌نویسه با ابهام سروکار دارد
	تلاش برای ورود ابهام کمتر در نوشتار متن
	ابهام در واژه‌های هم‌نگاره
	یا هم‌نویسه
	ابهام هم‌نویسه
۲. ابهام معنایی کلمات	ابهام در شباهت نوشتار کلمات
	ابهام در سطح جمله (چندمعنایی)
	چندمعنایی کلمات
۳. ابهام‌زایی و ابهام‌زدایی در متن	اشکال وجود ابهام معنایی در سیستم پردازش زبان طبیعی
	بستر هم به‌عنوان ابهام‌زا و هم به‌عنوان ابهام‌زدا
	بافت غیرزبانی
	متن فیزیکی
۴. دخالت انسان در ابهام	بافت موقعیت
	ارتباط نامطلوب ابهام به سیستم‌های ذخیره و بازیابی اطلاعات
	عدم توازن شناختی انسان‌ها با عدم توازن سامانه
	بودن یا نبودن صداقت فرد نویسنده در ابهام‌زایی
	نقش نویسنده ادبی در ابهام‌زایی
	توانایی انسان در کم کردن ابهام
	وجود ابهام در صورت عدم اطلاعات و دانش کاربر
	عدم آگاهی کاربر به موضوع مورد نظر خود
	ناتوانی شناختی، زبانی و یادگیری انسان
	عوامل بیرونی به‌عنوان عوامل ابهام‌زایی
۵. ابهام‌زایی ذاتی	اجتناب‌ناپذیر بودن ابهام
	عدم قطعیت به دلیل ابهام ذاتی
	وجود ابهام ذاتی در زبان
	ریشه‌یابی ابهامات
	وجود تعاریف دقیق و شفاف در متون علمی



نشریه مطالعات دانش پژوهی

صفحه ۴۳

شناسایی عوامل علی و
مداخله‌گر ابهام‌زا در
سیستم‌های ذخیره ...

مقوله‌ها	خرده مقوله
	ابهام‌زایی خودکار
۶. ابهام واژگانی	ابهام یک واژه یا عبارت بیشتر در جمله اتفاق می‌افتد
۷. ابهام ساختار جمله	ابهام در سطح جمله
۸. ارتباط معنایی متن	ابهام در سطح متن
	وجود بیشترین ابهام در ساختار جمله
	ابهام در متن
	تفاوت درک معنا توسط فردی که عامدانه تولید می‌کند با خواننده متن ابهام‌زا
۹. ابهام عامدانه	عدم اهمیت شاعران و ادیبان در ابهام‌زدایی
	مطلوب بودن ابهام
	ابهام برای برقراری ارتباط به مسیرهای خاص
۱۰. ابهام چندمعنایی کلمات	ابهام‌زایی در متون رشته‌های علوم انسانی به‌عنوان یک اجبار
	وجود ابهام به‌عنوان یک ضرورت از دید نویسنده ادبی
	ورود اطلاعات ابهام‌زا در سیستم‌های ذخیره و بازیابی اطلاعات توسط نویسندگان
	واژه‌های چندمعنایی

یافته‌های جدول ۳ نشان می‌دهد تعداد ۱۷ کد مفهومی به‌عنوان خرده مقوله استخراج شد.

۱-۴- پرسش اول پژوهش: عوامل علی ابهام‌زا در سیستم‌های ذخیره و بازیابی اطلاعات کدام‌اند؟
با توجه به نتایج کدگذاری محوری و انتخابی عوامل علی ابهام‌زا در سیستم‌های ذخیره و بازیابی اطلاعات به شرح زیر است:

جدول ۴، داده‌های مستخرج از عوامل علی ابهام‌زا در سیستم‌های ذخیره و بازیابی اطلاعات

