

=1000

**RECUPERAÇÃO DE DOCUMENTOS COM
CONSULTAS IMPRECISAS APLICADA A
DESCRIÇÃO DE FILMES**

RITA REZENDE BORGES DE LIMA

RECUPERAÇÃO DE DOCUMENTOS COM
CONSULTAS IMPRECISAS APLICADA A
DESCRIÇÃO DE FILMES

Projeto Orientado em Computação, elaborado como requisito para conclusão do curso de Ciência da Computação da Universidade Federal de Minas Gerais

ORIENTADOR: RODRYGO LUIS TEODORO SANTOS

Belo Horizonte
Novembro de 2023

© 2023, Rita Rezende Borges de Lima.
Todos os direitos reservados.

Rezende Borges de Lima, Rita

Recuperação de Documentos com Consultas Imprecisas
Aplicada a Descrição de Filmes / Rita Rezende Borges de Lima.
— Belo Horizonte, 2023
vii, 14 f. ; 29cm

Projeto Orientado a Computação — Universidade Federal de
Minas Gerais
Orientador: Rodrygo Luis Teodoro Santos

I. Título.

Para Sonaly, Bernardo, Helena e Luiz.

Resumo

A tarefa proposta, encontrar um determinado filme, aqui é contextualizada para uma tarefa de ranqueamento: dada uma consulta em linguagem natural com descrição sucinta e um modelo capaz de atribuir a cada filme um grau de similaridade para aquela descrição geramos um *ranking* o qual um dos filmes do topo, preferencialmente o primeiro, seja o filme descrito na consulta. No trabalho anterior, reproduzimos diferentes modelos para a tarefa de *ranking* descritas em Bhargav et al. [2022]. Neste estudo, introduzimos heurísticas específicas aplicadas ao conjunto de dados da tarefa além de aplicarmos técnicas de *reranking* Lin et al. [2021] com o objetivo de aprimorar os resultados prévios.

Palavras-chave: *Ranking*, Recuperação de Informação, Aprendizado Profundo, Processamento de Linguagem Natural, Recuperação de Filmes.

Lista de Tabelas

3.1	Acurácia de Predição para Consultas no Conjunto de Validação	7
3.2	Proporção de Consultas no Conjunto de Validação cuja Característica foi Encontrada	7
3.3	Resultados no Conjunto de Validação	10
3.4	Resultados no conjunto de Teste	11

Sumário

Resumo	v
Lista de Tabelas	vi
1 Introdução	1
1.1 Contextualização	1
1.2 Definição da tarefa	2
1.3 Motivação	2
1.4 Objetivo	2
2 Fundamentação e trabalhos relacionados	3
2.1 Conceitualização	3
2.2 Aprendizado profundo	3
2.3 Reranking	4
2.4 Modelos Densos Aplicados	4
2.5 Trabalhos relacionados	5
3 Contribuição	6
3.1 Conjunto de dados	6
3.1.1 Extração de Características dos Dados	7
3.1.2 Heurísticas Propostas	8
3.1.3 A tarefa de <i>Reranking</i>	8
3.2 Resultados	10
4 Conclusão	12
Referências Bibliográficas	13

Capítulo 1

Introdução

1.1 Contextualização

Formas de entretenimento, tal como músicas, livros, filmes, entre outros, são consumidas pelo indivíduo comum diariamente. A medida que o tempo passa, menos lembranças destas mídias são guardadas, ocasionando no sentimento de algo estar na **ponta da língua**. Aqui, definimos um item estar na ponta da língua, *Tip of the Tongue (ToT)* no inglês, quando esse já foi consumido por um usuário que não se recorda de atributos discriminativos o suficiente para recuperar esse item diretamente, como, por exemplo, lembrar do nome de um filme ou do compositor de uma música.

Como exemplo, considere um filme assistido há muito tempo, no qual o usuário só se lembra de certas cenas. O usuário poderia descrever o filme da seguinte maneira:

The ending of the film took place in a rooftop in New York. Out of nowhere a giant monster made of marshmallow appeared.

O usuário não se recorda do nome do filme para poder acessá-lo, ainda que se lembre de vários detalhes desse. Logo, podemos afirmar que o clássico de 1984 “Caça Fantasmas” estava na ponta de sua língua.

