

A Comparative Study of Bo1 and RM3 Query Expansion using DPH Retrieval

Albert Gesk¹, Lilian Erler¹

¹University Kassel, Germany

Abstract

Query expansion via pseudo-relevance feedback is a well-established technique to improve retrieval effectiveness in information retrieval systems. Among classical approaches, Bo1 and RM3 are frequently used but exhibit varying effectiveness depending on the retrieval setting. In this work, we empirically compare Bo1 and RM3 query expansion within a controlled Retriever-Rewriter-Retriever pipeline using the DPH retrieval model. Experiments are conducted on the *radboud-validation-20251114-training* dataset using nDCG@10 as evaluation metric. Statistical significance testing shows that neither a significant difference nor a significant improvement in retrieval effectiveness can be observed between Bo1 and RM3 under identical parameter settings.

Keywords

Information Retrieval, Query Expansion, Bo1, RM3, PyTerrier

1. Introduction

Query expansion using pseudo-relevance feedback (PRF) is a classical approach to mitigate vocabulary mismatch in information retrieval (IR) systems. The mechanism of pseudo-relevance feedback (PRF) involves extracting expansion terms from the initial top-ranked documents to formulate a new query for a second retrieval stage [1]. Among commonly used PRF techniques, Bo1 and RM3 are widely adopted in traditional IR systems and toolkits.

Despite their popularity, the relative effectiveness of Bo1 and RM3 is known to be sensitive to retrieval models, parameter settings, and datasets. This motivates a controlled comparison of both methods under identical experimental conditions.

In this paper, we investigate the following research question: *Does the choice of Bo1 versus RM3 query expansion lead to significant differences or improvements in retrieval effectiveness when applied within a DPH-based retrieval pipeline?*

Our contributions are: (i) a controlled experimental comparison of Bo1 and RM3 query expansion using identical retrieval and feedback settings, and (ii) a statistical analysis of their impact on nDCG@10 on the *radboud-validation-20251114-training* dataset.

2. Related Work

Pseudo-relevance feedback (PRF) is a well-established technique in information retrieval to improve query effectiveness by automatically expanding the original query with terms extracted from top-ranked documents [1].

Previous studies have shown that RM3 provides robust improvements over baseline retrieval models like BM25, particularly when the number of feedback documents and expansion terms are carefully tuned. However, the quality of the expansion terms can be limited because of its reliance on pseudo-relevant documents. This happens especially in scenarios where the top-ranked documents are not truly relevant. [2]

IR Lab Course, Winter Semester 2025

*Corresponding author.

[†]These authors contributed equally.

 albert.gesk@gmail.com (A. Gesk); uk076601@student.uni-kassel.de (L. Erler)

 <https://github.com/AlbertGesk> (A. Gesk); <https://github.com/lilianerler> (L. Erler)



These findings motivate a controlled comparison between RM3 and Bo1, as they represent alternative pseudo-relevance feedback methods based on distinct term selection principles. Bo1 is based on divergence-from-randomness (DRF) and selects terms that are statistically informative in the top-retrieved documents [3], while RM3 is a probabilistic relevance model [2].

Other works have explored hybrid or transformer-based retrieval pipelines that incorporate PRF or developed a demanding pipelines, such as ColBERT-PRF, which integrates semantic pseudo-relevance feedback into dense retrieval models [4]. But controlled comparisons of classical Bo1 and RM3 within identical DPH-based pipelines remain limited.

This motivates our work to conduct a systematic and controlled evaluation of Bo1 and RM3 under identical experimental conditions, assessing not only retrieval effectiveness but also statistical significance of the observed differences.

3. Methodology

This section depicts the experimental setup designed to answer the research question.

3.1. Dataset and Preprocessing

For the experiments, we utilized the *radboud-validation-20251114-training* dataset. The indexing of the documents was based on the standard textual representation (`default_text`) provided in the dataset. No additional preprocessing steps or filtering were applied, meaning only the standard tokenization performed by PyTerrier was used.

3.2. Experimental Design

The experiment use a retriever-rewriter-retriever pipeline as foundation. The DPH retrieval model [5] implemented via PyTerrier is setup with a thousand maximum number of results to return per query and with no metadata fields to return for each search result. The pseudo-relevance feedback models, both the Bo1 [6] and RM3 [7] are implemented via PyTerrier as well. For both models, all parameters are remained at default, the number of feedback terms to use is set to ten and the number of feedback documents to use is set to three. The interpolation weight between the original query and the feedback model for RM3 is set to 0.6.

