



Review Article

An Explanation of the False Drop in Information Retrieval in Human-Computer Interaction

Leila Jabbari^{*}, Hassan Mantegh, Mila Malekolkalami

Faculty of Management and Economics, Tarbiat Modares University, Tehran, Iran

Email address:

le_jabbari@yahoo.com (L. Jabbari), Hassan.mantegh@gmail.com (H. Mantegh), mila@yahoo.com (M. Malekolkalami)

^{*}Corresponding author

To cite this article:

Leila Jabbari, Hassan Mantegh, Mila Malekolkalami. An Explanation of the False Drop in Information Retrieval in Human-Computer Interaction. *American Journal of Information Science and Technology*. Vol. 5, No. 3, 2021, pp. 80-85. doi: 10.11648/j.ajist.20210503.14

Received: February 18, 2021; **Accepted:** August 30, 2021; **Published:** September 10, 2021

Abstract: Nowadays the field of information retrieval is quickly developing. Research on the human-computer interactions and designing of the interface is likely to lead to better results and greater empowerment for users and creators of information. One of the main and special issues related to the relationship between humans and computers in information storage and retrieval systems is the problem of accurate and correct information retrieval based on the user's information needs, due to there are still many false drops in the user's retrievals, although have the many advances come in this field. On the other hand, the human-computer interface environment is not well understood, compared to other forms of information retrieval, due to the complexity of human perception and detection compared to computer systems and the more difficult the detection and measurement of human motives and behaviors. In this paper, after reviewing the information storage and retrieval systems and their various evaluation metrics, the issue of the false drop and its effect on the human-machine interactions is discussed and finally, some solutions are presented. The principles of user interface design, providing informative feedback, reducing working memory load, providing alternative interface environments for novice and skilled users and visualizing information are among the proposed solutions of the present study.

Keywords: Information Retrieval, Human-Machine Interaction, False Drop, Information Needs

1. Introduction

When users log in to an information access system, they often have a vague understanding of how they access the information. Information retrieval can be called a sensitive and multidimensional process; because researchers agree that the user interface environment should ultimately help to better understand and articulate information needs. It should also help users formulate queries, choose from retrieved results, understand search results, and track search progress. Therefore, the human-computer interface requires more attention than other aspects of information retrieval; and this is largely due to the more complex cognitive features and perceptual power of man than computer systems, and consequently, the more difficult process of identifying and measuring his motivations and behaviors.

One of the most complex issues in the relationship between

humans and computers is the age-old problem of language, which acts as a key factor in their interaction in the exchange of information between them. Obviously, the internal language of a computer, that is, the language used to exchange information between computer components and to process information, must be chosen in a way that is compatible with the properties and characteristics of the computer components. However, no language can be the ideal language for direct communication between computers and humans just by relying on this feature. For this purpose, for the exchange of information, it is better to use language that is used in communication between humans and is used as the language of spoken and written words, the language of mathematical symbols, geometric shapes, and the like. The use of such language in human-computer communication will be suitable not only for entering basic data into the computer but also for retrieving data from the computer, as well as for the

programming language of computer operations [1].

Relationship feedback, in its original form, refers to an interactive cycle in which the user selects a small set of documents that he considers relevant to his query of the system. The system then uses the features taken from the relevant selected documents to review the main query. In the next step, this question is reviewed and then executed, and finally, a new set of documents is retrieved [2].

Undoubtedly, the main purpose of the information retrieval system is to retrieve information related to users' search. Another influential component in relevance is the use of Boolean algebra or set-based search and retrieval logic, which is well structured and easily exposed to computer processing. The basic premise of this logic is that the documents (including answers, facts, data, etc.) retrieved are the relevant ones to the user's question, and not retrieved ones are irrelevant. Even in some systems, the documents are evaluated, sorted and presented in a kind of order according to the degree of relevance and after reaching a certain threshold. Nevertheless, it is assumed that the basis for retrieving documents or not is whether they are relevant or not [3].

On the other hand, due to the lack of attention to the concept of relevance and the belief that relevance is not suitable for practical purposes, other terms have been proposed. Such as usefulness, appropriateness, application, etc which have been developed based on the idea of retrieving relevant documents.