طبقه	مقوله‌ها	خرده مقوله	فراوانی	درصد
عوامل علی ابهام‌زا	۱- ابهام نوشتاری	ورود ابهام در نوشتار متن	۴	16/5
		ارتباط ابهام واژگانی با سیستم نوشتاری		
		ابهام هم‌نویسه		
		ابهام در شباهت نوشتار کلمات		
۲- ابهام معنایی کلمات	۲- ابهام معنایی کلمات	عدم رعایت ترمینولوژی	۳	12/5
		چندمعنایی کلمات		
		ابهام واژگانی		
۳- ابهام عمدی	۳- ابهام عمدی	ابهام‌زایی نویسنده در آثار خود	۲	8/4
		تولید ابهام عمدی در آثار ادبی		
	۴- ابهام‌زایی ذاتی	نقص در داده‌ها	۶	25
		تفاوت بین پرسش و پاسخ کاربران		

طبقه	مقوله‌ها	خرده مقوله	فراوانی	درصد
		ناهم‌خوانی مفاهیم کلمات	۳	12/5
		ابهام در اثر عدم آگاهی موضوعی کاربران		
		وجود ذاتی ابهام در کلمات هم‌آوا		
		تفاوت در استفاده از واژگان بین کاربران		
	۵-ابهام گفتاری	ارتباط کلمات هم‌آوا با ابهام	۳	12/5
		بافت زبانی		
	۶-ابهام ساختاری جمله	ابهام مرجع	2	8/4
		ابهام در سطح نحوی		
	۷-ابهام ساختاری در داده‌ها	تغییرات مکرر در داده‌ها	4	16/7
		نوسانات مکرر در داده‌ها		
		ناهمگونی داده‌ها		
		پیچیدگی ساختار داده‌ها		
جمع	۷		24	100%

