

UNIVERSIDADE FEDERAL DO PARANÁ

MATEUS RAMBO STREY

RANQUEAMENTO DE RECURSOS EDUCACIONAIS EM UMA REDE
SOCIAL

CURITIBA PR

2016

MATEUS RAMBO STREY

RANQUEAMENTO DE RECURSOS EDUCACIONAIS EM UMA REDE
SOCIAL

Trabalho apresentado como requisito parcial à conclusão do Curso de Bacharelado em Ciência da Computação, setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Ciência da Computação*.

Orientador: Marcos Didonet Del Fabro.

CURITIBA PR

2016

*Dedico este trabalho à minha Mãe e
a família, que são a base da minha
vida.*

Agradecimentos

Agradeço à Robertha, por me acompanhar e suportar durante a graduação, e especialmente por me ajudar a colocar as ideias no papel. Daniel e Nanci, pelo suporte que sempre tive de vocês. Iago, Rafael e Cleyton, pelo apoio emocional nos momentos de necessidade. Maurício, por todas as discussões construtivas e ideias que tivemos nesse período da graduação. Walter, Willian, Luiz, Neto, Igor, Alejandro, Jones e tantos outros, por tornarem os meus dias mais humanos no Politécnico. Demais amigos e colegas que me ajudaram nessa jornada. Aos professores do C3SL, por acreditarem no meu potencial e me permitirem explorar soluções para problemas sociais através da computação.

Resumo

À partir da elaboração de um site de Rede Social voltado a Recursos Educacionais (RE), manifesta-se a necessidade de criar um mecanismo de busca que mostre ao usuário resultados relevantes. Por haver pouquíssimas fontes que abordassem a problemática, empiricamente descobriu-se que utilizar o classificador TF-IDF aliado ao Ranqueamento desenvolvido nesse trabalho, poderia satisfazer as necessidades presentes do portal. No desenvolvimento do trabalho são apresentados três sistemas de pesos para realizar o Ranqueamento e três formas de unir o ranqueamento ao TF-IDF. Como o portal ainda não foi inaugurado, foi ainda concebido um "Populador", que simula as interações dos usuários no site, permitindo comparar o efeito dos pesos na busca.

Após a execução das etapas acima comentadas, obteve-se algumas tabelas que mostravam como a posições dos resultados se alteravam conforme os diferentes pesos aplicados. Com estas, foi possível comparar a efetividade de cada mescla. Como resultado, concluiu-se que por mais que algumas mesclas e pesos pareçam mais indicados que outros, a validação do método só poderá ser testada, de fato, quando o site receber usuários de verdade.

Palavras-chave: classificação de documentos, recursos educacionais abertos, redes sociais.

Abstract

From the elaboration of a Social Network Site aimed at Educational Resources (ER), it is necessary to create a search mechanism that shows relevant results to the users. Because there are not much sources that addressed the problem, it was found empirically that using a TF-IDF classifier, coupled with the ranking system developed in this work, could satisfy the present needs of the portal. In the development of the work, three weight systems are presented to perform the ranking and three ways of linking the ranking with TF-IDF. As the portal has not yet been launched, a "Populator" was also conceived, which simulates the interactions of the users in the site, allowing to compare the effect of the weights in the search.

After performing the mentioned steps above, we obtained some tables that showed how the results positions changed according to the different weights applied. With these, it was possible to compare the effectiveness of each blend. As a result, it has been found that although some blends and weights appear to be better than others, method validation can only be tested when the site actually receives users. **Keywords:** Document classification, Open educational

resources, social networks.

Sumário

1	Introdução	1
2	Estado da Arte	3
2.1	Redes sociais para Recursos Educacionais	3
2.2	Recuperação de Informação e Relevância	5
2.3	Simulação de Redes Sociais	7
2.4	Abordagens	8
2.4.1	SNDocRank	8
2.4.2	MATF	8
3	Ranqueamento de Recursos Educacionais	9
3.1	PortalMEC	9
3.2	Ranqueamento	10
3.3	Busca via Elasticsearch	13
3.4	Populador	15
3.5	Implementação	18
3.6	Experimentos	19
3.6.1	Com recursos importados	19
3.6.2	Com recursos gerados	20
3.6.3	Resultados	20
4	Conclusões	22
	Referências Bibliográficas	24
A	Experimentos Realizados	27

Lista de Figuras

3.1	Modelo Entidade Relacionamento simplificado do PortalMEC	10
3.2	PortalMEC - Capa	11
3.3	PortalMEC - Busca de Recursos	11
3.4	PortalMEC - Recurso Educacional	12
3.5	Fluxo do cálculo do <i>score</i> no PortalMEC	18

Lista de Tabelas

2.1	Query, Documentos Encontrados e seus TF	6
2.2	Query, Documentos Encontrados e seus TF*IDF	6
2.3	Query, Documentos Encontrados e suas Relevâncias	6
3.1	Pesos do Ranqueamento de Recursos	14
3.2	Pesos da Reputação dos Usuários	14
3.3	Categoria dos Usuários	14
3.4	Mesclas possíveis de score/rank	15
3.5	Perfis de Usuário e suas Ações no Populador	18
A.1	Recursos importados do PortalMEC - 120 usuários - <i>query</i> 'dna' - top 15 resultados de 2826	28
A.2	Recursos importados - 120 usuários - <i>query</i> 'dna' - Comparativos	28
A.3	Recursos importados do PortalMEC - 120 usuários - <i>query</i> 'molecula' - top 15 resultados de 424	29
A.4	Recursos importados - 120 usuários - <i>query</i> 'molecula' - Comparativos	29
A.5	Recursos importados do PortalMEC - 500 usuários - <i>query</i> 'dna' - top 15 resultados de 2826	30
A.6	Recursos importados - 500 usuários - <i>query</i> 'dna' - Comparativos	30
A.7	Recursos importados do PortalMEC - 500 usuários - <i>query</i> 'molecula' - top 15 resultados de 424	31
A.8	Recursos importados - 500 usuários - <i>query</i> 'molecula' - Comparativos	31
A.9	Recursos importados do PortalMEC - 1000 usuários - <i>query</i> 'dna' - top 15 resultados de 2826	32
A.10	Recursos importados - 1000 usuários - <i>query</i> 'dna' - Comparativos	32
A.11	Recursos importados do PortalMEC - 1000 usuários - <i>query</i> 'molecula' - top 15 resultados de 424	33
A.12	Recursos importados - 1000 usuários - <i>query</i> 'molecula' - Comparativos	33
A.13	Recursos importados do PortalMEC - 5000 usuários - <i>query</i> 'dna' - top 15 resultados de 2826	34
A.14	Recursos importados - 5000 usuários - <i>query</i> 'dna' - Comparativos	34
A.15	Recursos importados do PortalMEC - 5000 usuários - <i>query</i> 'molecula' - top 15 resultados de 424	35
A.16	Recursos importados - 5000 usuários - <i>query</i> 'molecula' - Comparativos	35
A.17	Recursos importados do PortalMEC - 10000 usuários - <i>query</i> 'dna' - top 15 resultados de 2826	36
A.18	Recursos importados - 10000 usuários - <i>query</i> 'dna' - Comparativos	36
A.19	Recursos importados do PortalMEC - 10000 usuários - <i>query</i> 'molecula' - top 15 resultados de 424	37

A.20 Recursos importados - 10000 usuários - <i>query</i> 'molecula' - Comparativos . . .	37
A.21 Recursos gerados pelo populador - 120 usuários - <i>query</i> 'sint' - top 15 resultados de 398	38
A.22 Recursos gerados - 120 usuários - <i>query</i> 'sint' - Comparativos	38
A.23 Recursos gerados pelo populador - 120 usuários - <i>query</i> 'velit' - top 15 resultados de 306	39
A.24 Recursos gerados - 120 usuários - <i>query</i> 'velit' - Comparativos	39
A.25 Recursos gerados pelo populador - 500 usuários - <i>query</i> 'sint' - top 15 resultados de 935	40
A.26 Recursos gerados - 500 usuários - <i>query</i> 'sint' - Comparativos	40
A.27 Recursos gerados pelo populador - 500 usuários - <i>query</i> 'velit' - top 15 resultados de 721	41
A.28 Recursos gerados - 500 usuários - <i>query</i> 'velit' - Comparativos	41
A.29 Recursos gerados pelo populador - 1000 usuários - <i>query</i> 'sint' - top 15 resultados de 1641	42
A.30 Recursos gerados - 1000 usuários - <i>query</i> 'sint' - Comparativos	42
A.31 Recursos gerados pelo populador - 1000 usuários - <i>query</i> 'velit' - top 15 resultados de 1179	43
A.32 Recursos gerados - 1000 usuários - <i>query</i> 'velit' - Comparativos	43
A.33 Recursos gerados pelo populador - 5000 usuários - <i>query</i> 'sint' - top 15 resultados de 8270	44
A.34 Recursos gerados - 5000 usuários - <i>query</i> 'sint' - Comparativos	44
A.35 Recursos gerados pelo populador - 5000 usuários - <i>query</i> 'velit' - top 15 resultados de 6050	45
A.36 Recursos gerados - 5000 usuários - <i>query</i> 'velit' - Comparativos	45
A.37 Recursos gerados pelo populador - 10000 usuários - <i>query</i> 'sint' - top 15 resultados de 15625	46
A.38 Recursos gerados - 10000 usuários - <i>query</i> 'sint' - Comparativos	46
A.39 Recursos gerados pelo populador - 10000 usuários - <i>query</i> 'velit' - top 15 resultados de 11642	47
A.40 Recursos gerados - 10000 usuários - <i>query</i> 'velit' - Comparativos	47

Lista de Acrônimos

C3SL	Centro de Computação Científica e Software Livre
CERN	European Organization for Nuclear Research
DINF	Departamento de Informática
ER	Erdős Rényi random graph model
IDF	Inverted Document Frequency
IR	Recuperação de Informação
LOC	Library of Congress
MAS	Multi-Level Actor Similarity
MATF	Multi Aspect TF
MCTI	Ministério da Ciência, Tecnologia, Inovações e Comunicações
MEC	Ministério da Educação
MIT	Massachusetts Institute of Technology
NLP	Natural Language Processing
RE	Recursos Educacionais
REA	Recursos Educacionais Abertos
RN	Regular networks
SATA	Serial AT Attachment
SF	Scale-free networks
SGDB	Sistema de Gerenciamento de Banco de Dados
SN	Spatial networks
SVM	Support Vector Machines
SW	Small world models
TF	Term Frequency
TF-IDF	Term Frequency/Inverted Document Frequency
UFPR	Universidade Federal do Paraná
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura

Capítulo 1

Introdução

Existem diversos portais voltados ao compartilhamento de Recursos Educacionais (RE), como será apresentado no capítulo 2. Estes portais não possuem integração, o que dificulta a recuperação dos recursos, e também não possuem, ou possuem de maneira simples, alguma métrica para a avaliação dos recursos pela comunidade. O presente trabalho surge à partir do desenvolvimento de um portal com elementos de Redes Sociais voltado ao compartilhamento de RE, tendo como público alvo não só professores, mas a população brasileira como um todo. Neste projeto do Centro de Computação Científica e Software Livre (C3SL), em parceria com o Ministério da Educação (MEC), percebeu-se a importância de desenvolver um portal que reúna os RE de diversas fontes em conjunto com um sistema de busca de RE que retorne resultados relevantes, em relação ao tipo de usuário da plataforma em conjunto com as informações geradas através das relações dos mesmos.