In these cases, although these words do not contain "relevance", but the fact is that the same comprehension and concept of relevance prevail over them, and regardless of what words are used to express this concept, they all intend to offer new solutions for further development of the retrieval of relevant information by using new methods [3].

Of course, despite many advances that have been made in the field of relevance, there are still many issues regarding that; because relevance is a human-related concept, not just a system, and human concepts are hard and confusing.

Relevance is complex, multifaceted, and thought-provoking. Due to its human dimension same as other human concepts. Its continuous development requires continuous research and identification of new complex issues that arise in the future [3].

Therefore, it can be acknowledged that relevance is more of a human concept than a computer-oriented one. Thus, how human concepts can be articulated clearly and unambiguously is a constant question. The development of human concepts in the sciences of the twentieth century shows that based on the reductionism that is prevalent in the natural sciences, no clear answers can be given to the question of relevance. In other words, the meaning of effectiveness in systems related to human concepts is not the same as in engineering and commercial systems, and their equal study cannot bring the issue of relevance to its goal, and this is an important point that critics of the use of the concept of relevance have not considered [3].

In general, it should be acknowledged that in human-computer interaction, the retrieval of unwanted and unrelated information stems from various factors, which the

present study addresses. In other words, the main purpose of this study is to introduce the challenges of information retrieval systems in human-computer interaction. For this purpose, the following questions are raised in this research:

1. Why is there a false drop in data retrieval?
2. How can the false drop be reduced?

2. Literature Review

2.1. Information Storage and Retrieval Systems

Information retrieval systems represent, store, and retrieve information and knowledge related to users' specific needs and questions. The user expresses his needs with words from natural language in the form of a question. This query is compared with indexed documents by the system and based on a specific algorithm the most similar documents to the user query are presented. Ultimately, it is the user who measures the relevance to his needs [4]. Data retrieval is an ambiguous process. This ambiguity is due to the ambiguity of the elements involved in it, namely natural language, information needs, relevance, and so on [2].

The historical background of computer retrieval systems goes back to the development of automatic text indexing methods in the 1960s; this method is still widely used by information providers as well as in Internet services. In the 1970s, Boolean and vector space models and later probabilistic recovery models were developed. However, their commercial acceptance as partial compliance models (as opposed to the initial full compliance models) in Internet search engines, which present the results in a ranked list, was delayed until the 1990s [2].

Today, information retrieval systems play an important role in information concentration and knowledge development [5] as knowledge management is inconceivable in order to access a significant amount of information and extensive textual collections without this system. In our time, the use of the Web and search engines is considered part of everyday life, embedded with the methods, tools, and approaches of information retrieval systems. In this regard, like many other concepts and processes, human beings evaluate and comment on the quality of work.

2.2. Evaluation of Information Retrieval Systems

When it comes to evaluation, we mean judging quality and efficiency. According to Saraswick (1995), evaluation has a major influence on research, development, and information retrieval applications. In addition, Baiza-Yates and Ribero-Neto (2006) showed that in evaluating information retrieval systems, two general aspects of functional analysis and performance evaluation of the system are addressed. In the functional analysis section of the system, issues such as testing each function with the criteria of effectiveness (objectives), cost-efficiency (final costs), and cost-benefit (profit from services), as well as error analysis, are considered.

The problem of the false drop is of special importance in evaluating the efficiency of storage and processing systems;

thus, the first discussion about relevance was in the early 1950s. The discussion was essentially about irrelevance or false drop, that is, the unwanted information that the data retrieval system retrieves, not the relevance itself. The first was the recognition of the use of retrieving and accuracy as metrics to determine retrieving efficiency in which relevance was an underlying criterion [6].

Also, the performance of the system is evaluated in terms of space and time and other criteria, the most important of which is relevance. In the field of information retrieval texts, when discussing the evaluation of information retrieval systems, it mostly means the evaluation of system retrieval performance in terms of relevance [7].