1.2 Definição da tarefa

A tarefa proposta, encontrar um determinado filme, aqui é contextualizada para uma tarefa de ranqueamento. A entrada será constituída de uma consulta em linguagem natural em inglês com descrição sucinta e imprecisa, e um conjunto de filmes contendo seus principais atributos, tal como título, enredo, ano de publicação, entre outros. Dado um modelo capaz de atribuir a cada filme um grau de similaridade para aquela consulta geramos um ranking o qual um dos filmes do topo, preferencialmente o primeiro, seja o filme descrito na consulta.

1.3 Motivação

Encontrar filmes com descrições vagas e imprecisas é algo de interesse comum em usuários de internet como pode ser visto pela presença de múltiplos websites destinados a esta mesma finalidade tal como o site <https://whatismymovie.com/> e o fórum do site [reddit.com, r/tipofmytongue](https://www.reddit.com/r/tipofmytongue). Uma motivação para além do interesse da população em geral para o estudo e pesquisa na área de busca com consultas imprecisas é o desenvolvimento de ferramentas que possam auxiliar pessoas portadoras de doenças que acarretam em perda de memória, tal como o Alzheimer. Em particular esse trabalho foi inspirado em um dos desafios da *The Text REtrieval Conference* ou *TREC* de 2023.

1.4 Objetivo

Este trabalho é uma continuação do meu projeto orientado a computação anterior. Um de seus propósitos é explorar o conjunto de dados coletado por Bhargav et al. [2022] e analisar os resultados dos modelos implementados previamente, a fim de propor heurísticas para aprimorar o desempenho no referido conjunto de dados. Além disso, busca-se a aplicação de técnicas clássicas e mais robustas da área, como o *reranking* Lin et al. [2021], utilizando modelos densos como monoBERT Nogueira et al. [2019], monoT5 Pradeep et al. [2021] e duoT5 Pradeep et al. [2021].

Capítulo 2

Fundamentação e trabalhos relacionados

2.1 Conceitualização

De acordo com Baeza-Yates & Ribeiro-Neto [2011], a modelagem de uma tarefa de recuperação de informação é um processo complexo com intuito de produzir uma função de ranking, ou seja, uma função que atribui notas para documentos referentes a uma consulta. Esse processo consiste em duas etapas:

- A concepção de um framework para representar documentos e consultas.
- A definição de uma função para computar o ranking.

2.2 Aprendizado profundo

Aprendizado profundo, em inglês *Deep Learning*, refere-se a uma abordagem de aprendizado de máquina em que as representações dos dados são aprendidas automaticamente, eliminando a necessidade de extrair características manualmente. De acordo com LeCun et al. [2015], essa técnica tem se mostrado o estado-da-arte em várias áreas, como por exemplo processamento de linguagem natural.

As estratégias clássicas de ranqueamento, como o uso de TF-IDF, têm uma limitação ao considerar apenas a ocorrência exata dos termos nas consultas. Além disso, as abordagens tradicionais de aprendizado também apresentam limitações LeCun et al. [2015], uma vez que utilizam conjuntos fixos de características pré-determinadas para

descrever os dados. No entanto, uma estratégia de aprendizado profundo levaria em consideração o contexto de cada termo, o que auxiliaria na discriminação de documentos.

Desta forma, a utilização do aprendizado profundo oferece a possibilidade de explorar relações mais complexas e sutis entre os dados, permitindo uma compreensão mais aprofundada das informações e aprimorando o processo de ranqueamento. Ao considerar o contexto dos termos nas consultas, essa abordagem pode identificar características relevantes que distinguem diferentes enredos de filmes, proporcionando resultados de ranqueamento mais precisos.

2.3 Reranking

O *reranking* é uma técnica comumente empregada em sistemas de recuperação de informações para melhorar a precisão dos resultados gerados por modelos de *ranking* inicial. Após a geração da lista inicial de documentos ranqueados, o *reranking* envolve a reordenação dessa lista com base em critérios adicionais, como características específicas do usuário, contexto temporal ou outras heurísticas relevantes. Essa abordagem visa otimizar a relevância dos documentos mais relevantes, proporcionando uma resposta mais precisa às consultas dos usuários.