Inicialmente foram testadas buscas mais simples, puramente pelo sistema de gerenciamento de banco de dados, com resultados pouco relevantes ao que foi procurado. Após pesquisar as opções existentes em motores de busca e não encontrar trabalhos científicos sobre a temática (Busca em Redes Sociais para RE), encontrou-se abordagens para problemas similares, como a *SNDocRank*[Gou et al., 2010] e a *Multi Aspect TF*[Paik, 2013], apresentadas no capítulo 2. Consta-se que estas abordagens não contemplam as particularidades do domínio tratado. Cria-se então um sistema de ranqueamento que mede a relevância do RE na Rede Social, sendo este usado para complementar os resultados retornados na busca. O ranqueamento utiliza as interações da Rede Social (critérios com pesos ajustáveis), como a quantidade de visualizações de um RE, para inferir a relevância dos RE. Os critérios que serão apresentados foram discutidos em reuniões do projeto, dado que não existem fontes científicas ou mesmo informais que tratassem do problema encontrado.

Contudo, o portal está em processo de finalização e, portanto, não pode ser utilizado por usuários reais. Assim, para testar o método proposto nesse trabalho, é sugerida uma ferramenta que simula o comportamento de usuários em uma Rede Social para RE. Utiliza-se um modelo *Scale-free network* modificado para as necessidades do portal. É importante ressaltar que não foram encontradas informações relevantes a respeito sobre o tema (Simulação de Redes Sociais para RE) e o autor do trabalho optou por transformar a experiência em um escrito científico para uso futuro, expondo uma solução empírica para o problema.

A oportunidade de construir um portal para RE, além de tudo, é uma forma de contribuir à sociedade, pois, como será apresentada, a maior parte desses documentos estão dispersos em muitos repositórios, dificultando o processo de busca. Um portal assim, portanto, pode conectar produtores e consumidores de RE em um só lugar, e, quem sabe, inspirar novas contribuições.

Este trabalho, então, focará no ranqueamento de RE, apresentando, num primeiro momento, o atual cenário de pesquisas de Redes Sociais para RE, *Information Retrieval* (para busca em documentos) e Relevância, Simulação de Redes Sociais, e abordagens relacionadas ao tema do trabalho. A seguir, apresenta-se o portal desenvolvido, a questão do Ranqueamento de RE em si, Busca com o motor de busca escolhido, Simulador de Rede Social para os experimentos, implementação do Ranqueamento e do Simulador de Rede Social, e experimentos para validar a solução proposta. Termina-se, por fim, com as conclusões extraídas do trabalho e possibilidades de trabalhos futuros.

Capítulo 2

Estado da Arte

Este capítulo apresentará o estado da arte de Redes Sociais para Recursos Educacionais, com enfoque em *Recuperação de Informação* e Relevância, Simulação de Redes Sociais, além de algumas abordagens de outros pesquisadores para estes problemas.

Recursos Educacionais (RE) são gerados por diversas fontes e podem ser distribuídos ou divulgados de maneira livre. Logo, é possível compilá-los em um único lugar no qual as pessoas consigam compartilhar, avaliar ou até agrupar esse tipo de material. É daí que surge a ideia de criar um Site de Rede Social que tenha como objetivo disponibilizar o maior número de RE permissível e que os usuários encontrem conteúdos confiáveis.

Para encontrar esses conteúdos no Site de Rede Social, faz-se uso da área de *Recuperação de Informação* para documentos, que permite encontrar os recursos mais relevantes para o que for buscado por meio de *queries*. Como é necessário uma validação do retorno das buscas por recursos, utiliza-se uma Simulação de Rede Social que represente as interações dos usuários nessa rede. Cabe ressaltar que foram encontrados pouquíssimos trabalhos científicos sobre a relevância de busca de documentos num site de Rede Social de RE, tendo sido necessário, muitas vezes durante o processo de elaboração, buscar soluções de forma empírica.

2.1 Redes sociais para Recursos Educacionais

Sites de Rede Social são serviços *web* que permitem que indivíduos construam um perfil público (ou semi-público) e consigam se conectar a outros indivíduos, podendo, ainda, ver a lista de quem está conectado à si. Essa conexão (natureza e nomenclatura) pode variar conforme o site [boyd e Ellison, 2007]. Essa categoria de Sites pode ser classificada em *networking-oriented* - compartilhamento de conteúdo entre amigos - ou *knowledge-sharing oriented* - compartilhamento de conteúdo entre usuários interessados no assunto [Guo et al., 2009]. Assim sendo, um Site de Rede Social voltado a Recursos Educacionais pode ser classificado como *knowledge-sharing oriented*, visto que sua principal função é compartilhar Recursos Educacionais entre usuários interessados.

Recursos Educacionais Abertos (REA) são materiais de ensino, aprendizagem e investigação - digitais ou não - em domínio público ou com uma licença que permita acesso, uso, adaptação e distribuição livre ou com poucas restrições [UNESCO, 2012]. Neste trabalho, considera-se Recurso Educacional (RE) como um REA que pode ou não ser adaptado, apesar de manter a permissão ao acesso, uso e distribuição livre ou com poucas restrições.

Existem diversos portais, com abordagens distintas, voltados ao compartilhamento de REA (e alguns, inclusive, permitem Recursos com licenças mais restritivas). Abaixo listam-se

exemplos de portais que trabalham com a temática, sendo que todos armazenam e disponibilizam recursos de vários tipos (imagem, áudio, vídeo e afins):

- **Portal Domínio Público**[MEC, 2004]: Biblioteca digital do Ministério da Educação (MEC) para conteúdos que entraram em Domínio Público, desenvolvida com *software* livre. Não possui elementos de Rede Social;
- **UFPR - REA/PEA**[UFPR, 2015]: Portal para REA e Práticas Educacionais Abertas (PEA) desenvolvidos na UFPR. Permite envio de Recursos que serão disponibilizados após aprovação dos responsáveis pelo portal. O elemento de Rede Social é o envio de recursos.
- **Portal Dia a Dia Educação**[Seed-PR, 2011]: Portal da Secretaria de Estado da Educação do Paraná (Seed-PR), permite consulta de Recursos Educacionais e contém *links* para várias bibliotecas e sites com Recursos. O elemento de Rede Social é inclusão da quantidade de *downloads* e visualizações dos Recursos Educacionais;
- **Portal do Professor**[MEC e MCTI, 2008b]: Portal que permite compartilhamento de Planos de Aula e recursos multimídia (REA ou não). É possível apenas criar planos de aula (de forma individual ou coletiva). Usuários podem avaliar os recursos através de um comentário e uma nota (1 à 5 estrelas);
- **Banco Internacional de Objetos Educacionais**[MEC e MCTI, 2008a]: Portal do Ministério da Educação (MEC) que reúne Recursos Educacionais e permite o envio de recursos de diversos formatos como áudio, vídeo, animação, simulação e software educacional, disponíveis mediante aprovação dos responsáveis. O elemento de Rede Social é que usuários podem avaliar os recursos com uma nota (1 à 5 estrelas);
- **OER Commons**[ISKME, 2007]: Portal para REA, em língua inglesa, que permite envio e criação (pelo portal) de recursos. O elementos de Rede Social são: permitir avaliação por categorias dos recursos e criação de *tags*, permite criação de grupos de trabalho para compartilhar materiais relevantes dentro de um contexto;
- **Curriki**[Curriki, 2006]: Portal para REA, em língua inglesa, que também permite envio e criação (pelo portal) de recursos. Os elementos de Rede Social são: permitir avaliação dos recursos (texto e 1 à 5 estrelas); permitir criação de grupos de trabalho e permitir ter contatos (conexões) dentro da rede.
- **MIT OpenCourseWare**[MIT, 2002]: Portal que reúne video aulas, em língua inglesa, de cursos de nível de graduação do *Massachusetts Institute of Technology* (MIT) e os disponibiliza para a comunidade. O elemento de Rede Social presente é uma lista dos cursos mais visitados na plataforma.
- **Biblioteca Digital Mundial**[LOC e UNESCO, 2009]: Biblioteca criada pela *U.S. Library of Congress* em parceria com a UNESCO que agrupa recursos digitais (livros, manuscritos, gravações e outros) de várias bibliotecas ao redor do mundo. Não possui elementos de Rede Social.
- **OpenEd**[ACT, 2009]: Rede Social para RE (alguns são pagos), em língua inglesa, que visa a interação dos professores com os alunos via *web*. Os elementos de Rede Social são: professores podem enviar conteúdos e tarefas para seus alunos e/ou turmas. Os professores podem escolher suas áreas de interesse e receber sugestões de recursos relacionados.

Percebe-se que existem diversos REA disponíveis às pessoas, porém estes estão alocados em muitos sites distintos (vide lista acima), fazendo com que o usuário tenha certo grau de dificuldade em procurá-los. Não bastasse ter de procurar em diversos locais, ainda é necessário fazer uma seleção do material disponível (avaliando sua qualidade e coerência com o conteúdo buscado). Nos sites acima que permitem essa seleção, a busca retornada só irá levar em conta a avaliação se o resultado utilizar esta como filtro. A problemática é que em todo esse processo bons recursos perdem visibilidade e muitas pessoas podem não encontrar o que procuram, mas que, de fato, existem. A exceção é o *OpenEd*, que utiliza informações de uso dos professores e alunos para melhorar as buscas e sugestões de conteúdo (entretanto, está disponível apenas em língua inglesa, dificultando o acesso no Brasil).

2.2 Recuperação de Informação e Relevância

Um sistema de *Recuperação de Informação* (IR) interroga, através de uma questão (*query*), uma lista de documentos indexados e retorna um subconjunto destes como resposta, de acordo com um relacionamento - relevância - dos documentos com a questão [Goffman, 1964].

Esses documentos precisam ser relevantes em relação à *query* utilizada. A Relevância é um dos principais conceitos da *Information Science* e é aceito pela maioria dos pesquisadores que a satisfação do usuário está conectada com a relevância dos resultados dos sistemas de IR. Apesar disto, seu significado e métodos são debatidos constantemente [Kagolovsky e Moehr, 2003]. A relevância pode ser calculada de várias formas (*Document Classification*) como modelos Booleanos [Lee et al., 1994], TF-IDF, Support Vector Machines (SVM) [Thet et al., 2007], Processamento de Linguagem Natural (NLP) [Blosseville et al., 1992] e Redes Neurais [Benediktsson et al., 1990].

A técnica de TF-IDF é uma forma de se fazer a análise estatística dos documentos indexados. Cada palavra da *query* e dos campos dos documentos será considerada **termo**. Suponha-se que um certo sistema de *Recuperação de Informação* receba a *query* "hidrografia do Paraná", que será dividida em termos ("hidrografia", "do", "Paraná"). O sistema então procurará, em todos os documentos indexados, o TF (*Term Frequency*), que irá contar quantas vezes os termos da *query* aparecem. Geralmente os termos dos documentos são contados no momento da indexação, para consulta rápida no momento da busca. Definimos a frequência de um termo como $f_{w,d}$, onde f é a frequência (quantidade) do termo w no documento (ou conjunto de documentos) d [Ramos, 2003]. É interessante comentar aqui que os termos podem ou não ser *case-insensitive*. Neste exemplo, na Tabela 2.1 da página 6, consideraremos que o são.