Gradually, however, several other metrics were introduced in the field of relevance that could never find a place in this field. Accuracy and retrieval metrics have always remained the standard metrics for performance; This is because these metrics measure the degree of compliance between what the system has done to retrieve and build a connection (system relevance), or what it is unable to do and what the user perceives or evaluates as a connection (user relevance). Nevertheless, relevance in measuring information retrieval efficiency is an underlying criterion and still remains [3].

The purpose of explaining this section was to distinguish between evaluating the entire retrieval system, its performance, and retrieving the system in terms of relevance; however, in some of the proposed measures in addition to relevance, performance evaluation concerning aspects of time and space has been considered [7].

The purpose of evaluating the whole system is to be prepared to respond to funding sources and justify the system's performance and prove its value, to ensure the continuation of appropriate services and effectiveness (by comparing the desired level of performance and current level), to solve system problems (review strengths and weaknesses) And obtaining information to plan future performance or design and implement new systems [8].

In evaluating the performance of a system, the shorter the response time and the smaller the data space used, the better condition the system is in which is related to the performance of indexing structures, operating system interaction, communication channel delays, and the consequences of software layers running in the background [7].

In general, a good system is one that meets the needs of the user.

Meeting these needs depends on the good performance of the system in various dimensions. The quality of responses depends on the most important dimensions as information needs, response speed, and the system user interface; the evaluation of the first case is the most difficult among the three.

On the other hand, data retrieval does not solve the problem of retrieving information about a topic or category while providing a solution for users of a database system. For having an effective information retrieval system in meeting the user's information needs, it must somehow interpret the content of the information items in a work, and rank them based on how

relevant they are to the user's query. This content interpretation of a document requires the extraction of syntactic and semantic information from the text of the document and then uses it to adapt to the information needs of the user.

The problem is not only knowing how to extract this information but also knowing how to use it to determine relevance. Therefore, the issue of relevance is at the heart of information retrieval. In fact, the first goal of an information retrieval system is to retrieve all the documents that are relevant to the user's query while retrieving the least number of irrelevant documents.

3. Research Method

According to the objectives of this study, which uses printed and electronic texts and sources to identify issues and solutions to the questions raised, a review and library methods were used. In the review method, first, the backgrounds and literature related to the topic were searched and identified by searching in databases, and then a direct reference was made to the specialized book in the field of information retrieval and related articles. Indexing tool was also used for the conceptualization and compilation of theoretical literature.

4. Research Findings

4.1. *The Relationship Between Comprehensiveness, Obstruction and False Fall*

Criteria of comprehensiveness and constraint are widely used to compare the ranking results of non-interactive systems [7].

These two concepts are the basic concepts in information retrieval. There are two types of judgments to define comprehensiveness and constraint:

1- Relevant documents and 2- Irrelevant documents and in relation to these two judgments, another issue is:

"Relevance: The question versus the need for information" [9], that is, what is meant by the relevance or non-relevance of a document?

The exact relevance of the document to the question or to the needs of the researcher?

Sometimes the retrieved source is 100% relevant to the question and contains search keywords, both in the title and in the abstract, but has nothing to do with the user's information needs.

"Application satisfaction is measured only by the fact that the results obtained are relevant to her needs and not to her question," says Schutz (2013).

However, in most cases in the evaluation of relevance, judgment is still made on the question, not the need for information and is evaluated with two main criteria of obstruction and comprehensiveness [10].

Comprehensiveness is defined as the ratio of the number of documents retrieved to the total number of related documents in the file.

While obstruction or accuracy refers to the number of documents retrieved from the total number of documents retrieved from the file [8].

Comprehensiveness and obstruction are associated with false fall.

Especially when we expand the exploration in order to achieve high comprehensiveness, that is, whatever we expand the exploration and want to increase the comprehensiveness, we have difficulty in obstruction and the unwanted stagnation recovered increases, which is the false fall.

As a result, the greater the comprehensiveness, the higher the false fall, and the higher the obstacle, the lower the false fall [11].

Therefore, it can be said that comprehensiveness and barrier are not appropriate measures to investigate and control the false fall in data storage and retrieval systems.

