

Correction

# Correction: Mothe, J. Analytics Methods to Understand Information Retrieval Effectiveness—A Survey. *Mathematics* 2022, 10, 2135

Josiane Mothe 

INSPE, IRIT UMR5505 CNRS, Université Toulouse Jean-Jaurès, 118 Rte de Narbonne, F-31400 Toulouse, France; josiane.mothe@irit.fr; Tel.: +33-5-61556444

The author wishes to make the following corrections to this paper [1]:

1. In Abstract, (1) “It depicts how data analytics has been used in IR for a better understanding system effectiveness” should be “It depicts how data analytics has been used in IR to gain a better understanding of system effectiveness”; (2) “This review concludes lack of full understanding of system effectiveness according to the context although it has been possible to adapt the query processing to some contexts successfully” should be changed to “This review concludes that we lack a full understanding of system effectiveness related to the context which the system is in, though it has been possible to adapt the query processing to some contexts successfully”; (3) “This review also concludes that, even if it is possible to distinguish effective from non effective system on average on a query set” should be changed to “This review also concludes that, even if it is possible to distinguish effective from non-effective systems for a query set”.
2. In the third paragraph in the Introduction section, (1) “(it is so called because in this component each document will receive a score with regard to a given query)” should be changed to “(it is called this because in this component, each document will receive a score with regard to a given query)”; (2) The next sentence “the Ponte and Croft’ Language Modelling [4] and others” should add a comma before “and” to “the Ponte and Croft’ Language Modelling [4], and others”. (3) In the fourth paragraph, the sentence “Defining an information search process chain thus implies we decide which component will be used in each phase” should remove “thus”; (4) “For example, the number of terms to add to the query in the automatic query expansion phase, is one of the parameters to be decided” should remove “,” before “is”.
3. In the caption of Figure 1, change the number “4” to “four”.
4. The accessed date description “accessed 23 May 2022” behind every URL link should be changed to “accessed on 15 May 2022”.
5. The seventh paragraph was repeated, “For example, Terrier (<http://terrier.org/> accessed 23 May 2022 Terrier is an open source search engine that implements state-of-the-art indexing and retrieval functionalities adapted to use on TREC-like collections) popular platform [5] provides implementations of many weighting and query expansion models along with default values of their hyper-parameters ([http://terrier.org/docs/v4.0/configure\\_retrieval.html](http://terrier.org/docs/v4.0/configure_retrieval.html) accessed 23 May 2022). The optimal values of components hyper-parameters are obtained in an empiric way and different methods have been used for that, grid search is one of the most popular [6]. Optimal hyper-parameter values have been shown to be collection dependant” should be removed.
6. In the next paragraph, (1) the second sentence “For example, considering the query expansion component, Ayter et al. reported that for the query 442 from TREC7 reference collection, it is best to add two to five query terms to the initial query when



**Citation:** Mothe, J. Correction: Mothe, J. Analytics Methods to Understand Information Retrieval Effectiveness—A Survey. *Mathematics* 2022, 10, 2135. *Mathematics* 2022, 10, 3397. <https://doi.org/10.3390/math10183397>