Já a parte de IDF (*Inverted Document Frequency*) calcula o quão comum um termo é, baseado no índice de todos os documentos indexados (quanto mais aparece, mais comum). Calcula-se o IDF de todos os termos da *query* ("hidrografia", "do" e "paraná"). Usando a fórmula $idf(w) = \log \frac{|D|}{f_{w,D}}$, onde $|D|$ é a quantidade total de documentos, e o termo w está presente em $f_{w,D}$ deles [Ramos, 2003]. Os documentos do exemplo com o TF*IDF calculado estão na Tabela 2.2 na página 6.

Considera-se que $w_{w,d} = f_{w,d} * \log(|D|/f_{w,D})$ é o peso de um termo w em um documento d [Ramos, 2003]. Utilizando a teoria do *Vector Space Model* [Salton et al., 1975], constroi-se um vetor $|query|$ -dimensão (quantidade de termos na *query*) para cada documento e para a *query*. Cada posição do vetor recebe o peso $w_{i,d}$ do termo i neste documento (ou *query*) d . Para saber a relevância do documento, mede-se a distância entre o seu vetor d e o da *query* q , utilizando a

similaridade do cosseno. O exemplo encontra-se na Tabela 2.3. A relevância *score* é dada por [Manning et al., 2008]:

$$score_{d,q} = \frac{\sum_{i=1}^t w_{i,d} * w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,d}^2} * \sqrt{\sum_{i=1}^t w_{i,q}^2}} \quad (2.1)$$

Tabela 2.1: Query, Documentos Encontrados e seus TF

Query	Termos	TF
"hidrografia do Paraná"	["hidrografia", "do", "paraná"]	hidrografia: 1 do: 1 paraná: 1
Documentos Encontrados	Termos	TF
nome: "hidrografia do Brasil" descrição: "Mapa com suas bacias."	["hidrografia", "do", "brasil", "mapa", "com", "suas", "bacias"]	hidrografia: 1 do: 1 paraná: 0
nome: "hidrografia do Paraná" descrição: "Bacia do Paraná e seus afluentes"	["hidrografia", "do", "paraná", "bacia", "e", "seus", "afluentes"]	hidrografia: 1 do: 2 paraná: 2

Tabela 2.2: Query, Documentos Encontrados e seus TF*IDF

Query	TF * IDF	IDF
"hidrografia do Paraná"	hidrografia: 0 do: 0,301 paraná: 0,301	hidrografia: $\log \frac{2}{1} = 0$ do: $\log \frac{2}{1} = 0,301$ paraná: $\log \frac{2}{1} = 0,301$
Documentos Encontrados	TF * IDF	
nome: "hidrografia do Brasil" descrição: "Mapa com suas bacias."	hidrografia: 0 do: 0,301 paraná: 0	
nome: "hidrografia do Paraná" descrição: "Bacia do Paraná e seus afluentes"	hidrografia: 0 do: 0,602 paraná: 0,602	

Tabela 2.3: Query, Documentos Encontrados e suas Relevâncias

Query	Vector Space	
"hidrografia do Paraná"	[0; 0,301; 0,301]	
Documentos Encontrados	Vector Space	Cosine Simillarity (Score)
nome: "hidrografia do Brasil" descrição: "Mapa com suas bacias."	[0; 0,301; 0]	0,602/0,602 = 1
nome: "hidrografia do Paraná" descrição: "Bacia do Paraná e seus afluentes"	[0; 0,602; 0,602]	0,362404/0,362404 = 1

Existem alguns motores de busca que fazem uso do TF-IDF, por padrão ou de forma opcional, como o *Elasticsearch* [Elastic, 2015a] e o *Apache Solr* [Apache, 2006] (baseados

na biblioteca *Apache Lucene* [Apache, 2005]); *Indri* (faz parte da biblioteca *Lemur Project* [UMassAmherst e CMU, 2000]) e *Xapian* [Xapian, 2000].

2.3 Simulação de Redes Sociais

Existem alguns modelos abstratos, para simulação de Redes Sociais, que se baseiam em um agente (usuário, no caso deste trabalho). Vejamos alguns [Amblard et al., 2015]:

1. *Regular networks* (RN), no qual os agentes estão em um *grid* e interagem com seus vizinhos. Usado em autômatos celulares;
2. *Erdős Rényi random graph model* (ER), onde considera-se um grafo de N nós (agentes), conectado por n arestas escolhidas das possíveis $N(N - 1)/2$ arestas. Gera uma rede aleatória;
3. *Small world models* (SW), que mistura a estabilidade das RN (tem um sistema inicial) com a aleatoriedade das ER (as relações após o início são aleatórias). É usado para simular redes sociais por ser fácil de implementar, mas observa-se empiricamente que redes reais possuem distribuição desigual na conectividade dos nós, o que não acontece no SW;
4. *Scale-free networks* (SF), que utiliza um grafo em que os nós (agentes) mais populares - mais conectados - tem maior probabilidade de ser conectados, ou seja, a rede é composta de vários pequenos grupos ligados entre si (seguem uma *power law*). Seu uso para simular redes sociais é debatido pois, enquanto abordam o problema da distribuição de conectividade de nós do SW, possui propriedades que também vão contra o pressuposto de muitos sistemas sociais;
5. *Spatial networks* (SN), que considera um espaço geográfico ou social (em que as dimensões são atributos diferentes), ligando os agentes que estão mais perto dentro de um limite de distância. Pode ser usada em conjunto com outros modelos para reforçar algumas características desejadas;
6. Outros modelos que geralmente são *ad-hoc* aplicados em uma rede específica (e assim, sem possibilidade de adaptação em outras redes).

No caso de um Site de Rede Social para Recursos Educacionais, a abordagem (1) RN é inviável porque o usuário está mais interessado em encontrar recursos relevantes (e não apenas os recursos que sua rede de conexão publica); (2) ER é aleatória e ignora o domínio do problema tratado; (3) SW é uma mescla das duas abordagens já descartadas, e no modelo que desejamos, seria difícil modelar uma rede muitos nós diferentes (como usuário, recurso, e demais objetos gerados pela rede social, que tem conexões entre si); (4) SF utiliza uma propriedade interessante das redes sociais, que é a *power law*, entretanto, no caso de trabalho, significa que os usuários interagiriam apenas com suas conexões próximas também (além de também necessitar de mais de um tipo de nó). (5) SF seria interessante se existisse o desejo de expor conteúdos regionais para os usuários. Por fim, percebe-se a necessidade de um modelo *ad-hoc* que lide com os problemas específicos do domínio do trabalho.

Como falado anteriormente, apesar de existirem usuários ativos no tipo de Site de Rede Social tratado no trabalho, a maioria estará apenas consumindo o conteúdo já produzido, o que torna a propriedade de *power law* do SF interessante. Mas o SF por si só não contabiliza os diferentes tipos de nós necessários para a simulação, então utiliza-se um SF *ad-hoc*.

2.4 Abordagens

Não foram encontrados trabalhos científicos que relacionam *Recuperação de Informação* em Redes Sociais para RE. O que existe são: *frameworks* que levam em consideração as conexões de usuários da Rede Social (mas não a relevância do documento para a comunidade), como o *SNDocRank*; Abordagens que melhoram o TF-IDF, como o MATF e as abordagens que ele compara. Em relação à Simulação de Rede Social para RE, também não existem trabalhos científicos específicos para o tema, o que levou a uma abordagem *ad-hoc*, tratada no próximo capítulo.

2.4.1 SNDocRank

O *framework SNDocRank* [Gou et al., 2010] considera o *Multi-Level Actor Similarity* (MAS), que é o interesse do usuário baseado em suas conexões com outras pessoas da Rede Social. Para isso, o grafo de conexões do usuário é manipulado para se calcular o quão semelhante é o usuário que efetuou a busca em relação aos usuários que publicaram os recursos encontrados por esta busca. Para calcular a relevância de um documento, é usado a fórmula $SNDoc(v, t_i, d_j) = t_{f_{i,j}} * idf_i * (1 + pNorm(\log MAS_{vu}))$ onde *SNDoc* é o a relevância encontrada, *v* é o usuário que está buscando recursos, *t_i* é um termo da *query*, *d_j* é o documento que está sendo comparado, *p* é uma constante definida pelos pesquisadores e *MAS_{vu}* é a similaridade entre o usuário *v* e o usuário que publicou o recurso.

Esta abordagem, entretanto, presume que os usuários que publicam recursos estão em uma (ou algumas) categoria de interesse (como "artes", "música" e outras). Dessa forma, um usuário muito ativo em algumas áreas acaba restrito à elas, ignorando possíveis recursos bons de áreas aparentemente desconexas de seu gosto (gosto esse que foi estabelecido pelo algoritmo MAS). Para além disso, esta metodologia fica em desacordo com a nova proposta nacional de educação [MEC, 2016], pensando no quesito de incentivo à interdisciplinaridade dos conteúdos.

2.4.2 MATF

O *Multi Aspect TF* [Paik, 2013] (MATF) é uma modificação do TF-IDF que leva em consideração a *Advanced TF Hypothesis* (AD-TFH). A AD-TFH considera mais significativo um documento com TF 3 em relação a outro com TF 2, do que um documento com TF 26 em relação a outro com TF 25. Também relaciona a *Relative Intra-document TF* (TF do documento é medido em relação ao TF médio nos documentos) com a *Length Regularized TF* (normaliza documentos muito grandes, pois eles tem mais termos e resulta em uma chance maior de aparecer nos resultados). O MATF é comparado com outras modificações estado da arte de TF-IDF e tem resultados superiores à todas.

A abordagem é nitidamente superior às outras implementações do TF-IDF, mas no contexto do trabalho, não se leva em consideração nenhuma contribuição da comunidade (via Rede Social) em relação aos documentos retornados. Assim, apesar de agregar relevância aos resultados, continua sendo uma implementação que não trata do problema encontrado em Redes Sociais para RE.

Capítulo 3

Ranqueamento de Recursos Educacionais

A elaboração de uma solução para Ranqueamento de Recursos Educacionais apresentada nesse trabalho, faz parte do desenvolvimento do PortalMEC, um Site de Rede Social voltado a Recursos Educacionais. Este Ranqueamento será integrado com um sistema de IR que utiliza um método de TF-IDF modificado. Para a finalidade dos experimentos, criou-se um Simulador de Rede Social, chamado Populador, que simula usuários interagindo no portal. Em seguida, são apresentados os experimentos que mostram, de fato, como os pesos do ranqueamento interagem com a relevância dos RE encontrados.

3.1 PortalMEC

O PortalMEC, desenvolvido pelo C3SL (Centro de Computação Científica e Software Livre) para o MEC (Ministério da Educação), é um portal de Recursos Educacionais que agrega elementos de Redes Sociais. Nos portais existentes do MEC e de outras iniciativas, nota-se uma interface com o usuário defasada, visto que elas foram criadas a alguns anos e não foram atualizadas recentemente. Também todos os portais dependem de um processo burocrático de aprovação de novos recursos.