To evaluate the effectiveness of the retrieval pipelines, differences in per-topic nDCG@10 scores were assessed for statistical significance using a paired Student's t-test. The significance level was set to $\alpha = 0.05$, and Bonferroni correction was applied to account for multiple comparisons.

3.3. Hypotheses and Null-Hypotheses

Given the research question the following two hypotheses and corresponding null-hypotheses are formulated:

- H_1 Given query expansion via pseudo-relevance feedback using the top 3 retrieved documents and 10 expansion terms, there is a statistically significant difference in nDCG@10 on *radboud-validation-20251114-training* between (i) the DPH-based retrieval pipeline described in subsection 3.2 three with Bo1 query expansion and (ii) the same pipeline with RM3 query expansion. ($\alpha=0.05$)
- H_1^0 Given query expansion via pseudo-relevance feedback using the top 3 retrieved documents and 10 expansion terms, there is no statistically significant difference in nDCG@10 on *radboud-validation-20251114-training* between (i) the DPH-based retrieval pipeline described in subsection 3.2 three with Bo1 query expansion and (ii) the same pipeline with RM3 query expansion. ($\alpha=0.05$)

Table 1

Retrieval effectiveness and statistical significance (nDCG@10).

Method	nDCG@10	p-value
DPH-Bo1-DPH	0.4947	0.283
DPH-RM3-DPH	0.4744	

- H_2 Given query expansion via pseudo-relevance feedback using the top 3 retrieved documents and 10 expansion terms, there is a statistically significant improvement in nDCG@10 on radboud-validation-20251114-training between (i) the DPH-based retrieval pipeline described in subsection 3.2 three with Bo1 query expansion and (ii) the same pipeline with RM3 query expansion. ($\alpha=0.05$)
- H_2^0 Given query expansion via pseudo-relevance feedback using the top 3 retrieved documents and 10 expansion terms, there is no statistically significant improvement in nDCG@10 on radboud-validation-20251114-training between (i) the DPH-based retrieval pipeline described in subsection 3.2 three with Bo1 query expansion and (ii) the same pipeline with RM3 query expansion. ($\alpha=0.05$)

4. Results

Table 1 presents the retrieval effectiveness and statistical testing results. The DPH-Bo1-DPH pipeline achieves a higher mean nDCG@10 score than DPH-RM3-DPH. Though, Bo1 performs better than RM3. The p-value of 0.283 is significantly larger than 0.05. This means the observed difference is likely due to random variation across topics. We failed to reject our H_1^0 and H_2^0 , which means we can't prove our H_1 and H_2 .

5. Conclusion

Given the parameters of 3 retrieved documents and 10 expansion terms for Bo1 and RM3, both pseudo-relevance feedback-based transformer models behave similarly. Experimenting with more feedback documents and more expansion terms in the future could be meaningful.

6. Reference

7. Template parameters

There are a number of template parameters which modify some part of the `ceurart` document class. These parameters are enclosed in square brackets and are a part of the `\documentclass` command:

```
\documentclass[parameter]{ceurart}
```

Frequently-used parameters, or combinations of parameters, include:

- `twocolumn` : Two column layout. This option is not supported by CEUR-WS, hence only use it for papers not to be submitted to CEUR-WS.
- `hf` : Enable header and footer¹. This option shall also not be used for papers to be published by CEUR-WS.

¹You can enable the display of page numbers in the final version of the entire collection. In this case, you should adhere to the end-to-end pagination of individual papers.

8. Front matter

8.1. Title Information

The titles of papers should be either all use the emphasizing capitalized style or they should all use the regular English (or native language) style. It does not make a good impression if you or your authors mix the styles.

Use the `\title` command to define the title of your work. Do not insert line breaks in your title.

8.2. Title variants

`\title` command have the below options:

- `title`: Document title. This is default option.

```
\title [mode=title] {This is a title}
```

You can just omit it, like as follows:

```
\title {This is a title}
```

- `alt`: Alternate title.

```
\title [mode=alt] {This is a alternate title}
```

- `sub`: Sub title.

```
\title [mode=sub] {This is a sub title}
```

You can just use `\subtitle` command, as follows:

```
\subtitle {This is a sub title}
```

- `trans`: Translated title.

```
\title [mode=trans] {This is a translated title}
```

- `transsub`: Translated sub title.

```
\title [mode=transsub] {This is a translated sub title}
```

8.3. Authors and Affiliations

Each author must be defined separately for accurate metadata identification. Multiple authors may share one affiliation. Authors' names should not be abbreviated; use full first names wherever possible. Include authors' e-mail addresses whenever possible.