2.4 Modelos Densos Aplicados

MonoBERT é um modelo baseado na arquitetura *BERT* (Bidirectional Encoder Representations from Transformers). Desenvolvido por Nogueira et al. [2019], o *MonoBERT* se destaca por sua capacidade de realizar ranqueamento de documentos em estágios múltiplos. Ele utiliza a representação contextual bidirecional para entender o contexto das palavras em uma consulta, permitindo uma compreensão mais profunda da semântica. Essa abordagem de múltiplos estágios torna o *MonoBERT* eficiente em tarefas de recuperação de informações.

MonoT5 é uma variação do modelo T5 (Text-to-Text Transfer Transformer) adaptada para tarefas de ranqueamento. Desenvolvido por Pradeep Pradeep et al. [2021], o *MonoT5* enfatiza a simplicidade e a eficácia em tarefas específicas, incluindo *reranking*. Ao estruturar a tarefa de ranqueamento como uma tarefa de geração de texto, o

MonoT5 utiliza o contexto global da consulta e dos documentos para gerar uma representação textual que é então usada para a classificação.

DuoT5, também proposto por Pradeep Pradeep et al. [2021], é uma extensão do *MonoT5* projetada para operar em condições tanto monolínguas quanto multilínguas. Ele demonstra uma maior flexibilidade ao lidar com consultas e documentos em diferentes idiomas, tornando-se uma escolha valiosa em ambientes onde a diversidade linguística é uma consideração crucial.

2.5 Trabalhos relacionados

Arguello et al. [2021] discutem uma maneira de processar consultas ToT a respeito de filmes obtidas do site *I remember this movie*¹. Estas são divididas em sentenças que por sua vez são marcadas com códigos a respeito de seu conteúdo e forma, como exemplo a presença de comparações, descrições ou até incerteza. Um corpus é montado com títulos e enredos de filmes retirados da coleção de filmes *WikiPlot5* e esses são ranqueados para cada consulta por meio do algoritmo Okapi BM25, de Robertson et al. [1994].

Bhargav et al. [2022] os autores coletam e montam um novo conjunto de pares <consulta, item> para filmes e livros. Esses dados são compostos de perguntas do subreddit *Tip of my Tongue*² dentro do período de 2017 e 2020 com respostas marcadas como corretas pelo autor da pergunta. Nesse conjunto de dados são aplicadas diversas estratégias de ranking como a função BM25, *baseline* clássico para a tarefa, além de outras técnicas como o PL2, um modelo 2-Poisson Amati & Van Rijsbergen [2002] e *Dense Passage Retrieval* Karpukhin et al. [2020] .

¹<https://irememberthismovie.com/>.

²<https://www.reddit.com/r/tipofmytongue/>.

Capítulo 3

Contribuição

3.1 Conjunto de dados

O conjunto de dados utilizado neste estudo foi extraído por Bhargav et al. [2022] do fórum "Tip of My Tongue" do site *Reddit*, conforme descrito no artigo de referência. Cada *post* no fórum contém um título e uma descrição em inglês que se referem a um filme misterioso. Essas descrições geralmente abordam características do enredo, personagens e cenas relacionadas ao filme em questão. Já as respostas dos *posts* consistem em possíveis títulos para o filme descrito. É o autor do *post* quem marca se a resposta está correta ou não. Com essas respostas formamos pares de consulta e resposta, onde a consulta corresponde à descrição *tip of the tongue* do filme e o título do filme resposta. Estes dados também foram processados e disponibilizados pela *TREC*.

Informações da *Wikipedia* e do *IMDB* também foram utilizadas para complementar os dados resposta com características do filme, tais como enredo, data de publicação, diretor, gênero do filme, se é um filme para adultos, língua original, região entre outras. Além dos pares consulta/resposta, os autores também disponibilizaram um conjunto de dados com relações um-para-muitos de consulta para "hard negatives" (títulos que não correspondem ao filme em questão), a fim de fornecer exemplos de filmes que possuem características semelhantes à consulta, mas não são a resposta correta para auxiliar no treinamento.

3.1.1 Extração de Características dos Dados

Neste trabalho, empreguei a *API* da *LANGCHAIN*, que incorpora um modelo *GPT*, para realizar previsões de diversas características relacionadas a filmes para as consultas em linguagem natural. Fui capaz de classificar para cada consulta informações cruciais, como o gênero do filme, a indicação se o conteúdo é destinado a um público adulto, além de dados sobre a linguagem predominante e a região associada ao filme. A acurácia destas predições pode ser observadas na tabela 3.1.