Por tal razão, finda-se por criar uma barreira desnecessária ao acesso dos Recursos Educacionais. Adindo, o mecanismo de busca dos mesmos é muito simples, retornando resultados unicamente baseados no termo da pesquisa, sem considerar a relevância do retorno da busca. O PortalMEC (Figura 3.2 na página 11) visa solucionar os problemas comentados. Uma das formas de melhorar a busca de recursos relevantes é o Ranqueamento tratado no presente trabalho.

Contabilizando apenas os modelos necessários para o Ranqueamento e busca (o Modelo Entidade Relacionamento empregado está na Figura 3.1 da página 10), o PortalMEC é utilizado por Usuários [*User*], que manipulam os Recursos Educacionais [*LearningObject*], e Coleções [*Collection*] (agrupamento de Recursos Educacionais e Coleções, neste trabalho considera-se apenas as públicas). Dados os Recursos Educacionais e Coleções, os usuários poderão: buscá-los (Figura 3.3 na página 11); visualizá-los [*View*] (Figura 3.2 na página 11); favoritá-los [*Like*] (guarda os Recursos Educacionais e Coleções preferidas para acesso posterior); compartilhá-los em outras Redes Sociais [*Share*] (*Facebook*, *Twitter* e outras); baixá-los [*Download*]; relatar sua experiência após ter utilizado [*Review*] e, avaliar a experiência de outros usuários [*Rate*] (se experiência foi útil ou não para si).

O Ranqueamento é importante para o portal e dependerá dos usuários para alimentá-lo, o que requer a elaboração de algum tipo de incentivo para que esses cadastrem-se na plataforma. Para isso, usuários não cadastrados podem utilizar apenas a busca e visualizar

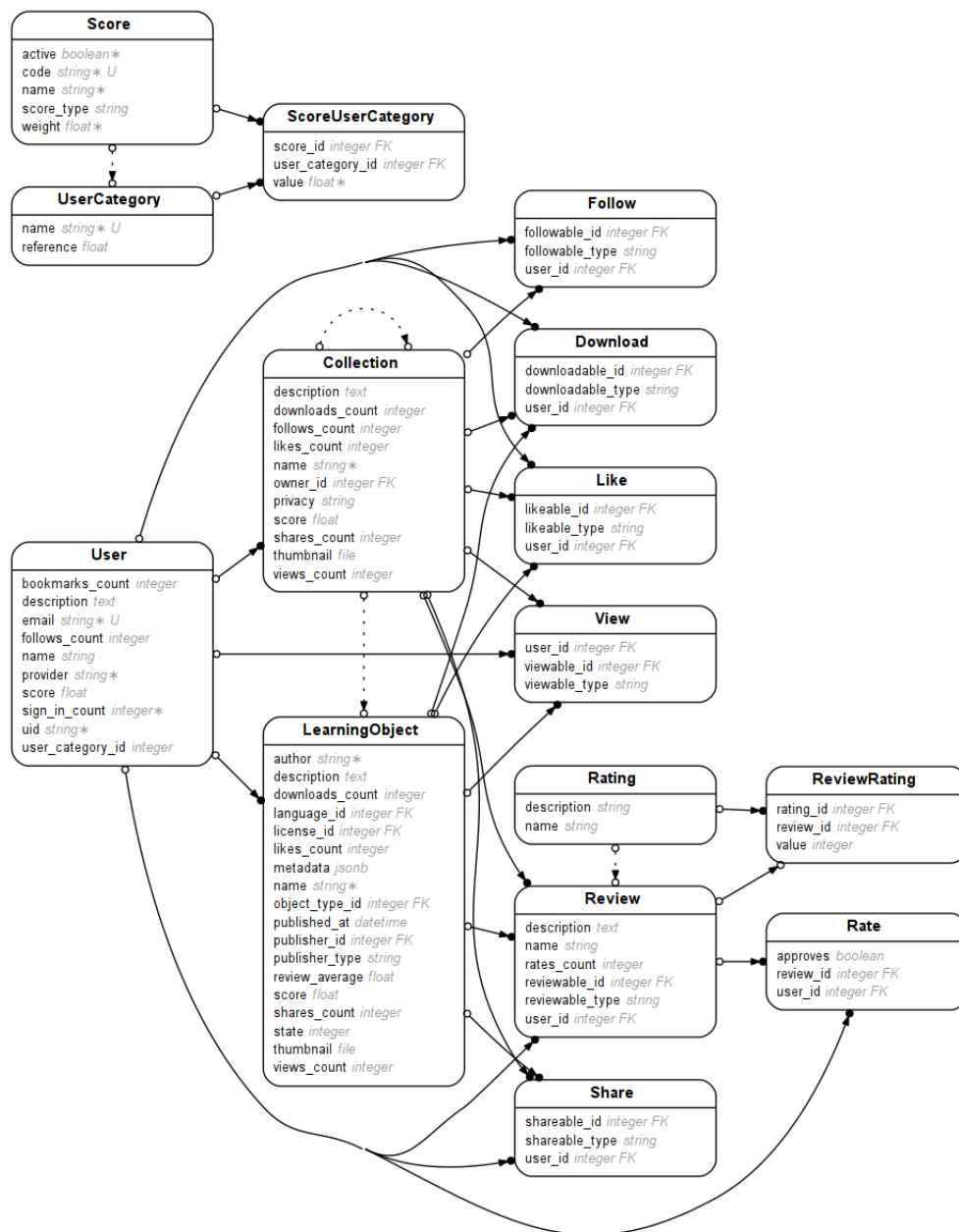


Figura 3.1: Modelo Entidade Relacionamento simplificado do PortalMEC

recursos educacionais, coleções e usuários, enquanto os usuários cadastrados podem utilizar todos os recursos da plataforma.

3.2 Ranqueamento

O Ranqueamento é uma forma de utilizar a inteligência e interações dos usuários da Rede Social a favor de um melhor compartilhamento de recursos relevantes. No PortalMEC, busca-se por Recursos Educacionais, Coleções e Usuários. Logo, todas essas entidades possuem um *score* (*rank* ou reputação) que será integrado com a busca. O sistema de ranqueamento será

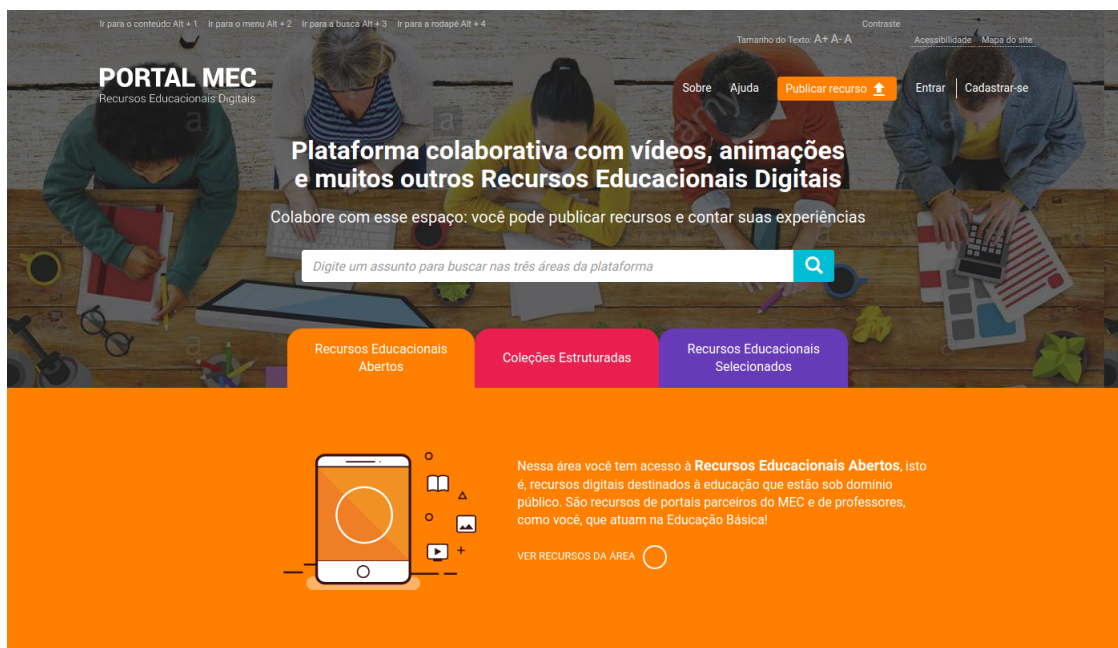


Figura 3.2: PortalMEC - Capa

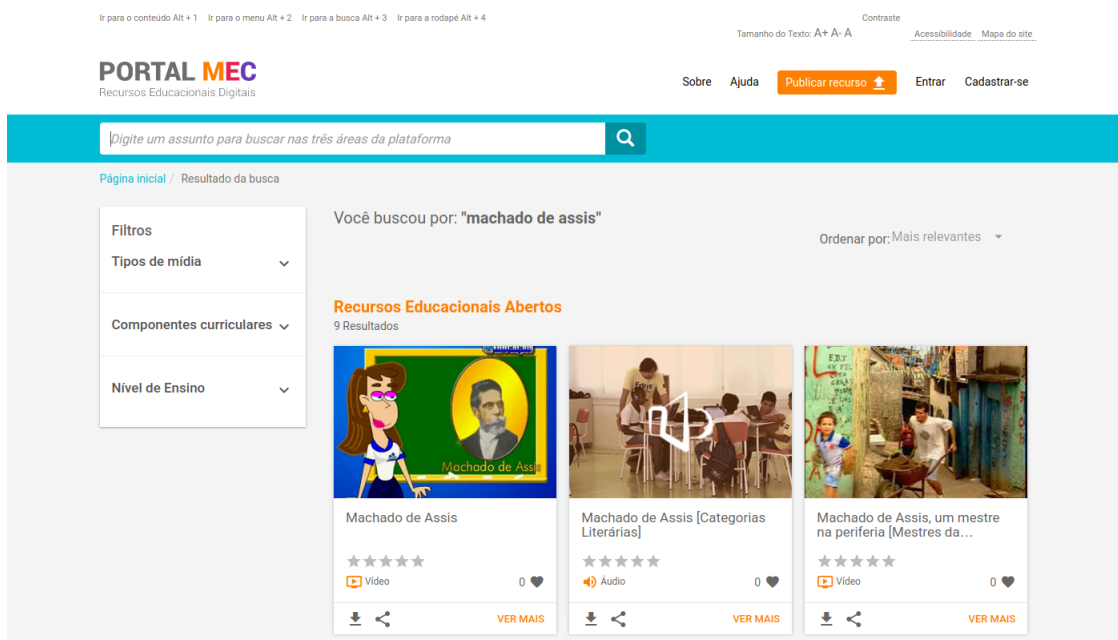


Figura 3.3: PortalMEC - Busca de Recursos

dividido no *score (rank)* dos Recursos Educacionais e Coleções com a reputação dos Usuários (que no modelo é chamada de *score* também para facilitar a implementação).

Levando em conta o domínio dos modelos com que se relaciona no portal, o Ranqueamento, portanto, utilizará critérios específicos para o propósito do PortalMEC. Esses critérios foram levantados em reuniões com os professores responsáveis pelo projeto, pelo autor deste trabalho e pelo colega Lucas Ernesto Kindinger.