`\author` command have the below options:

- `style` : Style of author name (chinese)
- `prefix` : Prefix
- `suffix` : Suffix
- `degree` : Degree
- `role` : Role
- `orcid` : ORCID
- `email` : E-mail
- `url` : URL

Author names can have some kinds of marks and notes:

- affiliation mark: `\author[<num>]`.

The author names and affiliations could be formatted in two ways:

1. Group the authors per affiliation.
2. Use an explicit mark to indicate the affiliations.

Author block example:

```
\author[1,2]{Author Name}[%  
    prefix=Prof.,  
    degree=D. Sc.,  
    role=Researcher,  
    orcid=0000-0000-000-0000,  
    email=name@example.com,  
    url=https://name.example.com  
]  
  
\address[1]{Affiliation #1}  
\address[2]{Affiliation #2}
```

8.4. Abstract and Keywords

Abstract shall be entered in an environment that starts with `\begin{abstract}` and ends with `\end{abstract}`.

```
\begin{abstract}  
    This is an abstract.  
\end{abstract}
```

The key words are enclosed in a `keywords` environment. Use `\sep` to separate keywords.

```
\begin{keywords}  
    First keyword \sep  
    Second keyword \sep  
    Third keyword \sep  
    Fourth keyword  
\end{keywords}
```

At the end of front matter add `\maketitle` command.

8.5. Various Marks in the Front Matter

The front matter becomes complicated due to various kinds of notes and marks to the title and author names. Marks in the title will be denoted by a star (*) mark; footnotes are denoted by super scripted Arabic numerals, corresponding author by an Conformal asterisk (*) mark.

8.5.1. Title marks

Title mark can be entered by the command, `\tnotemark[<num>]` and the corresponding text can be entered with the command `\tnotetext[<num>]{<text>}`. An example will be:

```
\title{A better way to format your document for CEUR-WS}  
  
\tnotemark[1]  
\tnotetext[1]{You can use this document as the template for  
    preparing your  
    publication. We recommend using the latest version of the ceurart  
    style.}
```

Table 2
Frequency of Special Characters

Non-English or Math	Frequency	Comments
\emptyset	1 in 1,000	For Swedish names
π	1 in 5	Common in math
\$	4 in 5	Used in business
Ψ_1^2	1 in 40,000	Unexplained usage

`\tnotemark` and `\tnotetext` can be anywhere in the front matter, but should be before `\maketitle` command.

8.5.2. Author marks

Author names can have some kinds of marks and notes:

- footnote mark : `\fnmark[<num>]`
- footnote text : `\fntext[<num>]{<text>}`
- corresponding author mark : `\cormark[<num>]`
- corresponding author text : `\cortext[<num>]{<text>}`

8.5.3. Other marks

At times, authors want footnotes which leave no marks in the author names. The note text shall be listed as part of the front matter notes. Class files provides `\nonumnote` for this purpose. The usage

`\nonumnote{<text>}`

and should be entered anywhere before the `\maketitle` command for this to take effect.

9. Sectioning Commands

Your work should use standard L^AT_EX sectioning commands: `\section`, `\subsection`, `\subsubsection`, and `\paragraph`. They should be numbered; do not remove the numbering from the commands.

Simulating a sectioning command by setting the first word or words of a paragraph in boldface or italicized text is not allowed.

10. Tables

The “ceurart” document class includes the “booktabs” package – <https://ctan.org/pkg/booktabs> – for preparing high-quality tables.

Table captions are placed *above* the table.

Because tables cannot be split across pages, the best placement for them is typically the top of the page nearest their initial cite. To ensure this proper “floating” placement of tables, use the environment `table` to enclose the table’s contents and the table caption. The contents of the table itself must go in the `tabular` environment, to be aligned properly in rows and columns, with the desired horizontal and vertical rules.

Immediately following this sentence is the point at which Table 2 is included in the input file; compare the placement of the table here with the table in the printed output of this document.

To set a wider table, which takes up the whole width of the page’s live area, use the environment `table*` to enclose the table’s contents and the table caption. As with a single-column table, this wide table will “float” to a location deemed more desirable. Immediately following this sentence is the point

Table 3
Some Typical Commands

Command	A Number	Comments
\author	100	Author
\table	300	For tables
\table*	400	For wider tables

at which Table 3 is included in the input file; again, it is instructive to compare the placement of the table here with the table in the printed output of this document.