Características	Acurácia
Gênero	0.8
Se o filme é adulto	0.9
Linguagem	0.6
Região	0.3

Tabela 3.1. Acurácia de Predição para Consultas no Conjunto de Validação

Adicionalmente, utilizei técnicas de expressões regulares, *REGEX*, para extrair informações temporais específicas de cada consulta, incluindo datas e anos relevantes. Essa abordagem permitiu uma compreensão mais profunda do contexto temporal das consultas, enriquecendo a análise e classificação dos filmes. Além disso, por meio da biblioteca *NLTK* extraí nomes próprios que foram validados com a *API* do *IMDb*. Isso possibilitou a verificação precisa de nomes de atores e diretores mencionados nas consultas. Essa combinação de métodos, envolvendo tanto a *API* da *LANGCHAIN* quanto técnicas de *REGEX* e a *API* do *IMDb*, proporcionou uma abordagem abrangente e precisa na extração e validação de dados relevantes para a tarefa de ranqueamento de filmes.

Características	% de consultas com detecção
Datas	0.65
Nome de Pessoas	0.05

Tabela 3.2. Proporção de Consultas no Conjunto de Validação cuja Característica foi Encontrada

3.1.2 Heurísticas Propostas

Uma das heurísticas adotadas no projeto consiste em limitar o tamanho das entradas, priorizando seções relevantes do corpus que tratam do enredo dos filmes. Foram selecionadas seções contendo palavras-chave específicas, como *plot*, *synopsis* e *description*, no corpus da Wikipedia. Essas seções foram concatenadas com o início do texto, oferecendo uma representação compacta e informativa do conteúdo relacionado ao enredo dos filmes, principalmente em comparação a forma anterior de processamento do corpus que se dava pelo truncamento arbitrário.

As consultas fornecidas pela *TREC* são segmentadas em sentenças, cada uma delas com anotações sobre seu conteúdo. Para refinar o processo, as sentenças marcadas com a anotação "social" foram filtradas, uma vez que geralmente contêm apenas comunicação do autor com o leitor, sendo consideradas menos relevantes para o escopo do projeto. Ambas abordagens visam aprimorar a eficácia do processamento, focando em informações mais substanciais para a tarefa em questão.

3.1.3 A tarefa de *Reranking*

Neste trabalho também foi utilizada a técnica de *reranking* para aprimorar a precisão dos resultados obtidos inicialmente com o *BM25* e as runs fornecidas pela *TREC*, que incluíam modelos como *DistilBERT* e *GPT-4*. O *reranking* foi executado utilizando os modelos *MonoBERT*, *MonoT5* e *DuoT5*, individualmente e em combinações. Além disso, incorporei as heurísticas e filtros baseadas nas características extraídas anteriormente das consultas, como gênero, classificação etária, linguagem e região. A combinação dessas técnicas proporcionou uma abordagem integrada e robusta para refinar o *ranking* inicial, incorporando tanto modelos tradicionais quanto novas heurísticas. Todos os experimentos realizados para este trabalho podem ser vistos abaixo:

- **ufmgBMmBQ:** um *reranking* com o modelo *monoBERT* de um ranking inicial feito com *BM25* com o corpus original truncado e consultas processadas.
- **ufmgBMmBQD:** um *reranking* com o modelo *monoBERT* de um ranking inicial feito com *BM25* com as consultas e o corpus processados.
- **ufmgBMmBQbF:** um *reranking* com o modelo *monoBERT* de um ranking inicial feito com *BM25* com o corpus original truncado e consultas processadas. O ranking inicial passa por um processo de filtragem comparando as características extraídas e previstas com informações do *IMDb*.

