

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