Os critérios que irão compor o *score (rank)* dos Recursos Educacionais e Coleções são:

- **presença de *thumbnail* no recurso:** por ser visualmente mais atrativo para quem busca um recurso;

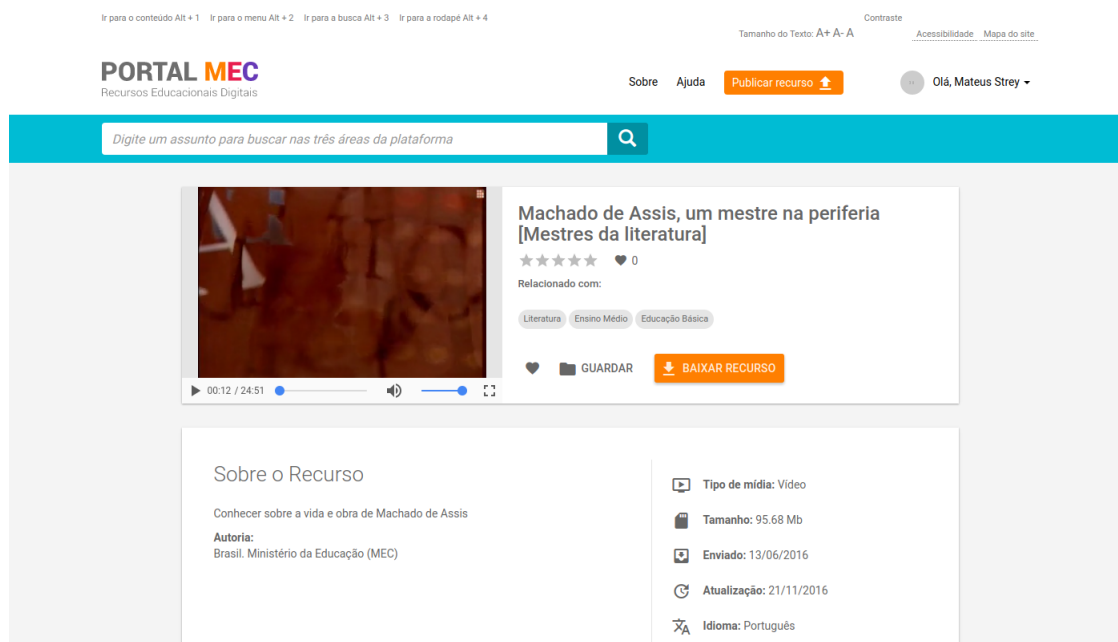


Figura 3.4: PortalMEC - Recurso Educacional

- **quantidade de visualizações do recurso [View]:** um recurso com alto índice de visualizações indica popularidade e provavelmente é relevante;
- **quantidade de favoritos [Like] do recurso:** se os usuários favoritaram em peso, pode deixá-lo mais atraente;
- **quantidade de downloads [Download] do recurso:** se muitos usuários baixaram, é provável que seja relevante;
- **quantidade de compartilhamentos em redes sociais [Share] do recurso:** se os usuários o compartilharam muitas vezes, pode indicar relevância;
- **média das experiências [Review] desse recurso:** a experiência dos usuários é muito importante para a avaliar a qualidade do recurso, pois ao relatarem se o aprovam ou não (e o porquê) podem auxiliar outros usuários a saberem se este recurso é coeso com suas necessidades;
- **presença de descrição no recurso:** a descrição deixa o recurso mais claro, pois pode resumir qual a finalidade dele e para qual público é indicado, por exemplo, tornando-o, assim, mais atrativo;
- **categoria do usuário que criou o recurso (baseada na sua reputação):** usuários bem reputados também o são com base em bons recursos que criaram; por isso, a confiabilidade sobre o recurso aumenta, tornando-o relevante;
- **adição do recurso em coleções dos usuários:** se os usuários guardaram nas suas coleções certo recurso, é porque talvez o consideraram válido.

Já os critérios que compõe a reputação dos Usuários são:

- **adição dos recursos educacionais do usuário em coleções de terceiros com score (rank) acima da média:** se seus recursos foram colecionados (principalmente em coleções bem reputadas), é relevante.

- **aprovação das experiências do usuário em recursos de terceiros (porcentagem das avaliações positivas):** se suas experiências foram verificadas e validadas por terceiros, é relevante.
- **avaliação média dos recursos do usuário:** se o usuário mantém um padrão de qualidade nos seus recursos publicados, é relevante.
- **melhor *score* (*rank*) entre os recursos do usuário:** se o usuário possui um recurso bem avaliado, é relevante.
- **quantidade de recursos enviados pelo usuário:** se o usuário é publica ou publicou bastantes recursos, é relevante.
- **quantidade de seguidores:** se o usuário é seguido por outros usuários, é relevante.
- **score médio das coleções públicas do usuário:** se o usuário mantém um padrão de qualidade nas suas coleções públicas, é relevante.
- **score médio dos recursos do usuário:** se o usuário mantém um padrão de qualidade nas suas coleções públicas, é relevante.
- **quantidade de submissões recentes do usuário:** se o usuário é ativo atualmente, é relevante.

Para se calcular o *score* (*rank* ou reputação), a nota obtida no critério pelo objeto é dividida pela nota máxima que esse critério alcançou (na classificação geral) e multiplicada por um peso ajustável. Por exemplo: um RE possui 15 visualizações, o RE mais visualizado possui 300, e o peso para visualização é 3, temos como *score* 0,15. O *score* (*rank* ou reputação) do objeto que está sendo avaliado é, então, calculado somando os valores retornados por todos os critérios:

$$score(rank|reputacao) = \sum pesodocriterio * \frac{notadocriterio}{notamaximadocriterio} \quad (3.1)$$

Os pesos cogitados estão nas tabelas 3.1 e 3.2 da página 14.

Como o *score* (*rank*) dos Recursos e Coleções se alimenta do *score* (reputação) dos Usuários e vice-versa, resultaria-se um ciclo em que quanto mais se aumenta o *rank*, mais aumentaria a reputação e assim por diante. Logo, usuários com muita reputação podem eclipsar usuários mais novos ou menos relevantes, bem como os usuários pouco reputados (inclusive possíveis perfis *fakes*, com auxílio de grupos organizados externamente) podem subir o *score* de recursos não relevantes. Implementou-se um sistema de categoria de usuários, no qual foram definidas quantidades médias em cada critério da reputação para cada categoria de usuário. Dessa forma, a reputação fica comportada dentro de uma categoria, e o RE ou a Coleção não ganharão *score* (*rank*) de forma cíclica, evitando os problemas citados. A tabela 3.3 da página 14 expõe as categorias de usuário e suas médias classificatórias. Essas categorias variam em 5 níveis, sendo que a C1 seria um usuário novo (possui cadastro a pouco tempo, publica pouco, interagem pouco, posta pouco) e a C5 seria um usuário experiente, participativo e mais confiável.

3.3 Busca via Elasticsearch

Elasticsearch é um motor de *analytics* e busca *full-text* (*near*) *real time*. Ainda é *opensource*, distribuído, com API Java e RESTful, orientado à documentos e construído em

Tabela 3.1: Pesos do Ranqueamento de Recursos

	Peso 1	Peso 2	Peso 3
Média das avaliações	10	5	7
Thumbnail	3	1	1
Quantidade de visualizações	3	2	2
Quantidade de likes	7	3	2
Quantidade de downloads	3	3	4
Quantidade de compartilhamento	2	3	4
Descrição	2	1	1
Categoria do usuário criador	7	4	5
Adições à coleções	4	3	5

Tabela 3.2: Pesos da Reputação dos Usuários

	Peso 1	Peso 2	Peso 3
Quantidade de envio de objetos	4	1	1
Avaliação média dos seus objetos	9	5	6
Adições a coleções bem avaliadas	7	3	3
Melhor score	3	1	1
Score médio	7	5	5
Aprovação das avaliações em objetos de terceiros	10	4	5
Quantidade de seguidores	6	1	2
Submissões recentes	4	1	2
Score médio das coleções públicas	3	3	3

Tabela 3.3: Categoria dos Usuários

	C1	C2	C3	C4	C5
Quantidade de envio de objetos	0	25	50	100	250
Avaliação média dos seus objetos	0	3,0	3,8	4,2	4,9
Adições a coleções bem avaliadas	0	10	20	40	80
Melhor score	0	0,5	0,55	0,6	0,68
Score médio	0	0,25	0,35	0,45	0,55
Aprovação das avaliações em objetos de terceiros	0	70%	85%	90%	95%
Quantidade de seguidores	0	50	250	1000	5000
Submissões recentes	0	5	15	25	35
Score médio das coleções públicas	0	0,4	0,6	0,75	0,8

cima do *Apache Lucene* [Elastic, 2016]. Empresas e projetos com escopos bem diferentes umas das outras o usam, como *Wikimedia* [Horohoe, 2014], *Netflix* [Loke e Kalantzis, 2014], CERN [Horanyi, 2014], GitHub [Pease, 2013], Quora [Ecoffet, 2013]. Foi optado por também utilizá-lo no PortalMEC devido à robustez, rapidez no retorno da busca e facilidade de uso.

Para efetuar uma busca por documentos, o Elasticsearch (1) usa o modelo booleano do Apache Lucene para encontrar os documentos que possuem os termos da *query*, (2) calcula a relevância dos documentos encontrados com uma técnica TF/IDF modificada (adiciona *coordination factor*, normalização do tamanho do campo, e boost no peso do termo), e (3) retorna os documentos ordenados pela relevância (score) de forma decrescente. [Elastic, 2015b]

No PortalMEC, inicialmente, buscas retornadas pelo *Elasticsearch* eram ordenadas de forma decrescente pelo *score* do ranqueamento (*rank* ou reputação). Esse método se mostrou inefetivo devido aos *scores* (*rank*) muito próximos (sem o lançamento oficial da plataforma, a base estava quase sem interações de usuários) e por ignorar totalmente a relevância (*score*) gerada pelo *Elasticsearch*, dando prioridade a recursos pouco relevantes à *query* do usuário.

Foi mesclado então o *score* do ranqueamento (*rank* ou reputação) dos recursos com o *score* gerado pelo *Elasticsearch* de forma que a *query* seja respeitada enquanto também considera o ranqueamento do portal. As mesclas cogitadas estão na Tabela 3.4 da página 15.

Tabela 3.4: Mesclas possíveis de score/rank

	fórmula
mescla 1 (m1)	$score * rank$
mescla 2 (m2)	$score * \log(rank + 1)$
mescla 3 (m3)	$\log(score + 1, 1) * (rank + 1)$

Cada mescla possui uma motivação, ressaltando que o *score* $[0, \infty[$ (não pode ser normalizado, visto que não expõe o *score* máximo da *query* utilizada) e o *rank* $[0, 1]$:

- **m1**: mescla integralmente o *score* com o *rank*. É bem influenciado pelo *rank*.
- **m2**: mescla o *score* com o *log* do *rank* (o +1 serve para não retornar um valor menor que 1 e aumentar a prioridade de quem tem pouco ranqueamento). Atenua a influência do ranqueamento no resultado.
- **m3**: mescla o *log* do *score* (+1,1 serve para não retornar 0 quando o *score* for 0) e o *rank* (+1 serve para não multiplicar por 0). Atenua valores extremos do *score* enquanto leva em consideração a reputação.