- **ufmgBMmBQDbF:** um *reranking* com o modelo *monoBERT* de um ranking inicial feito com *BM25* com as consultas e o corpus processados. O ranking inicial passa por um processo de filtragem comparando as características extraídas e previstas com informações do *IMDb*.
- **ufmgBMmBQaF:** um *reranking* com o modelo *monoBERT* de um ranking inicial feito com *BM25* com o corpus original truncado e consultas processadas. O ranking final passa por um processo de filtragem comparando as características extraídas e previstas com informações do *IMDb*.
- **ufmgBMmBQDaF:** um *reranking* com o modelo *monoBERT* de um ranking inicial feito com *BM25* com as consultas e o corpus processados. O ranking final passa por um processo de filtragem comparando as características extraídas e previstas com informações do *IMDb*.
- **ufmgDBmBQ:** um *reranking* com o modelo *monoBERT* de um ranking inicial feito com *DistilBERT* disponibilizado pela *TREC*. O *reranking* foi feito com o corpus original truncado e consultas processadas.
- **ufmgDBmBQD:** um *reranking* com o modelo *monoBERT* de um ranking inicial feito com *DistilBERT* disponibilizado pela *TREC*. O *reranking* foi feito com as consultas e o corpus processados.
- **ufmgDBdTQD:** um *reranking* com o modelo *duoT5* de um ranking inicial feito com *DistilBERT* disponibilizado pela *TREC*. O *reranking* foi feito com as consultas e o corpus processados.
- **ufmgDBmBdTQD:** um *reranking* com o modelo *monoBERT* seguido de outro *reranking* com o modelo *duoT5* de um ranking inicial feito com *DistilBERT* disponibilizado pela *TREC*. O *reranking* foi feito com as consultas e o corpus processados.
- **ufmgG4mBQD:** um *reranking* com o modelo *monoBERT* de um ranking inicial feito com *GPT-4* disponibilizado pela *TREC*. O *reranking* foi feito com as consultas e o corpus processados.
- **ufmgG4dTQD:** um *reranking* com o modelo *duoT5* de um ranking inicial feito com *GPT-4* disponibilizado pela *TREC*. O *reranking* foi feito com as consultas e o corpus processados.

3.2 Resultados

As métricas utilizadas na trilha da TREC são MRR (Mean Reciprocal Rank), sucesso e NDCG (Normalized Discounted Cumulative Gain). Essas métricas específicas para sistemas de recuperação de informações aqui também foram escolhidas como indicadores principais para avaliar o desempenho dos experimentos.

Os resultados dos experimentos conduzidos são apresentados nas Tabelas 3.3 e 3.4. A Tabela 3.3 detalha os resultados obtidos para o conjunto de validação, fornecendo métricas como MRR, sucesso e NDCG para todos os experimentos descritos na Seção 3.1.3. Já a Tabela 3.4 destaca os experimentos destinados ao conjunto de teste, enviados para a trilha da TREC. Os resultados desses experimentos são computados e fornecidos pela trilha.

Run	Submitted	nDCG	Success	MRR
BM25		0.1314	0.4067	0.0881
ufmgBMmBQ		0.1148	0.2533	0.0860
ufmgBMmBQD		0.1297	0.3139	0.0872
ufmgBMmBQbF		0.1203	0.3033	0.0892
ufmgBMmBQDbF		0.1187	0.3033	0.0873
ufmgBMmBQaF		0.1312	0.2533	0.0954
ufmgBMmBQDaF		0.1373	0.3139	0.0903
Distilbert		0.1627	0.6600	0.0743
ufmgDBmBQ	✓	0.1984	0.4133	0.1450
ufmgDBmBQD	✓	0.2045	0.4267	0.1499
ufmgDBmBdTQD	✓	0.1412	0.3990	0.7870
ufmgDBdTQD		0.1812	0.4089	0.1424
GPT-4		0.2407	0.3200	0.2180
ufmgG4mBQD	✓	0.1979	0.3200	0.1638
ufmgG4mTQD		0.1780	0.3200	0.1484
ufmgG4dTQD	✓	0.1872	0.3200	0.1511

Tabela 3.3. Resultados no Conjunto de Validação

A implementação das heurísticas de processamento de consulta e documento demonstrou impacto positivo nos resultados obtidos no conjunto de validação. Ao refinar a representação das consultas e documentos, as heurísticas contribuíram para uma melhor compreensão do contexto e das intenções dos usuários, resultando em melhorias significativas na precisão do ranking. Essa abordagem efetiva realça a importância de considerar características específicas dos dados na etapa de pré-processamento para

Run	Submitted	nDCG	Success	MRR
ufmgDBmBQ	✓	0.2090	0.3933	0.1636
ufmgDBmBQD	✓	0.1998	0.4067	0.1507
ufmgDBmBdTQD	✓	0.1108	0.4067	0.0505
ufmgG4mBQD	✓	0.2404	0.3733	0.2002
ufmgG4dTQD	✓	0.1668	0.3733	0.1189

Tabela 3.4. Resultados no conjunto de Teste

otimizar o desempenho do sistema.