Information retrieval systems create relevance, receive a query, process it using several algorithms, and present what they identify as relevant. Individuals extract relevant information or information entities and link and interpret information or information entities to the issue and other factors. Information retrieval systems adapt queries to the information entities in their repositories to find query-related items, possibly rank them, and output results. Users get results and extract what they think is relevant; But users' inference from the results, questions, and information entities used by the system for adaptation goes beyond merely matching between noun phrases. Moreover, they can find other intelligence entities or other information related to their problem that the system has not retrieved for various reasons, such as not being included in the initial question [12].

Harter (1992) provides excellent examples of extracting relevance beyond the retrieved position. He specifically presents his favorite examples and then analyzes several articles that are not directly related to the topics as discussed, but they are somehow related. With the help of this example, he shows that the extraction of relevance from these articles is very different in terms of individual cognitive status (psychological relevance) from the thematic relevance desired by the system. Thematic relevance is limited to the limited format of language and causes the user to forget [12].

One of the issues related to information relevance is the issue of false drop and the negative relationship between these two. In other words, the increase in false drop indicates a decrease in the relevance of the retrieved information based on user needs. False drop with unwanted and irrelevant information retrieval is one of the most important challenges of information retrieval systems. Nizo defines false drop as retrieving irrelevant evidence. While the terms index and user request match, their meanings are not related due to misalignment or incorrect combination of terms [11].

Hartley and other people consider Intra-tissue adaptation to be the main cause of false drops. Therefore, recognizing the types of extra-tissue matches is essential to improve search methods, and some of them that have been identified are:

4.1.1. Multiple Meanings

False drops occur when a term in the evidence index is semantically different from a similar term in the application. These false drops, which result from polysemous words, occur more frequently in continuous search. So when the same word has different meanings in the document and the request, the possibility of a false drop increases.

4.1.2. Terms Outside the Phrase

Such false drops occur when the phrase in the application or document is not considered as a single unit. Thus, a single word may appear in the phrase as the title of the document or as a separate term in the request (or vice versa). Wrong matching occurs when a term matches a word phrase outside its context.

4.1.3. Incorrect Assignment of Index Term

The third category of intra-tissue adaptation is when the adaptive term in the document and the application has more or less the same meaning, but the term in the document does not indicate its main subject. Manually assigning subject headings can prevent this from happening, but it reduces the degree of comprehensiveness and precision because only the most general subject of each document is indexed. However, the above factors lead to a false drop or, in other words, the retrieval of irrelevant documents [13]. Therefore, various reasons cause false drop, including:

- 1) The use of natural language in information retrieval due to the variety of words used to express the subject of the document and creating a low precision causes a false drop. So using controlled languages can be helpful. This challenge occurs when the index term of the document and the keywords searched by the user are similar but semantically different. Thus, not using polysemous words can prevent false occurrence.
- 2) When a single word is matched with similar words in a phrase, then the retrieved documents are related to that phrase and not to the single word the user wants. Thus a false drop occurs.
- 3) When the index term assigned to the document is incorrect and does not reflect the thematic content of the document, in this case, although the match between the index term and the user keyword is done correctly, the retrieved document is irrelevant. In this case, it is necessary to pay attention to the correct indexing and understanding of the content of the document.

Thus, one should be very careful in choosing indexing terms and indexing according to the policies of the system and the existing standards in order to reduce the amount of false drop and to achieve the efficiency of the information retrieval system.

4.2. Suggested Solutions to Reduce the Amount of False Drop

In general, the results of the present study show that the following operational strategies can be considered to reduce the false drop in human-computer interaction:

Comprehensiveness and precision are based on the assumption that the set of relevant documents for a query is the same at all times, and is user-independent. But different users may have different interpretations of which document is relevant and which is not. To solve this problem, some user-friendly metrics can be used, which are:

- 1) Expected length of exploration, for working with poorly arranged sets of documents;
- 2) Satisfaction: Satisfaction that only pays attention to relevant documents; and "failure" which pays attention only to irrelevant documents [7].
- 3) -Novelty ratio: is defined with a retrieved connection deduction that is unknown for use [7].
- 4) Coverage ratio: Coverage ratio is defined in the form of a deduction of known related documents (for the user) that have been practically retrieved [7].
- 5) Relative comprehensiveness: Relative comprehensiveness is obtained from the ratio of the number of documents with the relevance found by the system and the number of documents with the relevance that the user expects [7].
- 6) Achievement of comprehensiveness: Achievement of comprehensiveness is obtained by the ratio of the number of documents with the relevance that the user expects and the number of documents that are examined in the attempt to find the expected relevant documents [7].