Received: 25 July 2022

Accepted: 30 July 2022

Published: 19 September 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

- expanding it, while for query 432, it is best too add twenty terms [7]” should be “For example, considering the query expansion component, Ayter et al. reported that for the query 442 from the TREC7 reference collection, it is best to add 2 to 5 query terms to the initial query when expanding it while for query 432, it is best to add 20 terms [7]; (2) in the last sentence “The relationship between search components and system performance is thus worth studying in order to better understand the impact of the choice one makes when designing a query processing chain”, remove “thus”.
7. Remove the next repeated paragraph “Some studies have shown that the optimal hyper-parameters are not only collection dependent but query dependent as well. For example, considering the query expansion component, Ayter et al. reported that for the query (Evaluation forums distinguish between topics and queries. In evaluation frameworks, a topic consists of a title which is generally used as the query to submit to the search engine, a description of one or two sentences that explains the title, and a narrative that depicts what a relevant document is and what information will not be relevant. Because queries are usually formed from topic title, topic and query are often used interchangeably) 442 from TREC7 reference collection (<http://trec.nist.gov>, accessed 23 May 2022; TREC, Text REtrieval Conference provides collections for information retrieval evaluation. A collection consists of topics, documents and relevance assessments.), it is best to add two to five query terms to the initial query when expanding it, while for query 432, it is best too add twenty terms [7]. The relationship between search components and system performance is thus worth studying in order to better understand the impact of the choice one makes when designing a query processing chain”.
  8. In the following paragraph, (1) “With regard to better understanding of IR thanks to the system features”, should be changed to “With regard to a better understanding of IR thanks to the system features”; (2) “The strength of such evaluation campaigns is that a series of systems, each with different query processing chains, are using the same queries, the same document sets to search in and the same evaluation measures. Because the results are obtained on the same bases, they are comparable and analyses are made possible” should be changed to “The strength of such evaluation campaigns is that a series of systems, each with different query processing chains, are using the same queries, the same document sets to search in, and the same evaluation measures. Because the results are obtained on the same basis, they are comparable and analyses are made possible”; (3) In the last two sentences of the same paragraph, “These studies suffer, however, from the lack of structured and easily automatically exploitable descriptions on the query processing chains the participants used. The components used in the indexing and query processing as well as the values of their hyper-parameters are described in papers and working notes both in verbose ways and with different levels of detail” should be changed to “These studies suffer, however, from the lack of structured and easily exploitable descriptions of the query processing chains that the participants used. The components used in the indexing and query processing and the values of their hyper-parameters are described in papers and working notes both in verbose ways and with different levels of detail”.
  9. In the next paragraph, (1) “Here, the data collections from shared tasks is also used, but rather than considering real participants’ systems, a huge number of systems, from a few hundred to several thousands, are generated thanks to some tools [18,19]” should be changed to “Here, the data collections from shared tasks are also used, but rather than considering real participants’ systems, a huge number of systems, from a few hundred to several thousands, are generated thanks to some tools [18,19]”; (2) “Such chains have the advantage of being deterministic in the sense the components they used are fully described and known, and thus allowed deeper analyses. The effect of individual components or hyper-parameters can be studied” should be corrected to “Such chains have the advantage of being deterministic in the sense that the

- components they used are fully described and known, and, thus, allowed deeper analyses. The effect of individual components or hyper-parameters can be studied”.
10. Please remove the space before “?” in the sentence “Can we understand better the IR system effectiveness, that is to say successes and failures of systems, using data analytics methods?”.
  11. Please remove the space before “?” in the sentence “Did data driven analysis, based on thorough examination of IR components and hyper-parameters, lead to different or better conclusions?”.
  12. Please remove the space before “?” in the sentence “Did we learn from query performance prediction?”.
  13. In the next sentence, “The more long-term challenge is then”, “then” should be removed.
  14. In the second paragraph from the end on page 3, “It does not cover query performance prediction strictly speaking either”, remove “strictly speaking”.
  15. In the last paragraph of Introduction, (1) the sentence “Section 4 reports on the analysis conducted on result participants obtained at evaluation campaigns” should be changed to “Section 4 reports on the results of analyses conducted on participants obtained at evaluation campaigns”; (2) “Section 7 discusses the reported work also in terms of its potential impact for IR and concludes this paper” should be changed to “Section 7 discusses the reported work in terms of its potential impact for IR and concludes this paper”.
  16. In the first paragraph of Section 2, “Related work mainly consists in surveys that study a particular IR component. Other related studies are on relevance in IR, query difficulty and query performance prediction and fairness and transparency in IR” should be changed to “Related work mainly consists of surveys that study a particular IR component. Other related studies are of relevance in IR, query difficulty and query performance prediction, and fairness and transparency in IR”.
  17. In the first paragraph in Section 2.1, (1) add a comma before the last “and” in the sentence “More precisely, the criteria they used are as follows: the data source used in the expansion (e.g., Wordnet, top ranked documents, . . . ), candidate feature extraction method, feature selection method, and the expanded query representation”; (2) “With regard to effectiveness, they report h1Mean Average Precision on TREC collections (sparse results)” should be corrected to “With regard to effectiveness, they report mean average precision on TREC collections (sparse results)”; (3) “It is the area under the precision-recall curve which, in practice is replaced with an approximate based on precision at every position in the ranked sequence of documents” should be changed to “It is the area under the precision–recall curve which, in practice, is replaced with an approximate based on precision at every position in the ranked sequence of documents”; (4) “The authors concluded that for query expansion, linguistics techniques are considered as less effective than statistical-based methods. Specially, local analysis seems to perform better than corpus based. The authors also mentioned that the methods seem to have a high degree of complementary that should be exploited more. Their final conclusion is that the best choice depends on many factors among which the type of collection being queried, the availability and features of the external data, and the type of queries” should be “The authors concluded that for query expansion, linguistics techniques are considered as less effective than statistic-based methods. In particular, local analysis seems to perform better than corpus based. The authors also mentioned that the methods seem to be complementary and that this should be exploited more. Their final conclusion is that the best choice depends on many factors among which the type of collection being queried, the availability and features of the external data, and the type of queries”.
  18. In the second paragraph in Section 2.1, “They considered mainly rule-based stemmers and classified the stemmers according to their features, such as their strength, the aggressiveness with which the stemmer clears the terminations of the terms, the number of rules and suffixes considered, their use of recoding phase, partial-matching