11. Math Equations

You may want to display math equations in three distinct styles: inline, numbered or non-numbered display. Each of the three are discussed in the next sections.

11.1. Inline (In-text) Equations

A formula that appears in the running text is called an inline or in-text formula. It is produced by the `math` environment, which can be invoked with the usual `\begin{...}\end{...}` construction or with the short form `$... $`. You can use any of the symbols and structures, from α to ω , available in L^AT_EX [?]; this section will simply show a few examples of in-text equations in context. Notice how this equation: $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$, set here in in-line math style, looks slightly different when set in display style. (See next section).

11.2. Display Equations

A numbered display equation—one set off by vertical space from the text and centered horizontally—is produced by the `equation` environment. An unnumbered display equation is produced by the `displaymath` environment.

Again, in either environment, you can use any of the symbols and structures available in L^AT_EX; this section will just give a couple of examples of display equations in context. First, consider the equation, shown as an inline equation above:

$$\lim_{n \rightarrow \infty} \frac{1}{n} = 0. \quad (1)$$

Notice how it is formatted somewhat differently in the `displaymath` environment. Now, we'll enter an unnumbered equation:

$$S_n = \sum_{i=1}^n x_i,$$

and follow it with another numbered equation:

$$\lim_{x \rightarrow 0} (1 + x)^{1/x} = e \quad (2)$$

just to demonstrate L^AT_EX's able handling of numbering.

12. Figures

The “figure” environment should be used for figures. One or more images can be placed within a figure. If your figure contains third-party material, you must clearly identify it as such, as shown in the example below.

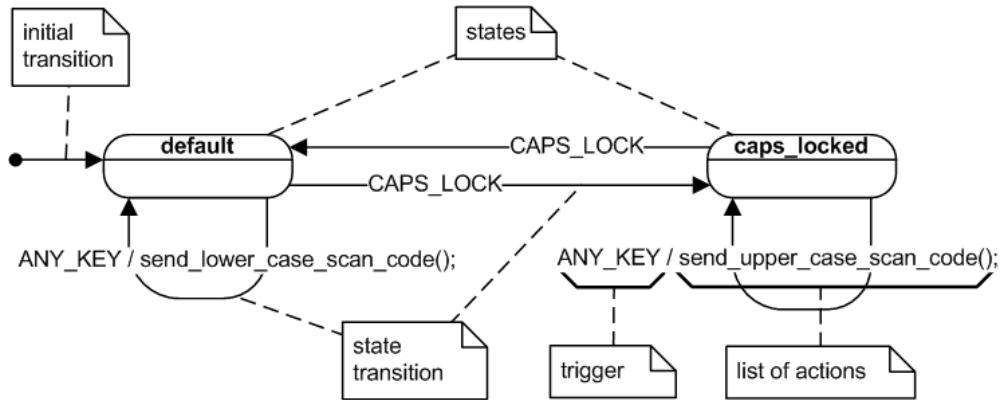


Figure 1: Sample UML state machine (source: https://en.wikipedia.org/wiki/UML_state_machine, license CC BY-SA 3.0).

Your figures should contain a caption which describes the figure to the reader. Figure captions go below the figure. Your figures should also include a description suitable for screen readers, to assist the visually-challenged to better understand your work.

Figure captions are placed below the figure.

13. Citations and Bibliographies

The use of BibTeX for the preparation and formatting of one's references is strongly recommended. Authors' names should be complete – use full first names ("Donald E. Knuth") not initials ("D. E. Knuth") – and the salient identifying features of a reference should be included: title, year, volume, number, pages, article DOI, etc.

The bibliography is included in your source document with these two commands, placed just before the `\end{document}` command:

```
\bibliography{bibfile}
```

where "bibfile" is the name, without the ".bib" suffix, of the BibTeX file.