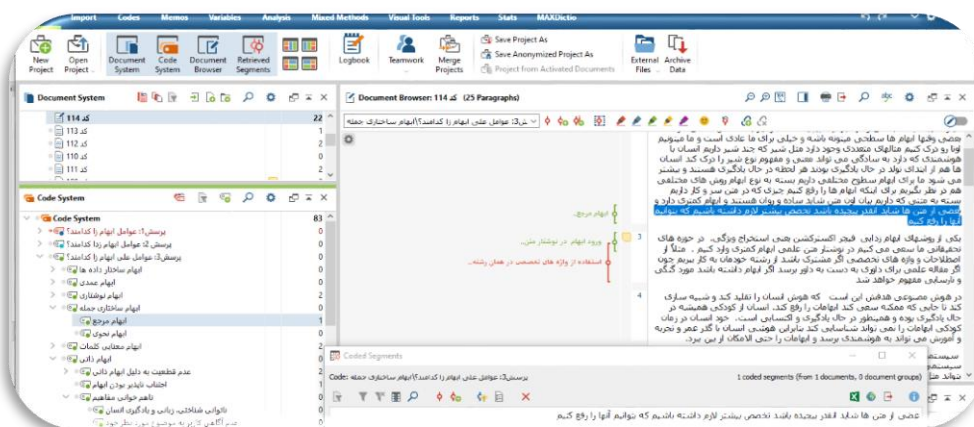


نشریه مطالعات دانش پژوهی

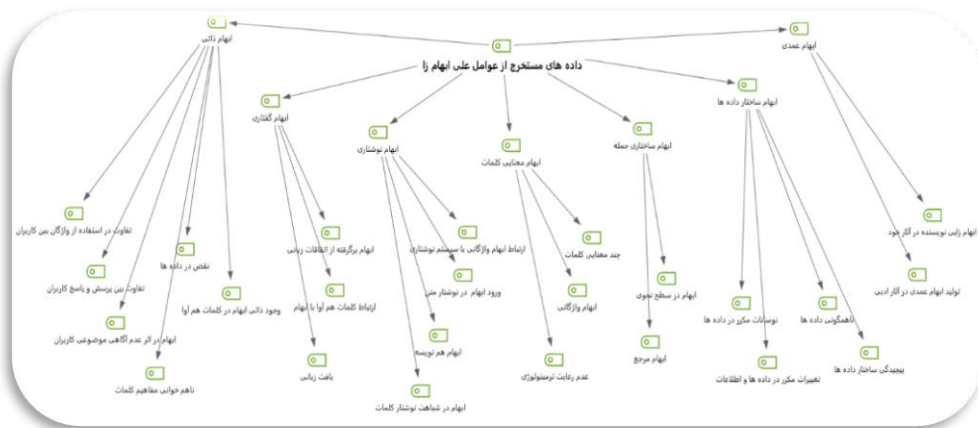
صفحه ۴۵

شناسایی عوامل علی و مداخله‌گر ابهام‌زا در سیستم‌های ذخیره ...

جدول ۴، نشان می‌دهد که بیشترین مقوله اصلی مربوط به مقوله ابهام‌زایی ذاتی (۶ خرده مقوله، ۲۵٪) و کمترین مقوله اصلی مربوط به ابهام عمدی و ابهام ساختاری جمله (۲ خرده مقوله، ۸/۴٪) هستند.



شکل ۱، نمونه‌ای از داده‌های مستخرج پرسش اول: «عوامل علی ابهام‌زا»



شکل ۲. داده‌های مستخرج از پرسش اول پژوهش: «عوامل علی ابهام‌زا»

۲-۴- پرسش دوم پژوهش: عوامل مداخله‌گر ابهام‌زا در سیستم‌های ذخیره و بازیابی اطلاعات کدام‌اند؟

با توجه به نتایج کدگذاری محوری و انتخابی عوامل مداخله‌گر ابهام‌زا در سیستم‌های ذخیره و بازیابی اطلاعات به شرح زیر است:



جدول ۵، عوامل مداخله‌گر ابهام‌زا در سیستم‌های ذخیره و بازیابی اطلاعات

طبقه	مقوله‌ها	خرده مقوله	فراوانی	درصد فراوانی
عوامل مداخله‌گر ابهام‌زا	۱- ابهام عمدی	عدم توجه به ابهام به‌عنوان مشکل در متون ادبی	۷	41
		دخالت انسان در بروز ابهام		
		راز ماندگاری و زیبایی آثار ادبی		
		زیبایی اثربخش ابهام در آثار ادبی		
		تأثیر مؤلف در ابهام		
		دخالت سیاست در بروز ابهام		
	۲- ابهام غیرعمدی	استفاده عبارات دوپهلو در جملات	۷	41
		ابر عامل در ابهام‌زایی		
		مداخله فرهنگ در ابهام‌زایی		
			ابهام در اثر کم‌سوادگی کاربران واردکننده اطلاعات در اینترنت	
تولید ناخواسته ابهام در قالب اطلاعات توسط کاربران اینترنتی				
عدم آگاهی اکثر مردم به ابهام زبان				
مداخله ویژگی زبان فارسی در ابهام				
	۳- الگوریتم‌های بازیابی اطلاعات	زمینه تولید ابهام برحسب خوانش یک متن	۲	۱۲
		عدم مهارت از تکنیک‌های جستجو در سیستم‌های اطلاعاتی		
		تغییر نامناسب در الگوریتم‌های بازیابی اطلاعات		
	۴- منابع نامعتبر	افزایش ابهام‌زایی باوجود منابع نامعتبر	۱	۶
جمع	۴		۱۷	100%