- and constraint rules. They also compared the algorithms according to their conflation rate or Index Compression Factor” should be changed to “They considered mainly rule-based stemmers and classified the stemmers according to their features, such as their strength, the aggressiveness with which the stemmer clears the terminations of the terms, the number of rules and suffixes considered, their use of recoding phase, partial-matching, and constraint rules. They also compared the algorithms according to their conflation rate or index compression factor”.
19. In the third paragraph in Section 2.1, “We can also mention Kamphuis et al.’ study in which they considered 8 variants of the BM25 scoring function [26]. The authors considered 3 TREC collections and, based on average precision and precision at 30 documents” should be changed to “We can also mention the study by Kamphuis et al. in which they considered 8 variants of the **BM25 scoring function** [26]. The authors considered 3 TREC collections and used average precision in 30 documents”.
  20. The second paragraph in Section 2.2 should be changed to “Mizzaro [27] studied different kinds of relevance in IR, for which he defined several dimensions. He concluded that common practice to evaluate IR is to consider: (a) the surrogate, a representation of a document; (b) the query, the way the user expresses their perceived information need; and (c) the topic, which refers to the subject area the user is interested in. He also mentioned that this is the lowest level of relevance consideration in that it does not consider the real user’s information need nor the perceived information need, nor the information the user creates or receives when reading a document. Ruthven [28] studied how various types of TREC data can be used to better understand relevance and found that factors, such as familiarity, interest, and strictness of relevance criteria, may affect the TREC relevance assessments”.
  21. In the second sentence of Section 2.3, “made” should be changed to “performed”.
  22. In the caption of Figure 2, the first sentence should be changed to “A common practice in IR literature is to analyse the effect of hyper-parameters on the overall system effectiveness and to present the results under the form of tables or graphs”; Add comma before “and” in the third sentence to “Here, a deep learning-based model was used and comparisons are reported on the different training types, encoders, and batch sizes; using different effectiveness measures (nDCG@10, MRR@10, and R@1K), on different collections (here TREC DL’19, TREC-DL’20, and MSMARCO DEV)”; The next two sentences should be “The best results are highlighted in bold font. The bottom part is a typical graph to compare different variants’ or hyper-parameters’ effect on effectiveness. Here, the lines represent different combinations of hyper-parameters”.
  23. In Section 3, the first sentence should be changed to “The analysis of effectiveness for better comprehensive understanding of IR relies on data analysis methods and analysable data that we describe in this section”.
  24. Add a comma in the first paragraph of Section 3.1 to “System effectiveness analyses rely on different statistical analysis methods, including, but not limited to, machine learning”.
  25. Change the second paragraph in Section 3.1 to “**Boxplot** is a graphical representation of a series of numerical values that shows their locality, spread, and skewness based on their quartiles. Whiskers extend the Q1–Q3 box, indicating variability outside the upper and lower quartiles. Beyond the whiskers, outliers that differ significantly from the rest of the dataset are plotted as individual points. Effectiveness under different conditions (different queries, different values of a component parameter) is a typical series that can be represented under the form of a boxplot”.
  26. In the third paragraph in Section 3.1, (1) there should be a dot in the third sentence, and it should be “The most familiar measure of correlation is the Pearson product-moment correlation coefficient, which is a normalised form of the covariance. Covariance between two random variables measures their join distance to their expected values which”; (2) “Kendall correlation measures the correlation on ranks, that is the similarity of the orderings of the data when ranked by each of the variable values” should