4.3. Principles of User Interface Design

In a way, what makes the human-computer interface environment efficient? Expert Ben Schneiderman writes: "Properly designed, efficient computer systems create a positive sense of accomplishment, capability, proficiency, and transparency in the user community. When an interactive system is well-designed, the interface almost disappears, resulting in "Users can focus on their work, their development or their entertainment." Schneiderman (1997) cites the principles of user interface design as steps towards achieving these goals. These are the principles that are especially important in accessing information [7]:

- 1) Provide informative feedback: This principle is especially important for information access interface environments. If the user has control over how and when the feedback is provided, then the system provides an internal control situation.
- 2) Reduce working memory load: Access to information is a recurring process whose purpose is to change direction as it encounters information. One of the main ways in which information access interface environments can affect memory load is to provide mechanisms for continuous monitoring of choices made during the exploration process, allowing users to retry. Turn to strategies that have been temporarily set aside, move from one strategy to the next, and remember and not lose information and content during multiple exploration sessions. Another tool to help memory is to provide browsable information related to the current stage of the information access process. This includes tips on related

metadata or terms, and starting points for exploration.

- 3) Provide alternative interface environments for novice and skilled users: An important balance in the design of all interface environments is the balance between simplicity versus power. Learning to work with simple interface environments is easier and that easily comes at the cost of less flexibility and sometimes inefficient use. Powerful interface environments allow the informed user to do more work and have more control over the operation of the interface environment; But it can be time consuming to learn, and more difficult to remember for those who use the system from time to time. A common solution is to use a scaffolding approach [14]: provide the novice user with a simple interface environment that can be learned quickly, and this environment provides the basic efficiency of the application; But it is limited in terms of flexibility and power. Other interface environments are offered to more experienced users, giving them more control, more options, and more features, or perhaps even completely different interactive models. Proper design of the user interface creates Intrinsic and intuitive bridges between simple and advanced interface environments. Information access interface environments must challenge certain types of simplicity / power balance. One of these balances is the amount of information that is displayed about the functions of the search system itself. Users who are just starting out with a particular system or collection may not know enough to choose from the complex features of the system or collection-specific domain. They may not know the best way to weight words, or they may not be aware of the effects of word weighting on relevance feedback. On the other hand, users who have worked with a system and gained a sense of a subject are likely to be able to choose from the suggested words and consciously add that word to their query. Determining the amount of information that should be shown to the system user is one of the main design options for information access interface environments.
- 4) Information visualization: A growing but lesser-known field is the field of information visualization, which seeks to provide visual drawings of very large information spaces. Humans have adapted a lot to images and visual information. Images and graphics are attractive, especially if they are well designed. With visual representation, some types of information can be conveyed much faster and more efficiently than any other method. The prevalence of high-speed graphics processors and high-resolution color displays increases the interest in visualizing information. Visualization allows for two- or three-dimensional representations of physical phenomena. An example of scientific visualization is a color image of the design of ocean floor peaks and valleys, which provides a view of physical phenomena that (currently) cannot be photographed and the image is made up of data that represents subsurface phenomena [15].