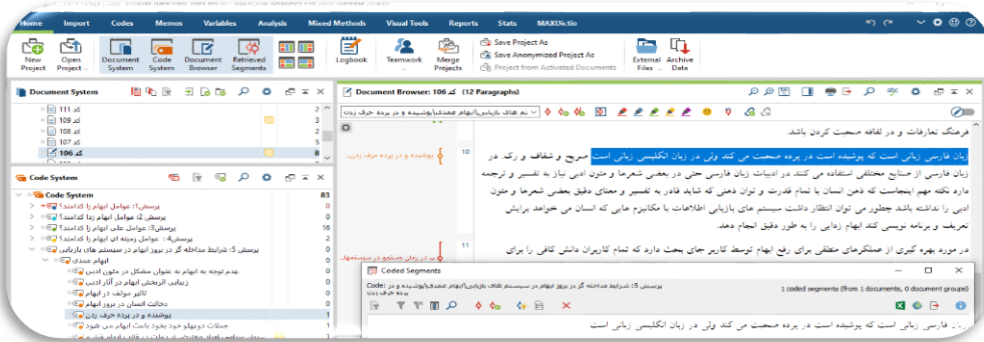
نشریه مطالعات دانش پژوهی

صفحه ۴۶

دوره ۳، شماره ۳

پیاپی ۹

جدول ۵، نشان داده می‌دهد که بیشترین مقوله‌های اصلی را مقوله‌های ابهام عمدی و ابهام غیرعمدی (خرده مقوله، ۴۱٪) و کمترین مقوله را منابع نامعتبر (۱ خرده مقوله، ۶٪) تشکیل داده‌اند.

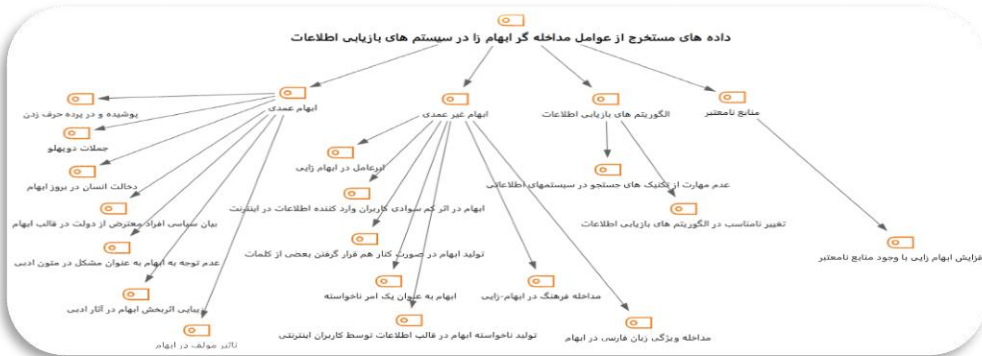


شکل ۳، نمونه‌ای از داده‌های مستخرج پرسش دوم: «عوامل مداخله‌گر ابهام‌زا»



نشریه مطالعات دانش پژوهی

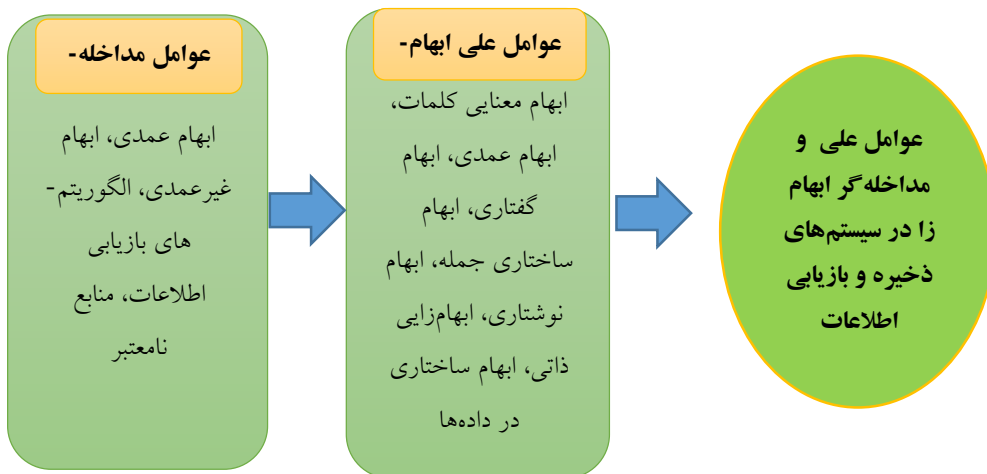
صفحه ۴۷



شکل ۴، داده‌های مستخرج از پرسش دوم پژوهش: «عوامل مداخله‌گر ابهام‌زا»

شناسایی عوامل علی و مداخله‌گر ابهام‌زا در سیستم‌های ذخیره ...