3.4 Populador

Com o não lançamento do PortalMEC neste ano (2016), criou-se uma biblioteca que popula o portal a fim de realizar os experimentos deste trabalho. Esta biblioteca, chamada Populador e é baseada em perfis de usuários que realizam ações, num período de tempo, na rede. Instanciamos o Populador com a duração total da geração, duração de cada período da geração, data inicial, quantidade total de usuários gerados e distribuição dos perfis para os usuários. Exemplificando abaixo com um pseudo-código:

Populador

```

inicializa() {
    data inicial: 01 de janeiro de 2015
    duração total da geração: 1 ano
    duração de cada período da geração: 1 mês
    usuários gerados: 1000
    distribuição dos perfis: 80\% lurker e 20\% poster

    // - Depósito: responsável por armazenar todos os usuários
    //       e objetos gerados pelo Populador.
    // - Gerador de Usuários: responsável por criar os usuários

```

```

//      do período corrente, definindo seus perfis.
// - Despachante de Ações: realiza as ações de todos os
//      usuários gerados, conforme seus perfis.
Depósito.inicializa()
GeradorDeUsuários.inicializa()
DespachanteDeAções.inicializa()

    gera()
}

gera() {
    periodos = duração total da geração /
              duração de cada período da geração

    para cada periodo dos periodos {
        Depósito.incrementa_periodo()
        GeradorDeUsuários.popula_periodo()
        DespachanteDeAções.popula_periodo()
    }
}

```

As ações são, por exemplo, a criação de um recurso ou visualização de uma coleção, e podem acontecer múltiplas vezes durante o período conforme o perfil. Cria-se as que fazem parte do cálculo de *score (rank)* do Ranqueamento. É possível atrelar os objetos gerados por uma ação a outras ações (filhos), e no nosso do portal, as visualizações (*Views*) são base para as ações de criação de *download*, favoritar, compartilhar, experiência e adição à coleções de RE e coleções.

```

DespachanteDeAções
    inicializa() {
        // recebe uma instância do Depósito
        Depósito.instância()
    }

    popula() {
        // O Depósito armazena usuários separando-os por perfil
        para cada perfil de um grupo de usuários do Deposito {
            para cada ação do perfil {
                // A Ação executa as ações (cria os objetos)
                // para o tipo de perfil escolhido com o
                // grupo de usuários fornecido.
                Ação.itera(grupo de usuários do perfil)
            }
        }
    }
}

Ação
    inicializa() {
        // filhos são Ações que dependem dos objetos gerados
        // pela presente Ação
    }

```

```

    filhos: nenhum
    quantidade de objetos: 1 até 100
    objetos_ação: Vetor
    Deposito.instancia()
}

itera(usuários) {
    para cada usuário dos usuários {
        objetos = cria_acao(usuario)
        objetos_ação recebe objetos
    }

    Deposito.importa(objetos_ação)

    se existem filhos {
        itera_filhos(usuários)
    }
}

cria_acao(usuario) {
    n = número aleatório entre a quantidade de objetos
    objetos = Vetor
    n vezes {
        // cada Ação implementa sua forma de criar o objeto
        // que a representa
        objeto = cria_objeto_ação(usuario)
        objetos recebe objeto
    }
    retorna objetos
}

itera_filhos(usuários) {
    para cada filho dos filhos {
        para cada usuário dos usuários {
            para cada objeto dos objetos_ação {
                filho.itera(usuario, objeto)
            }
        }
    }
}

```

No final da execução, tem-se uma base contendo usuários e suas interações no portal, permitindo que se realizem os experimentos do presente trabalho. Os perfis escolhidos serão tratados por **Poster** (usuário que cria e interage com recursos) e **Lurker** (usuário que interage com os recursos, raramente criando algum). Na Tabela 3.5 da página 18 é possível observar as ações realizadas por perfil, o tipo de objeto criado e quantas vezes elas podem ocorrer em um período.

Tabela 3.5: Perfis de Usuário e suas Ações no Populador

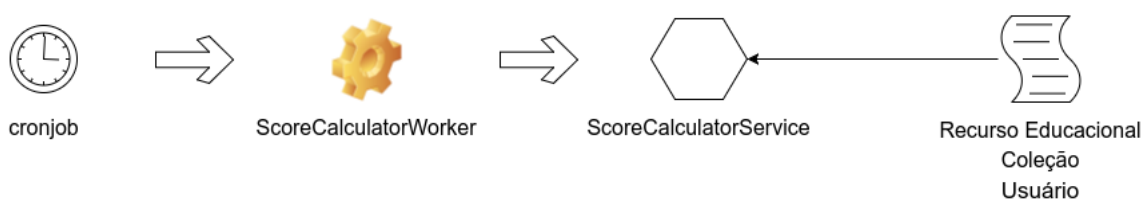
	Objeto Criado	Dependência	Perfil Objetos/Período	
			<i>Poster</i>	<i>Lurker</i>
Publicar Recurso Educacional	LearningObject	-	1 até 20	0 até 1
Publicar Coleção	Collection	-	1 até 5	0 até 5
Visualizar RE ou Coleção	View	LearningObject Collection	10 até 100	10 até 100
Favoritar RE ou Coleção	Like	View	0 até 20	1 até 50
Baixar RE ou Coleção	Download	View	0 até 20	1 até 30
Compartilhar RE ou Coleção	Share	View	0 até 5	1 até 30
Seguir quem publicou RE ou Coleção	Follow	View	0 até 10	0 até 15
Adicionar RE ou Coleção em uma Coleção	CollectionItem	View	0 até 20	0 até 50
Relatar experiência com um RE ou Coleção	Review	View	0 até 10	0 até 20
Avaliar experiência em um RE ou Coleção	Rate	View	0 até 10	0 até 50

3.5 Implementação

O *backend* do PortalMEC é desenvolvido em *Ruby* com o *framework Rails* e funciona como uma *API RESTful*, enquanto seu *frontend* é desenvolvido em *Javascript* com o *framework Angular*. Utiliza o *PostgreSQL* como sistema de gerenciamento de banco de dados (SGDB), *Dspace* para armazenamento persistente dos RE, *Sidekiq* como servidor de processamento em *background*, e a gema (biblioteca em *Ruby*) *Searchkick* para comunicar com o *Elasticsearch*.

Para o Ranqueamento, criamos *concerns - pattern* que expõe métodos genéricos - para os modelos utilizados (*Scoreable*: RE e Coleções; *Reputationable*: Usuário), responsáveis por retornar as notas de cada critério utilizado. Essas notas são utilizadas pelo *ScoreCalculatorService*, serviço que calcula o *score* (*rank* ou reputação) dos RE, Coleções e Usuários, conforme explicado na seção sobre o Ranqueamento.

O servidor de tarefas *Sidekiq* funciona alheio ao processo do servidor *web* (pois evita problemas de performance para quem utiliza o site) e utiliza uma fila de processos (*workers*) que serão executados assim que possível e conforme sua prioridade. Os *scores* são calculados a cada período de tempo (atualmente 6 horas, decidido em reunião) através de um *cronjob* (tarefas agendadas para sistemas *Unix-like*) que chama um *ScoreCalculatorWorker* para cada RE, Coleção e Usuário do portal. Este *worker* é responsável por instanciar o *ScoreCalculatorService*. O fluxo do cálculo do *score* está representado na Figura 3.5 da página 18.

Figura 3.5: Fluxo do cálculo do *score* no PortalMEC

Em relação à busca, o *Elasticsearch* permite o uso de *scripts* para alterar o *score* calculado pelo programa (TF-IDF). Cria-se os *scripts* para mesclar o *score* (*rank* ou reputação) do Ranqueamento com o *score* do *ElasticSearch*, conforme a Tabela 3.4 na página 15. Comunica-se com o *Elasticsearch* pela gema *Searchkick*, que não suporta o uso de *scripts*, sendo necessária a construção parcial das *queries* enviadas para o *Elasticsearch* através do *SearchService* (serviço de busca).

Utiliza-se a estrutura do PortalMEC para implementar o Populador, aproveitando a comunicação com o banco de dados e os modelos dos objetos. Para geração dos campos dos objetos, a gema *Faker* inventa nomes, e-mails, endereços, frases e afins. A gema *parallel* permite a utilização de várias *threads*, aumentando a performance da geração de objetos que não tem dependência entre si.

3.6 Experimentos

A máquina utilizada possui processador *Intel(R) Core(TM) i5-4460 3.20GHz*, 2 memórias RAM *DIMM DDR3 8GB 1600Mhz*, disco rígido interno SATA de 1TB, rodando o sistema operacional *Fedora 24 Xfce Workstation*. Os experimentos foram realizados em duas situações: usando os recursos importados do PortalMEC (19.000 RE, sem gerar novos RE) e gerando RE (com conteúdo em latim). Para cada uma delas, utilizamos o Populador com: duração total de 1 ano; período de 1 mês; distribuição de usuários em 99% *lurker* e 1% *poster*; geração de 120, 1.000, 5.000 e 10.000 usuários; para cada quantidade usuários, efetua-se duas buscas por RE. Optou-se por tal proporção de *posters* e *lurkers* porque é sabido que muitos usuários preferem consumir passivamente os RE, sem compartilhar sua experiência com outros usuários [Sclater, 2016]. Cada busca compara os 15 primeiros RE retornados apenas com o TF-IDF, com 3 pesos diferentes de ranqueamento e 3 formas de mesclar o *score* do TF-IDF-*Elasticsearch* (*Score*) com o *score* do ranqueamento (*Rank*), exibindo a nova colocação do recurso. As tabelas com os Experimentos encontram-se no Apêndice A.

3.6.1 Com recursos importados

A *query* "dna" possui o primeiro elemento muito díspar (em relação ao *Score*) dos elementos seguintes. Percebe-se que este elemento não altera sua posição independente do peso ou mescla utilizados (em parte por seu *Rank* nunca estar entre os menores). No mais, verifica-se que as mesclas m1 e m2 são menos conservadoras com a posição do RE ordenado apenas pelo TF-IDF do que a mescla m3.

Já a *query* "molécula" possui elementos próximos em relação ao *Score*. Nesta é possível perceber como o *Rank* influencia a ordenação do resultado, onde provavelmente teríamos vários RE com campos com conteúdo parecidos, mas com a relevância dos RE muito diferente. Também verifica-se que a mescla m3 é muito mais conservadora que as mesclas m1 e m2.