5. Conclusion

Today, the field of information retrieval is developing rapidly. Research into how human-computer interaction and advances in interface design are likely to lead to better exploration results and greater empowerment for users and information creators. Research on human-computer interaction is difficult; this is because the field is relatively new and it is not easy to get strong results when studying users. Also, how human-computer interaction has been and is a topic in storage and retrieval systems. The issue of relevance is one of the issues in these systems. To determine the relationship between the retrieved information and the needs of the questioners in storage and retrieval systems, the criterion of comprehensiveness and precision to evaluate the performance of storage and retrieval information systems is of particular importance. However, there are still challenges to the proper functioning of the data retrieval system. These challenges arise from the retrieval of unwanted and irrelevant information by the computer (s). False drop is a major issue for data retrieval systems. The pioneers in the field of relevance (Moyers, 1950; Perry, 1951; Tab, 1955) were well aware that whatever was retrieved was not necessarily relevant. Their focus was more on irrelevance and unwanted retrieval. They found that false drops, scratches, false collusion, and over-retrieval were caused by internal malfunctions, inefficiencies of any representation (indexing language, coding, and classification), and inadequate methods of using them. (Saraswick, 1389: 34). False drop comes from a variety of factors. In the meantime, intra-tissue adaptation is the main cause of false drop. Polysemy, terms that are irrelevant to the context of the text, as well as incorrect assignment of the index term, are examples of types of extrinsic matching that lead to disruption and retrieval of unwanted information or false drop. Of course, the problem of false drop can be reduced by providing operational solutions. Principles of user interface design, providing informative feedback, reducing working memory load, providing alternative interface environments for novice and skilled users and visualizing information are among the proposed solutions.

References

- [1] Lerner, A. (1987). *Fundamentals of Cybernetics*. Translated by Kiomars Pariani. Tehran: Daneshpajooh.
- [2] Kiani, Mohammad Reza (2012). Approaches to evaluating information retrieval systems: backgrounds and perspectives ahead. *Library and Information Science*, Volume 16, Number 2, p. 243- 258.
- [3] Saraswick, Tafko (2010). *Relevance in Information Science*. Translated by Haidar Mokhtari, Abbas Mirzaei. Tehran: Chapar.
- [4] Mandal, T. (2008). Recent developments in the Evaluation of information Retrieval Systems: moving towards Diversity and practical Relevance. *Informatica*, 32, 27- 38.
- [5] Wang, Y. D. & Forgionne, G. (2006). A Decision – theoretic Approach to the evaluation of information Retrieval Systems. *Information processing & management*, 42, 863- 874.
- [6] Kent, A., Berry, M., Leuhns, F. U., and Perry, J. W. (1955). *Machine literature searching VIII. Operational criteria for designing information retrieval systems*.
- [7] Biza-Yates, Ricardo; Ribro-Neto, Bertie (2006). *New realms in information retrieval*. Translated by Ali Hossein Ghasemi. Tehran: Chapar.
- [8] Pao, Miranda Lee (2000). *Information retrieval concepts*. Translated by Assadollah Azad and Rahmatollah Fattahi, Mashhad, Ferdowsi University Press, Printing and Publishing Institute.
- [9] Schutze, H. 2013. *Introduction to Information Retrieval: Relevance Feedback and Quarry Extention*. <http://www.cis.uni-muenchen.de/~hs/teach/13s/ir/pdf/09expand.pdf> (accessed Aug. 21, 2013).
- [10] Hariri, Najla, Bab al-Hawaiji, Fahima; Farzandipour, Mehrdad and Somayeh Naderi (2014). Relevance evaluation criteria in information retrieval systems: Knowledge and Unknowledge. *Journal of Information Processing and Management*. 30 (1): 199-221.
- [11] Yousefi, Ahmad (1997). False drop in data storage and retrieval. *Notices*. 13 (1).
- [12] Harter, S. P. (1992). Psychological relevance and information science. *Journal of the American Society for information Science* 53 (4), 257- 270.
- [13] Hartley, J. et al (1996). Obtaining information accurately and quickly: are structured abstracts more efficient? *Journal of information science*, 22, 349-356.
- [14] Rosson, Mary Beth; Carroll, John M.; Bellamy, K. E. (1990). Smalltalk Scaffolding: A Case study of minimalist instruction. In *proc. Of the ACM Conference on human Factors in Computing Systems*, p. 423- 429.
- [15] Keller, P. R.; Keller, M. M. (1993). *Visual Cues: Practical Data Visualization*. IEEE Computer Society Press.