- be “Kendall correlation measures the correlation on ranks, that is the similarity of the orderings of data when ranked by each of the variable values”.
27. In the sixth paragraph in Section 3.1, “In agglomerative clustering, to start with, each individual corresponds to a cluster, at each processing step, the two closest clusters are merged; the process ends when there is a single cluster. Ward criterion when used to choose the pair of clusters to merge is the minimum value of the error sum of squares [35]” should be changed to “In agglomerative clustering, each individual corresponds to a cluster; at each processing step, the two closest clusters are merged; the process ends when there is a single cluster. The minimum value of the error sum of squares is used as the ward criterion to choose the pair of clusters to merge [35]”.
  28. In the third paragraph to the end in Section 3.1, the sentence “It is used for example in query performance prediction” should be changed to “It is used, for example, in query performance prediction”.
  29. In the last paragraph in Section 3.1, the first two sentences should be “In this study, we do not consider **deep learning** methods as means to analyse and understand information retrieval effectiveness. Deep learning is more and more popular in IR but still these models lack interpretability”.
  30. In Section 3.2, “In this paper, the studied papers focused on TREC challenge which is pioneer. TREC considered many different languages although when it began and nowadays, it is mainly focused on English” should be changed to “In this paper, the studied papers focused on the pioneering TREC challenge. TREC considered many different languages, but when it began and nowadays, it is mainly focused on English”.
  31. In the next paragraph, add a comma before “namely” to “System performance analyses (presented in Sections 4–6) share the same type of data structures, namely matrices”.
  32. In the third paragraph, add a comma after “In general”.
  33. In the caption of Figure 3, add “the” before “3D matrices” at the beginning.
  34. In the first paragraph in Section 4, “hat opened the opportunity to mine these results” should be changed to “This opened the opportunity to mine these results”.
  35. In the fourth paragraph in Section 4, “At that time, it has not been possible to understand the reasons of the variability in results; it was stated as depending on three factors: the query, the relationship between the query and the documents, and the system parameters and was considered as a difficult problem to understand [14]. This difficulty remains” should be changed to “At that time, it was not possible to understand the reasons of the variability in results; it was stated as being dependent on three factors: the query, the relationship between the query and the documents, and the system parameters; it was considered as a difficult problem to understand [14] and this difficulty remains”.
  36. In the seventh paragraph in Section 4, “This analysis showed that the two measures correlates which means that a system A that is better than a system B on one of the two measures is also better when considering the second measure. Effective systems are effective, the measure that is used does not matter” should be changed to “This analysis showed that the two measures correlate, which means that a system A that is better than a system B on one of the two measures is also better when considering the second measure. When effective systems are effective, the measure that is used does not matter”.
  37. In the eighth paragraph, change “query” to “queries” in “Both some easy queries, for which the median effectiveness is high, and hard query”.
  38. In the caption of Figure 8, “and very similar ones (e.g., topic 285)—Web track 2014—Topics are ordered accoring to decearsing err@20 of the best system” should be changed to “and very similar ones (e.g., topic 285)—Web track 2014—topics are ordered according to decreasing the err@20 of the best system”.
  39. The caption of Figure 9 should be corrected to “System failure and effectiveness depend on queries—not all systems succeed or fail on the same queries. The visualisation

- shows the two first principal components of a Principal Component Analysis, where the data of the system effectiveness is obtained for each topic by each participants' run. MAP measure of TREC 12 Robust Track participants' runs. Figure reprinted with permission from [10], Copyright 2007, John Wiley and Sons".
40. In the caption of Figure 10, "brown triangles is for query cluster 2 and green crosses for cluster 3" should be changed to "brown triangles are for query cluster 2, and green crosses for cluster 3".
  41. In the first paragraph of Section 5, "The system factor is the one that has been highlighted the first in challenges: systems do not perform identically. Thanks to participants' results to challenges" should be changed to "The system factor is the factor that has been mentioned first in shared tasks: systems do not perform identically. Thanks to the results the participants' system obtained in shared tasks".
  42. In the second paragraph of Section 5, "Google scholar was used to find what were the first pieces of work related to automatically generated IR chains in the objective to analyse the component effects. Although the two cited work did not obtain many citations, they mark the starting point of this new research track" should be changed to "Google Scholar was used to find what were the first pieces of work related to automatically generated IR chains with the objective of analysing the component effects. Although the two cited works did not obtain many citations, they mark the starting point of this new research track".
  43. In the third paragraph of Section 5, in "The analyses based on synthetic data is in line with the idea developed in", "is" should be changed to "are".
  44. The fifth paragraph of Section 5 should be "Compared to using participants' systems, generated query processing chains give the ability to know the exact components and hyper-parameters used and thus make deeper analysis possible".
  45. In the sixth paragraph of Section 5, (1) "One of the pioneer studies that analysed a huge number of automatically generated systems is Ayter et al. [7]" should be "One of the pioneer studies that analysed a huge number of automatically generated systems is Ayter et al. [7]"; (2) "into" should be changed to "in" in the second sentence; (3) "Among their findings, the authors concluded the choice of the stemmer component had little if not influence while the weighting model had an impact on the results (see Figure 11). Other findings were that diricheletlm is the weaker search model among the 7 studied while BB2 is among the best; this when considering also all the other parameters. Their analyses also confirmed that systems behave differently and that the choice of the components at each phase of the retrieving process as well as the component hyper-parameters are an important part of system successes and failures" should be "Among their findings, the authors concluded that the choice of the stemmer component had little to no influence while the weighting model had an impact on the results (see Figure 11). Other findings were that dirichletLM is the weaker search model among the seven studied while BB2 is among the best; this when also considering all the other parameters. Their analyses also confirmed that systems behave differently and that the choice of the components at each phase of the retrieving process and the component hyper-parameters are an important part of system successes and failures".
  46. In the caption of Figure 11, the first sentence should be "The choice of the weighting model has more impact than the stemmer used".
  47. In the seventh paragraph of Section 5, "They show that variants of the stop word list used during indexing has not a huge impact" should be changed to "They show that variants of the stop word list used during indexing does not have a huge impact".
  48. In the eighth paragraph of Section 5, "diricheletlm" should be changed to "dirichletLM".
  49. In the fourth and the fifth paragraphs in Section 6.1, (1) add a colon symbol after "but" to "A single query feature cannot explain a single system effectiveness, but:"; (2) the word "Reversely" should be changed to "Inversely".