۳-۴- مدل نهایی عوامل علی ابهام‌زا



شکل ۵، مدل نهایی پژوهش

۵- بحث و نتیجه گیری

یافته‌های جمعیت شناختی مصاحبه‌شوندگان نشان داد که از نظر جنسیت تعداد متخصصان مرد بیش از زن و اکثراً دارای مدرک دکتری بودند. همچنین متخصصان در مرتبه استادیار بیش از دیگر مرتبه‌های علمی بودند.

در ادامه پرسش‌های پژوهش بررسی می‌شود.

بررسی نتایج به دست آمده در مورد عوامل علی ابهام‌زایی نشان می‌دهد که بر اساس نظر متخصصان مقوله‌های ابهام‌زا شامل ابهام نوشتاری، ابهام معنایی کلمات، ابهام عمدی، ابهام ذاتی، ابهام ساختاری جمله و داده‌ها هستند و «ابهام ذاتی» با ۶ زیر مقوله شامل نقص در داده‌ها، تفاوت بین پرسش و پاسخ کاربران، ناهمخوانی مفاهیم کلمات، ابهام در اثر عدم آگاهی موضوعی کاربران، وجود ذاتی ابهام در کلمات هم‌آوا، تفاوت در استفاده از واژگان بین کاربران، نسبت به دیگر مقوله‌ها، نقشی قوی‌تر در بروز ابهام دارد. جهت رفع این ابهام‌ها می‌توان به استفاده از الگوریتم‌های پردازش زبان طبیعی، بهبود ساختار جملات و داده‌ها، و طراحی رابط کاربری کارآمد اشاره کرد.

در مورد پرسش دوم نتایج نشان می‌دهد که از نظر خبرگان عوامل مداخله‌گر ابهام‌زا عبارت‌اند از: ابهام عمدی، ابهام غیر عمدی، الگوریتم‌های بازیابی اطلاعات و منابع نامعتبر. لازم به ذکر است در این پژوهش، عوامل علی ابهام‌زایی در مقایسه با عوامل مداخله‌گر به طور مستقیم باعث تولید ابهام شده و عوامل مداخله‌گر تأثیر غیرمستقیم بر این امر دارند. همچنین برخی زیر مقوله‌ها از جمله ابهام عمدی هم به صورت مستقیم و هم غیرمستقیم بر ایجاد ابهام مؤثرند؛ اما با شیوه‌هایی متفاوت تأثیر گذاری خود را اعمال می‌کنند. به نحوی که در عامل علی ابهام عمدی از نویسندگان و تولیدکنندگان آثار نشئت می‌گیرد و در بخش مداخله این زیر مقوله از بعد کاربری به دلیل کم‌سواد اطلاعاتی (اینترنتی) و مداخله‌های فرهنگی حاصل می‌شود.