13.1. Some examples

A paginated journal article [?], an enumerated journal article [?], a reference to an entire issue [?], a monograph (whole book) [?], a monograph/whole book in a series (see 2a in spec. document) [?], a divisible-book such as an anthology or compilation [?] followed by the same example, however we only output the series if the volume number is given [?] (so series should not be present since it has no vol. no.), a chapter in a divisible book [?], a chapter in a divisible book in a series [?], a multi-volume work as book [?], an article in a proceedings (of a conference, symposium, workshop for example) (paginated proceedings article) [?], a proceedings article with all possible elements [?], an example of an enumerated proceedings article [?], an informally published work [?], a doctoral dissertation [?], a master's thesis: [?], an online document / world wide web resource [? ? ?], a video game (Case 1) [?] and (Case 2) [?] and [?] and (Case 3) a patent [?], work accepted for publication [?], prolific author [?] and [?]. Other cites might contain 'duplicate' DOI and URLs (some SIAM articles) [?]. Multi-volume works as books [?] and [?]. A couple of citations with DOIs: [? ?]. Online citations: [? ? ? ?].

14. Acknowledgments

Identification of funding sources and other support, and thanks to individuals and groups that assisted in the research and the preparation of the work should be included in an acknowledgment section, which is placed just before the reference section in your document.

This section has a special environment:

```
\begin{acknowledgments}
    These are different acknowledgments.
\end{acknowledgments}
```

so that the information contained therein can be more easily collected during the article metadata extraction phase, and to ensure consistency in the spelling of the section heading.

Authors should not prepare this section as a numbered or unnumbered \section; please use the “acknowledgments” environment.

15. Appendices

If your work needs an appendix, add it before the “\end{document}” command at the conclusion of your source document.

Start the appendix with the “\appendix” command:

```
\appendix
```

and note that in the appendix, sections are lettered, not numbered.

Acknowledgments

Thanks to the developers of ACM consolidated LaTeX styles <https://github.com/borisveysman/acmart> and to the developers of Elsevier updated L^AT_EX templates <https://www.ctan.org/tex-archive/macros/latex/contrib/els-cas-templates>.

Declaration on Generative AI

Either:

The author(s) have not employed any Generative AI tools.

Or (by using the activity taxonomy in ceur-ws.org/genai-tax.html):

During the preparation of this work, the author(s) used X-GPT-4 and Gramby in order to: Grammar and spelling check. Further, the author(s) used X-AI-IMG for figures 3 and 4 in order to: Generate images. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication’s content.

References

- [1] W. Junmei, A knowledge-based approach for pseudo-relevance feedback by exploiting semantic relevance, Ph.D. thesis, 2025. URL: <https://doi.org/10.1007/s10115-025-02581-5>.
- [2] N. Sinhababu, R. Khatun, Leq: Large language models generate expanded queries for searching, in: 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2024, pp. 1–4. doi:[10.1109/ICCCNT61001.2024.10725314](https://doi.org/10.1109/ICCCNT61001.2024.10725314).
- [3] L. de Swart, Performance comparison of different query expansion and pseudo-relevance feedback methods (2024).
- [4] X. Wang, C. MacDonald, N. Tonellootto, I. Ounis, Colbert-prf: Semantic pseudo-relevance feedback for dense passage and document retrieval, ACM Trans. Web 17 (2023). URL: <https://doi.org/10.1145/3572405>. doi:[10.1145/3572405](https://doi.org/10.1145/3572405).
- [5] G. Amati, Frequentist and bayesian approach to information retrieval, in: M. Lalmas, A. MacFarlane, S. M. Rüger, A. Tombros, T. Tsikrika, A. Yavlinsky (Eds.), Advances in Information Retrieval, 28th European Conference on IR Research, ECIR 2006, London, UK, April 10-12, 2006, Proceedings,

- volume 3936 of *Lecture Notes in Computer Science*, Springer, 2006, pp. 13–24. URL: https://doi.org/10.1007/11735106_3. doi:10.1007/11735106_3.
- [6] G. Amati, Probability models for information retrieval based on divergence from randomness, Ph.D. thesis, University of Glasgow, UK, 2003. URL: <http://theses.gla.ac.uk/1570/>.
- [7] N. A. Jaleel, J. Allan, W. B. Croft, F. Diaz, L. S. Larkey, X. Li, M. D. Smucker, C. Wade, Umass at TREC 2004: Novelty and HARD, in: E. M. Voorhees, L. P. Buckland (Eds.), Proceedings of the Thirteenth Text REtrieval Conference, TREC 2004, Gaithersburg, Maryland, USA, November 16-19, 2004, volume 500-261 of *NIST Special Publication*, National Institute of Standards and Technology (NIST), 2004. URL: <http://trec.nist.gov/pubs/trec13/papers/umass.novelty.hard.pdf>.

A. Online Resources

The sources for the ceur-art style are available via

- GitHub,
- Overleaf template.