Quanto às filtragens aplicadas com base em características antes do *reranking*, observou-se que inicialmente essas filtragens não produzem diferenças notáveis nos resultados para o conjunto de validação. No entanto, ao realizar a filtragem após a aplicação do *reranking*, houve uma melhoria substancial no ranking final, especialmente em métricas como *NDCG* e *MRR*. Esse fenômeno sugere que a filtragem inicial pode não ter um impacto direto na eficiência do *reranking*, em parte por este não ser perfeito, mas consegue qualificar mais um *ranking* final.

É interessante destacar que ao realizar o *reranking* com os modelos densos *MonoBERT* e *DuoT5*, houve melhorias significativas no ranking inicial gerado pelo modelo denso *DistilBERT* no conjunto de validação. No entanto, esse comportamento não foi replicado ao tentar aplicar *reranking* a um *ranking* inicial gerado pelo modelo *GPT-4*. Essa observação ressalta nuances na interação entre modelos densos e a eficácia do *reranking* em diferentes contextos, apontando para a necessidade de considerar as características específicas dos modelos ao desenvolver estratégias de otimização.

Capítulo 4

Conclusão

Na realização da tarefa de ranqueamento para descrições imprecisas de filmes, empreguei técnicas clássicas da literatura, como o *reranking* com modelos densos como *MonoBERT*, *MonoT5* e *DuoT5*. Além disso, propusemos preprocessamentos simples e extrações e predições de características cruciais das consultas, visando utilizá-las para aprimorar a eficiência do *ranking*.

Ao longo do projeto, explorei diversas bibliotecas, *frameworks* e técnicas de ranqueamento. Essa experiência proporcionou a aplicação prática dos conhecimentos adquiridos durante a graduação, especialmente nas áreas de recuperação de informação, aprendizado profundo e processamento de linguagem natural. A implementação de *language models* representou uma oportunidade valiosa para consolidar e expandir meu entendimento em tópicos essenciais dessas disciplinas.

Referências Bibliográficas

- Amati, G. & Van Rijsbergen, C. J. (2002). Probabilistic models of information retrieval based on measuring the divergence from randomness. *ACM Trans. Inf. Syst.*, 20(4):357–389.
- Arguello, J.; Ferguson, A.; Fine, E.; Mitra, B.; Zamani, H. & Diaz, F. (2021). Tip of the tongue known-item retrieval: A case study in movie identification. Em *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, pp. 5–14.
- Baeza-Yates, R. & Ribeiro-Neto, B. (2011). *Modern Information Retrieval: The Concepts and Technology behind Search*. Addison-Wesley Publishing Company, USA, 2nd edição.
- Bhargav, S.; Sidiropoulos, G. & Kanoulas, E. (2022). 'it's on the tip of my tongue': A new dataset for known-item retrieval. Em *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, WSDM '22, p. 48–56, New York, NY, USA. Association for Computing Machinery.
- Karpukhin, V.; Oguz, B.; Min, S.; Lewis, P.; Wu, L.; Edunov, S.; Chen, D. & Yih, W.-t. (2020). Dense passage retrieval for open-domain question answering. Em *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6769–6781, Online. Association for Computational Linguistics.
- LeCun, Y.; Bengio, Y. & Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.
- Lin, J.; Nogueira, R. & Yates, A. (2021). Pretrained transformers for text ranking: Bert and beyond.
- Nogueira, R.; Yang, W.; Cho, K. & Lin, J. (2019). Multi-stage document ranking with bert.
- Pradeep, R.; Nogueira, R. & Lin, J. (2021). The expando-mono-duo design pattern for text ranking with pretrained sequence-to-sequence models.

-
- Robertson, S. E.; Walker, S.; Jones, S.; Hancock-Beaulieu, M. & Gatford, M. (1994).
Okapi at trec-3. Em *Text Retrieval Conference*.