نتایج به دست آمده با پژوهش‌های ابیسیری و اردانا (۲۰۲۴)، شیری (۱۳۹۰) و ژانگ و همکاران (۲۰۲۰) در خصوص عوامل مداخله‌گر ابهام‌زا همسو هستند. در مورد حذف حروف اضافه که از مقوله‌های عوامل علی ابهام‌زایی است، با پژوهش گونن و گولدبرگ (۲۰۱۶) و مظفری و همکاران (۱۳۹۷) همسو و مشترک است. در پژوهش بوتنکو (۲۰۲۳)، در مورد ابهام و تأثیرات آن در بازیابی اطلاعات با پژوهش حاضر همسو است. همچنین پژوهش نصری و مدرس خیابانی (۱۴۰۲) در مورد ابهام واژگانی که از خرده مقولات و زیرمجموعه ابهام معنایی کلمات است، با پژوهش حاضر همسو و مشترک است. پژوهش مهرآوران و قاسمی (۱۳۹۹) و نصری و مدرس خیابانی (۱۴۰۲) در مورد عوامل بیانی و دستوری همسو و مشترک هستند. شناسایی عوامل علی و مداخله‌گر ابهام‌زا در سیستم‌های ذخیره و بازیابی اطلاعات نقش بسزایی در بهبود کارایی و دقت این سیستم‌ها دارد. با توجه به اهمیت این مسئله، پژوهش‌های بیشتری در این زمینه پیشنهاد می‌شود. این پیشنهادها از قبیل:

- ارزیابی تأثیر ابهام‌زدایی بر رضایت کاربران در سیستم‌های ذخیره و بازیابی اطلاعات.
- ارزیابی بازخورد کاربران در استراتژی‌های ابهام‌زدایی برای سیستم‌های ذخیره و بازیابی اطلاعات با تأکید بر زبان فارسی.



نشریه مطالعات دانش پژوهی

صفحه ۴۸

دوره ۳، شماره ۳

پیاپی ۹

- توسعه سیستم‌های هوش مصنوعی برای تحلیل و شناسایی ابهامات در متون و اطلاعات به منظور ارائه راه‌حل‌های مناسب برای حل آن‌ها.
- همچنین روش‌های بهینه‌سازی برای شناسایی عوامل علی و مداخله‌گر ابهام‌زا توسعه یابند و بهبود کارایی سیستم‌های ذخیره و بازیابی اطلاعات را فراهم آورند.

۶- سپاسگزاری

این مقاله مستخرج از رساله دکتری نویسنده اول، با عنوان «کشف مطالعاتی عوامل ابهام‌زا و ابهام‌زدا در سیستم‌های ذخیره و بازیابی اطلاعات» است. نویسندگان مقاله مراتب سپاس خود را از معاونت محترم علمی دانشگاه آزاد اسلامی - واحد بابل، ابراز می‌دارند. همچنین از داوران محترم به خاطر ارائه نظرهای ساختاری و علمی سپاسگزاری می‌شود.