Comentando cada grupo de usuários:

- **120 usuários:** As mesclas m1 e m2 afetam a ordenação em todos os pesos enquanto a m3 afeta pouco, principalmente nos recursos com *Rank* abaixo de 0.2. Nota-se que o *Rank* é mais homogêneo entre os recursos por não existirem muitas interações;
- **500 usuários:** O *Rank* varia entre 0,19 e 0,43 no Peso 1 da *query* "molécula" (e de forma similar nos outros pesos), permitindo a observação de forma mais precisa da variação causada na ordenação. Por exemplo, o 6º recurso, com *Rank* abaixo de 0,2 em todos os pesos, alcançou a 21ª 22ª posição para m1 e m2, enquanto ficou em 13ª com m3;

- **1.000 usuários:** Na duas *queries* o Rank varia entre 0,4 e 0,2 (no Peso 1) e observa-se variações grandes de ordenação em todas as mesclas (com Score homogêneo e heterogêneo);
- **5.000 usuários:** O Rank observado nos recursos das duas *queries* está bem homogêneo, variando entre 0,38 e 0,28 no Peso 1. Verifica-se pouca alteração na ordem dos resultados;
- **10.000 usuários:** O Rank continua homogêneo, provavelmente porque os recursos são todos de alguns publicadores (importados de alguns portais) e os usuários acabam todos interagindo com os mesmos recursos sempre (não há criação de recursos). De maneira similar aos 5.000 usuários, a ordenação é pouco alterada;

3.6.2 Com recursos gerados

De maneira similar aos recursos importados, os resultados com Score mais homogêneos são os que mais sofrem influência do *Rank*. A mescla m1 é a que mais sofre influência do *Rank*, seguida da mescla m2. A mescla m3 é a mais conservadora em relação a ordenação do TF-IDF.

Comentando cada grupo de usuários:

- **120 usuários:** A query "*sint*" possui Score mais homogêneo que a query "*verit*". Por exemplo, o 14º recurso desta query foi rebaixado para 38º lugar nos Pesos 1 e 3 nas mesclas m1 e m2. Na query "*verit*", o 9º recurso foi rebaixado para a 43ª posição no Peso 1 e mescla m1. Apesar de possuir umas mudanças maiores em relação a posição de alguns recursos, a query "*verit*" ainda foi menos afetada em quantidade de recursos que a outra, já que possui um Score mais heterogêneo;
- **500 usuários:** A mudança de posição é mais agressiva, como no 13º recurso da query "*velit*" que foi rebaixado para a 88ª posição no Peso 1 com a mescla m1. Isso reflete uma combinação de Score homogêneo com Rank inferior a todos os outros recursos listados;
- **1.000 usuários:** O Score homogêneo nos recursos da query "*sint*" os torna muito influenciados pelo Rank, como observado no 4º recurso, que possui o menor Rank entre os recursos retornados, sendo muito rebaixado. De forma parecida, na query "*velit*" os 6º e 7º recursos são rebaixados em muitas posições (mais de 100 para o 7º, nas mesclas m1 e m2);
- **5.000 usuários:** Com mais usuários e mais recursos criados, existe uma saturação de recursos com Score e Rank parecidos, de forma que na query "*sint*", com exceção do 1º recurso, todos os recursos foram rebaixados para posições abaixo da 15ª;
- **10.000 usuários:** Seguindo a lógica do aumento de usuários e recursos, RE com Rank menores ou iguais à 0,2 são muito rebaixados, como o 6º e 10º recursos da query "*sint*" ou o 5º recurso da query "*velit*". Os resultados estão muito afetados pelo Rank, provavelmente pelo Score estar bem homogêneo para a maior parte dos resultados (15 mil para a query "*sint*" e 11 mil para a query "*velit*");

3.6.3 Resultados

Em relação aos pesos, é necessário a validação da comunidade após o uso, ou mesmo discussão à respeito de quais critérios serão mais considerados. Pelos experimentos, é possível perceber que o Peso 1 é o que atribui maior *Rank*, enquanto o Peso 2 e 3 ficam bem próximos um do outro. Em compensação, as mesclas são um pouco mais claras em relação à suas utilidades: a mescla m1 utiliza integralmente o *Rank* e pode trazer muitos resultados pouco relevantes em

relação à *query* (apesar de serem bem ranqueados) para o topo dos resultados; a mescla m2 é mais moderada e é uma boa candidata a união que será implementada no portal, necessitando validação para confirmar sua utilidade; a mescla m3 é a mais conservadora em relação à posição dos resultados sem Ranqueamento e pode servir como substituta da m2, caso não seja boa em considerar a relevância da *query*.

Capítulo 4

Conclusões

Após o desenvolvimento de todas as etapas propostas, percebeu-se, primeiramente, a baixa disponibilidade de fontes científicas que abordem especificamente a problemática do trabalho, sendo necessária a criação de um método *ad-hoc* inspirado em algumas soluções e empirismo.

Ainda, após várias pesquisas e reuniões, confirmou-se necessária a existência de um portal que reunisse os RE dispersos na *web*, e para tal, a elaboração de um sistema que recupere os RE mais relevantes para os usuários (possivelmente professores), a fim de estimulá-los a frequentar e contribuir com a Rede.

Como o PortalMEC ainda não foi inaugurado, o Populador serviu como uma ótima ferramenta de teste e possibilitou uma melhor compreensão de como a busca se comportará em relação ao Ranqueamento proposto. Porém, esse recurso, obviamente, não tira a necessidade de realizar a validação do mecanismo com usuários reais no sistema. Quando o portal estiver funcionando, será possível testar os diferentes pesos e decidir qual é o mais indicado para a comunidade.

Uma observação bastante relevante é que os experimentos indicam que a comunidade precisará ser ativa ao interagir com os objetos para o ranqueamento ter sentido. Existem algumas iniciativas que permitem o diálogo e o uso colaborativo dos REA (como o *OpenEd*) e todas aumentam o conhecimento associado a eles. Entretanto, é sabido que muitos usuários preferem consumir passivamente os conteúdos online, sem compartilhar sua experiência com outros usuários [Sclater, 2016]. De forma que, mesmo que só existam alguns membros ativos no portal, a comunidade como um todo terá uma busca com mais sentido e relevância. Fica assim, como desafio futuro, pensar em possibilidades de como fazer com que uma parcela significativa de usuários participem ativamente no portal.

Falando sobre trabalhos futuros, é possível também: trocar a fórmula TF-IDF utilizada pelo *ElasticSearch* (como a BM25, disponível na versão mais recente da ferramenta); adicionar, remover ou modificar critérios para os pesos do Ranqueamento como forma de apurá-lo; encontrar novas maneiras de mesclar o TF-IDF com o Ranqueamento; integrar a área de interesse do usuário na busca (cuidando para não privá-lo de outras áreas e acabar com a interdisciplinaridade desejada); unir interesses do usuário retornados por uma solução em *Machine Learning*, desenvolvida como parte do projeto, junto ao cálculo de relevância.

Fica, ademais, um incentivo para que outros cientistas da computação ou de áreas afins, contribuam para a questão aqui tratada e pensem em novas possibilidades ou, inclusive, transformem o que é aplicado atualmente, em trabalhos ou artigos científicos, a fim de compartilhar as diferentes metodologias criadas, mas que são brevemente divulgadas.

O modelo de busca aqui desenvolvido mostrou-se bastante eficaz para resolver o problema presente no quase debutado PortalMEC, projeto esse que acredita-se ser de grande utilidade à sociedade civil brasileira, em especial à área de Educação, por facilitar o contato entre pessoas interessadas e conhecimentos já construídos em diversas regiões do país (mas escondido por não haver meios melhores de recuperar informação no contexto apresentado). Dessa maneira, o autor acredita que a produção e disponibilização desse conhecimento, mesmo não sendo uma grande descoberta, é uma forma também de retribuir à sociedade o investimento a ele feito quando ingressou numa universidade pública (UFPR).

Referências Bibliográficas

- [ACT, 2009] ACT (2009). Opened. <https://www.opened.com/>. Acessado em 30/11/2016.
- [Amblard et al., 2015] Amblard, F., Bouadjio-Boulic, A., Gutiérrez, C. S. e Gaudou, B. (2015). Which models are used in social simulation to generate social networks? a review of 17 years of publications in jasss. Em *2015 Winter Simulation Conference (WSC)*, páginas 4021–4032.
- [Apache, 2005] Apache (2005). Apache lucene. <https://lucene.apache.org/>. Acessado em 30/11/2016.
- [Apache, 2006] Apache (2006). Apache solr. <https://lucene.apache.org/solr/>. Acessado em 30/11/2016.
- [Benediktsson et al., 1990] Benediktsson, J. A., Swain, P. H. e Ersoy, O. K. (1990). Neural network approaches versus statistical methods in classification of multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):540–552.
- [Blosseville et al., 1992] Blosseville, M. J., Hébrail, G., Monteil, M. G. e Pénot, N. (1992). Automatic document classification: Natural language processing, statistical analysis, and expert system techniques used together. Em *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '92*, páginas 51–58, New York, NY, USA. ACM.
- [boyd e Ellison, 2007] boyd, d. m. e Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230.
- [Curriki, 2006] Curriki (2006). Curriki. <http://www.curriki.org/>. Acessado em 30/11/2016.
- [Ecoffet, 2013] Ecoffet, A. L. (2013). What technology does quora use for its full-text search infrastructure? <https://www.quora.com/Full-Text-Search-Quora-feature/What-technology-does-Quora-use-for-its-full-text-search-infrastructure/answer/Adrien-Lucas-Ecoffet>. Acessado em 30/11/2016.
- [Elastic, 2015a] Elastic (2015a). Elasticsearch. <https://www.elastic.co/>. Acessado em 30/11/2016.
- [Elastic, 2015b] Elastic (2015b). Elasticsearch guide: Scoring theory. <https://www.elastic.co/guide/en/elasticsearch/guide/current/scoring-theory.html>. Acessado em 30/11/2016.

- [Elastic, 2016] Elastic (2016). Elasticsearch github. <https://github.com/elastic/elasticsearch>. Acessado em 30/11/2016.
- [Goffman, 1964] Goffman, W. (1964). A searching procedure for information retrieval. *Information Storage and Retrieval*, 2(2):73 – 78.
- [Gou et al., 2010] Gou, L., Zhang, X. L., Chen, H.-H., Kim, J.-H. e Giles, C. L. (2010). Social network document ranking. Em *Proceedings of the 10th Annual Joint Conference on Digital Libraries*, JCDL '10, páginas 313–322, New York, NY, USA. ACM.
- [Guo et al., 2009] Guo, L., Tan, E., Chen, S., Zhang, X. e Zhao, Y. E. (2009). Analyzing patterns of user content generation in online social networks. Em *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, páginas 369–378, New York, NY, USA. ACM.
- [Horanyi, 2014] Horanyi, G. (2014). Needle in a haystack. <https://medium.com/@ghoranyi/needle-in-a-haystack-873c97a99983#.yevgl7xo2>. Acessado em 30/11/2016.
- [Horohoe, 2014] Horohoe, C. (2014). Wikimedia moving to elasticsearch. <https://blog.wikimedia.org/2014/01/06/wikimedia-moving-to-elasticsearch/>. Acessado em 30/11/2016.
- [ISKME, 2007] ISKME (2007). Oer commons. <https://www.oercommons.org/>. Acessado em 30/11/2016.
- [Kagolovsky e Moehr, 2003] Kagolovsky, Y. e Moehr, J. R. (2003). Terminological problems in information retrieval. *Journal of Medical Systems*, 27(5):399–408.
- [Lee et al., 1994] Lee, J. H., Kim, M. H. e Lee, Y. J. (1994). Ranking documents in thesaurus-based boolean retrieval systems. *Information Processing Management*, 30(1):79 – 91.
- [LOC e UNESCO, 2009] LOC e UNESCO (2009). Biblioteca digital mundial. <https://www.wdl.org/pt/>. Acessado em 30/11/2016.
- [Loke e Kalantzis, 2014] Loke, S. e Kalantzis, C. (2014). Introducing raigad - an elasticsearch sidecar. <http://techblog.netflix.com/2014/11/introducing-raigad-elasticsearch-sidecar.html>. Acessado em 30/11/2016.
- [Manning et al., 2008] Manning, C. D., Raghavan, P. e Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.
- [MEC, 2004] MEC (2004). Portal domínio público. <http://www.dominiopublico.gov.br/pesquisa/PesquisaObraForm.jsp>. Acessado em 30/11/2016.
- [MEC, 2016] MEC (2016). Base nacional comum curricular - proposta preliminar - segunda versão. <http://basenacionalcomum.mec.gov.br/documentos/bncc-2versao.revista.pdf>. Acessado em 30/11/2016.
- [MEC e MCTI, 2008a] MEC e MCTI (2008a). Banco internacional de objetos educacionais. <http://objetoseducacionais2.mec.gov.br/>. Acessado em 30/11/2016.
- [MEC e MCTI, 2008b] MEC e MCTI (2008b). Portal do professor. <http://portaldoprofessor.mec.gov.br/index.html>. Acessado em 30/11/2016.