50. In the third paragraph from the end in Section 6.1, change “non effective” to “non-effective”.
51. In the caption of Figure 14, “4 data sets” should be changed to “four datasets”.
52. The first paragraph of Section 6.2 should be changed to “In Section 5, we reported studies on system variabilities and the effect of the components used at each search phase on the results when averaged over query sets, not considering deeply the query effect”.
53. In the fourth paragraph of Section 6.2, (1) the first sentence should be changed to “In Figure 17a, where the easy queries only are analysed, we can see that the most influential parameter is the query expansion model used because this is the one where the tree first split, here for the value  $c$ ”; (2) the last two sentences should be “The main overall conclusion is that the influential parameters are not the same for easy and hard queries; giving the intuition that obtaining the best performance cannot be by applying the same process whatever the queries are. This was further analysed in [53], where more TREC collections were studied with the same conclusions”.
54. In the second paragraph of Section 7, (1) change “non effective” to “non-effective”; (2) Add “a” before “similar profile” in the last sentence to “Some systems have a similar profile, they fail/success on the same queries”.
55. In the paragraph “Regarding C4” in Section 7, “dirichletlm” should be changed to “dirichletLM”.
56. In the caption of Figure 18, the semicolon behind “some obtained 0.8” should be changed to a comma.
57. In the fifth paragraph from the end of Section 7, “Up to now, the accuracy of features or feature combinations has not demonstrate they can explain system effectiveness” should be changed to “Up to now, the accuracy of features or feature combinations has not demonstrated that they can explain system effectiveness”.
58. In the fourth paragraph from the end of Section 7, (1) the first sentence should be changed to “Although we do not yet understand well the factors of system effectiveness, the studies show that not a single system, while effective in average on a query set, is able to answer all the queries well”; (2) “SQE has not been proven to be very effective, certainly due both to the limited number of configurations used at that time (2 different query processing chains)” should be changed to “SQE has not been proven to be very effective, certainly due to both the limited number of configurations used at that time (two different query processing chains)”; (3) “This very large number of combinations however makes it difficult to use in real world systems” should be changed to “However, this very large number of combinations makes it difficult to use in real world systems”; (3) remove “in” after “among” in the sentence “Mothe and Ullah [91] present. . .”
59. In the second paragraph from the end of Section 7, (1) “I am tempt to answer” should be changed to “I am tempted to answer”; (2) “I am convinced that data analytics methods can further been investigated to analyse the amount of data that has been generated by the community, both in shared tasks and in labs while tuning systems” should be changed to “I am convinced that data analytics methods can further be investigated to analyse the amount of data that have been generated by the community, both in shared tasks and in labs while tuning systems”.

The authors apologize for any inconvenience caused and state that the scientific conclusions are unaffected. The original article has been updated.

## Reference

1. Mothe, J. Analytics Methods to Understand Information Retrieval Effectiveness—A Survey. *Mathematics* **2022**, *10*, 2135. [[CrossRef](#)]