۷- منابع و مآخذ

- ایمانی، محسن (۱۳۹۲). *ابهام‌زدایی و ارزیابی اطلاعات استخراج شده از متن زبان طبیعی*. پایان‌نامه مقطع کارشناسی ارشد. دانشگاه علم و صنعت ایران.
- تدین، محمدعلی (۱۳۹۱). مطالعه مسائل جاری رفع ابهام در ترجمه ماشینی مبتنی بر مجموعه متون و ارائه راهکارهای جدید. پایان‌نامه کارشناسی ارشد. دانشگاه شیراز.
- چارماز، کتی (۲۰۰۶). *ساختن نظریه مبنایی، راهنمایی کاربردی برای تحلیل کیفی (ترجمه مهیار کاظمی موحد؛ حمیدرضا ایرانی، ۱۳۹۷)*. رسا.
- خدابخشیان، وحید (۱۳۹۶). *بهبود رفع ابهام معنایی کلمات در وردنت با استفاده از تشکیل گراف ارتباط حس کلمات*. پایان‌نامه کارشناسی ارشد. دانشگاه آزاد اسلامی شاهرود.
- رضایی دینانی، مینا؛ کربلا آقایی کامران، معصومه؛ و میرزاییان، وحیدرضا (۱۴۰۱). *تاثیر برچسب‌گذاری معنایی در رفع ابهام هم‌نویسه‌ای تخصصی از نظر ری‌زش کاذب در بازیابی متون علمی*. *مطالعات کتابداری و سازماندهی اطلاعات*، ۱(۳۳)، ۱۰۷-۱۲۴. doi:10.30484/nastinfo.2021.2914.2063
- زمانی، محمد؛ قیاسی، میترا؛ و طهماسبی، صفیه (۱۴۰۲). *ابهام و ابهام‌زدایی در سیستم‌های بازیابی اطلاعات (مرور نظام‌مند)*. *آستان قدس رضوی*، ۲۶(۴)، ۵۷-۸۱. doi:10.30481/lis.2024.429511.2125
- زمانی، محمد (۱۴۰۳). *کشف مطالعات ابهام‌زا و ابهام‌زدا در سیستم‌های ذخیره و بازیابی اطلاعات با تأکید بر زبان فارسی*. رساله منتشر نشده دکتری. دانشگاه آزاد واحد بابل.
- ستوده، هاجر؛ و هنرجویان، زهره (۱۳۹۱). *ربط در علم اطلاع‌رسانی*. *مطالعات ملی کتابداری و سازماندهی اطلاعات*، ۶۴، ۱۳۰-۱۳۲.
- شیری، قهرمان (۱۳۹۰). *اهمیت و انواع ابهام در پژوهش‌ها*. *فنون ادبی*، ۳(۲)، ۳۶-۱۵. doi:20.1001.1.20088027.1390.3.2.4.5
- صادقی، وحید؛ و محمدی نافچی، نگین (۱۴۰۱). *بازیابی واژه‌ها در بازشناسی گفتار خوانداری فارسی: تاثیر نوای گفتار*. *زبان پژوهشی دانشگاه الزهراء*، ۱۴(۴۴)، ۷۵-۵۹. doi:10.22051/jlr.2022.38936.2127
- عبدالله زاده، امیرحسین؛ مقیمی، سید محمد؛ و مزاری، ابراهیم (۱۴۰۳). *مفهوم‌پردازی رهبری اینترگرانه در سازمان‌های عمومی: رویکردی مبتنی بر نظریه داده‌بنیاد*. *مطالعات وقف و امور خیریه*، ۲(۱)، ۲۰۳-۲۲۶. doi: 10.22108/ecs.2024.140244.1084
- عباسی، مهسا (۱۳۹۷). *ابهام‌زدایی از لغات زبان فارسی، با استفاده از الگوریتم یادگیری عمیق*. پایان‌نامه کارشناسی ارشد. دانشگاه محقق اردبیلی.



قادرزاده، امید؛ و غلامی، فاطمه (۱۳۹۴). تحلیل تجربه‌های زنان سنندجی از فالگیری (مطالعه کیفی دلایل و پیامدهای فالگیری). *جامعه شناسی کاربردی*، ۲۶(۲)، ۱۰۷-۱۲۴.

doi: 20.1001.1.20085745.1394.26.2.7.3

مظفری، زهرا؛ تاکی، گیتی؛ صباغ‌جعفری، مجتبی؛ و یوسفیان، پاکزاد (۱۳۹۷). سامانه رفع ابهام از حروف اضافه در زبان فارسی با استفاده از قالب‌های معنایی. *پژوهش‌های زبانی*، ۹(۱)، ۹۹-۱۱۷. doi: 10.22059/jolr.2018.66652

مظفری، نیلوفر (۱۴۰۰). ارائه روشی مبتنی بر ژنتیک برای رفع ابهام نام نویسندگان مقالات. *پژوهش و مدیریت اطلاعات*، ۳۶(۳)، ۷۹۱-۸۱۶. doi: 10.52547/jipm.36.3.791

منصوریان، یزدان (۱۳۹۳). روش تحقیق در علم اطلاعات و دانش‌شناسی. سمت.

نصری، محمد؛ نصرت‌زاده، حیدر؛ و مدرس خیابانی، شهرام (۱۴۰۲). با هم‌آیی و ابهام واژگانی و مشکلات نارسایی‌های مربوط به معادل‌گزینی (سازی) آن‌ها در متون بانکی با استفاده از نظریه‌ی معناشناسی ساختارگرا. *پژوهش‌های زبان‌شناختی در زبان‌های خارجی*، ۱۳(۲)، ۱۷۷-۱۹۹. doi: 10.22059/jflr.2022.349034.989

Abeysiriwardana, M., & Sumanathilaka, D. (2024). A Survey on Lexical Ambiguity Detection and Word Sense Disambiguation. *arXiv preprint arXiv:2403.16129*. <https://doi.org/10.48550/arXiv.2403.16129>

AlMousa, M., Benlamri, R., & Khoury, R. (2022). A novel word sense disambiguation approach using WordNet knowledge graph. *Computer Speech & Language*, 74, 101337. <https://doi.org/10.1016/j.csl.2021.101337>

Bar-Hillel, Y. (1960). The present status of automatic translation of languages. *Advances in computers*, 1, 91-163. <https://doi.org/10.1016/j.csl.2021.101337>

Berendsen, R., Kovachev, B., Nastou, E. P., De Rijke, M., & Weerkamp, W. (2012). Result disambiguation in web people search. In *Advances in Information Retrieval: 34th European Conference on IR Research, ECIR 2012, Barcelona, Spain, April 1-5, 2012. Proceedings* 34 (pp. 146-157). Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-28997-2_13

Butenko, I. (2023, October). Disambiguation in information retrieval in scientific and technical texts. In *AIP Conference Proceedings* (Vol. 2833, No. 1). AIP Publishing. <https://doi.org/10.1063/5.0151708>.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods Approaches* (3rd ed.). Sage Publications, Inc.

Empson, W. (1930). *Seven Types of Ambiguity*. Chatto.

Fakhrahmad, S. M., Rezapour, A. R., Zolghadri Jahromi, M., & Sadreddini, M. (2011). A new word sense disambiguation system based on deduction. In *Proceedings of the World Congress on Engineering* (Vol. 2, pp. 1276-1281).

Faust, M. (Ed.). (2011). *The handbook of the neuropsychology of language*. <https://onlinelibrary.wiley.com/doi/book/10.1002/9781118432501>

Glaser, B. G., & Holton, J. (2005). Staying open: The use of theoretical codes in grounded theory. *The Grounded Theory Review*, 5(1), 1-20.

Chen, P., Bowes, C., Ding, W., & Choly, M. (2012). Word sense disambiguation with automatically acquired knowledge. *IEEE Intelligent Systems*, 27(04), 46-55. <https://doi.org/10.1109/MIS.2010.134>

Hristea, F., & Colhon, M. (2020). The long road from performing word sense disambiguation to successfully using it in information retrieval: An overview of the unsupervised approach. *Computational Intelligence*, 36(3), 1026-1062. <https://doi.org/10.1111/coin.12303>

Lyons, J. (1970). *Semantics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511620614>



نشریه مطالعات دانش پژوهی

صفحه ۵۰

دوره ۳، شماره ۳

پیاپی ۹

- Zhang, D., Zhang, H., Tang, J., Hua, X. S., & Sun, Q. (2020). Causal intervention for weakly-supervised semantic segmentation. *Advances in Neural Information Processing Systems*, 33, 655-666. <https://doi.org/10.48550/arXiv.2009.12547>
- Zuheros, C., Tabik, S., Valdivia, A., Martínez-Cámara, E., & Herrera, F. (2019). Deep recurrent neural network for geographical entities disambiguation on social media data. *Knowledge-Based Systems*, 173, 117-127. <https://doi.org/10.1016/j.knosys.2019.02.030>



نشریه مطالعات دانش پژوهی

صفحه ۵۱

شناسایی عوامل علی و
مداخله‌گر ابهام‌زا در
سیستم‌های ذخیره ...