- [MIT, 2002] MIT (2002). Mit opencourseware. <https://ocw.mit.edu/>. Acessado em 30/11/2016.
- [Paik, 2013] Paik, J. H. (2013). A novel tf-idf weighting scheme for effective ranking. Em *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '13, páginas 343–352, New York, NY, USA. ACM.
- [Pease, 2013] Pease, T. (2013). A whole new code search. <https://github.com/blog/1381-a-whole-new-code-search>. Acessado em 30/11/2016.
- [Ramos, 2003] Ramos, J. (2003). Using tf-idf to determine word relevance in document queries. Em *Proceedings of the first instructional conference on machine learning*.
- [Salton et al., 1975] Salton, G., Wong, A. e Yang, C. S. (1975). A vector space model for automatic indexing. *Commun. ACM*, 18(11):613–620.
- [Sclater, 2016] Sclater, N. (2016). Moocs, open educational resources and social networking: bridging the gap between informal and formal learning.
- [Seed-PR, 2011] Seed-PR (2011). Portal dia a dia educação. <http://www.diaadia.pr.gov.br/index.php>. Acessado em 30/11/2016.
- [Thet et al., 2007] Thet, T. T., Na, J.-C. e Khoo, C. S. G. (2007). Filtering product reviews from web search results. Em *Proceedings of the 2007 ACM Symposium on Document Engineering*, DocEng '07, páginas 196–198, New York, NY, USA. ACM.
- [UFPR, 2015] UFPR (2015). Ufpr - rea/pea. <http://acervodigital.ufpr.br/handle/1884/35989>. Acessado em 30/11/2016.
- [UMassAmherst e CMU, 2000] UMassAmherst e CMU (2000). Lemur project. <http://www.lemurproject.org/>. Acessado em 30/11/2016.
- [UNESCO, 2012] UNESCO (2012). Declaração rea de paris em 2012. http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/Events/Portuguese_Paris_OER_Declaration.pdf. Acessado em 21/11/2016.
- [Xapian, 2000] Xapian (2000). Xapian. <http://xapian.org/>. Acessado em 30/11/2016.

Apêndice A

Experimentos Realizados

A seguir, cada página demonstrará um experimento com uma *query* de busca e um número específico de usuários criados no processo. A primeira tabela da página demonstra os quinze recursos melhor avaliados encontrados utilizando apenas o *ElasticSearch* (TF-IDF). Na segunda tabela, compara-se em qual colocação estes recursos encontrariam-se com diferentes pesos e mesclas do *Rank*.

Tabela A.1: Recursos importados do PortalMEC - 120 usuários - *query* 'dna' - top 15 resultados de 2826

#	Score	Nome do Recurso
1º	6.026632	DNA
2º	3.4740183	DNA Replication
3º	3.4740183	DNA Transcrição
4º	3.1023014	Stretching DNA
5º	3.1023014	DNA 01
6º	3.1023014	Molécula DNA
7º	3.0942023	DNA: Replicação
8º	3.013316	DNA orbit animated small
9º	3.013316	Do DNA à proteína
10º	2.9690452	DNA: transcrição
11º	2.9690452	DNA Replicação
12º	2.9690452	DNA recombinante
13º	2.9478023	Cell Transcription and Translation
14º	2.7792146	Extração de DNA
15º	2.7792146	DNA simple 2

Tabela A.2: Recursos importados - 120 usuários - *query* 'dna' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.40671	1º	1º	1º	0.325062	1º	1º	1º	0.35343	1º	1º	1º
2º	0.215221	2º	2º	2º	0.177597	3º	2º	2º	0.171963	3º	3º	2º
3º	0.191489	4º	4º	3º	0.16129	5º	4º	3º	0.162162	4º	4º	3º
4º	0.191489	11º	11º	7º	0.16129	12º	12º	7º	0.162162	12º	12º	7º
5º	0.191489	12º	12º	8º	0.16129	13º	13º	8º	0.162162	13º	13º	8º
6º	0.192717	10º	10º	6º	0.162531	11º	11º	6º	0.163202	11º	11º	6º
7º	0.201227	8º	8º	5º	0.172208	10º	8º	5º	0.176715	9º	8º	5º
8º	0.276046	3º	3º	4º	0.237079	2º	3º	4º	0.241877	2º	2º	4º
9º	0.191489	14º	13º	11º	0.16129	14º	14º	12º	0.162162	14º	14º	11º
10º	0.223112	7º	7º	10º	0.190181	9º	7º	11º	0.186368	10º	10º	12º
11º	0.191489	15º	15º	13º	0.16129	15º	15º	13º	0.162162	16º	15º	13º
12º	0.217045	9º	9º	12º	0.196951	6º	6º	10º	0.206712	6º	6º	10º
13º	0.263865	5º	5º	9º	0.225842	4º	5º	9º	0.223196	5º	5º	9º
14º	0.213993	16º	16º	14º	0.176356	17º	16º	14º	0.170924	18º	17º	14º
15º	0.191489	18º	18º	15º	0.16129	18º	18º	15º	0.162162	19º	18º	15º

Tabela A.3: Recursos importados do PortalMEC - 120 usuários - *query* 'molécula' - top 15 resultados de 424

#	Score	Nome do Recurso
1º	3.7604115	Motores moleculares
2º	3.2713356	Molecular Screening
3º	3.2713356	Molécula DNA
4º	3.019493	Molecular Formulas
5º	3.013316	Forma da Molécula
6º	2.7019122	Moléculas e luz
7º	2.7019122	Geometria molecular (2) - Trigonal
8º	2.7019122	Bases moleculares da fenilcetonúria
9º	2.632288	Geometria molecular (4) – Bipiramidal trigonal
10º	2.632288	The Drosophila Molecular Clock Model
11º	2.632288	Biologia molecular e doenças genéticas
12º	2.6170685	Geometria molecular (5) – Octaédrica
13º	2.6170685	Geometria molecular (1) – Linear
14º	2.6170685	Molécula del agua
15º	2.4155946	Geometria molecular (3) - Tetraédrica

Tabela A.4: Recursos importados - 120 usuários - *query* 'molécula' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.191489	2º	2º	1º	0.16129	2º	2º	1º	0.162162	2º	2º	1º
2º	0.305179	1º	1º	2º	0.254697	1º	1º	2º	0.262043	1º	1º	2º
3º	0.192717	4º	4º	3º	0.162531	4º	4º	3º	0.163202	4º	4º	3º
4º	0.191489	7º	7º	4º	0.16129	7º	7º	4º	0.162162	7º	7º	4º
5º	0.191489	8º	8º	5º	0.16129	8º	8º	5º	0.162162	8º	8º	5º
6º	0.222504	9º	9º	6º	0.186033	10º	10º	7º	0.184437	10º	10º	8º
7º	0.191489	14º	14º	9º	0.16129	16º	16º	9º	0.162162	15º	14º	9º
8º	0.211574	10º	10º	7º	0.187274	9º	9º	6º	0.193199	9º	9º	6º
9º	0.211574	12º	12º	11º	0.187274	11º	11º	10º	0.193199	11º	11º	10º
10º	0.215221	11º	11º	10º	0.177597	13º	12º	11º	0.171963	14º	13º	11º
11º	0.192717	17º	17º	14º	0.162531	17º	17º	13º	0.163202	17º	17º	12º
12º	0.263865	6º	5º	8º	0.23045	6º	5º	8º	0.24559	6º	5º	7º
13º	0.213993	13º	13º	12º	0.176356	15º	14º	12º	0.170924	16º	16º	13º
14º	0.192717	18º	18º	15º	0.162531	18º	18º	14º	0.163202	18º	18º	15º
15º	0.191489	19º	19º	17º	0.16129	20º	20º	18º	0.162162	20º	20º	18º

Tabela A.5: Recursos importados do PortalMEC - 500 usuários - *query* 'dna' - top 15 resultados de 2826

#	Score	Nome do Recurso
1º	6.026632	DNA
2º	3.4740183	DNA Replication
3º	3.4740183	DNA Transcrição
4º	3.1023014	Stretching DNA
5º	3.1023014	DNA 01
6º	3.1023014	Molécula DNA
7º	3.0942023	DNA: Replicação
8º	3.013316	DNA orbit animated small
9º	3.013316	Do DNA à proteína
10º	2.9690452	DNA: transcrição
11º	2.9690452	DNA Replicação
12º	2.9690452	DNA recombinante
13º	2.9478023	Cell Transcription and Translation
14º	2.7792146	Extração de DNA
15º	2.7792146	DNA simple 2

Tabela A.6: Recursos importados - 500 usuários - *query* 'dna' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.415503	1º	1º	1º	0.332599	1º	1º	1º	0.360804	1º	1º	1º
2º	0.413801	2º	2º	2º	0.330879	2º	2º	2º	0.359363	2º	2º	2º
3º	0.247212	6º	4º	3º	0.214435	5º	4º	3º	0.223818	5º	4º	3º
4º	0.193191	18º	18º	10º	0.163011	18º	18º	9º	0.163604	18º	18º	9º
5º	0.198165	16º	16º	7º	0.170866	16º	16º	7º	0.173344	16º	16º	7º
6º	0.197145	17º	17º	8º	0.165272	17º	17º	8º	0.164627	17º	17º	8º
7º	0.301497	8º	8º	4º	0.245191	8º	8º	4º	0.251556	8º	8º	5º
8º	0.197996	19º	19º	14º	0.166133	19º	19º	14º	0.165347	20º	20º	14º
9º	0.326561	7º	6º	6º	0.279717	6º	6º	6º	0.290318	7º	6º	6º
10º	0.227704	15º	15º	13º	0.201035	15º	14º	12º	0.203026	15º	15º	13º
11º	0.377742	3º	3º	5º	0.309932	3º	3º	5º	0.334161	3º	3º	4º
12º	0.197314	20º	20º	15º	0.170006	20º	20º	15º	0.172623	19º	19º	15º
13º	0.29171	9º	9º	9º	0.243747	9º	9º	10º	0.250733	10º	9º	11º
14º	0.390318	5º	7º	12º	0.319699	7º	7º	13º	0.341829	6º	7º	12º
15º	0.397515	4º	5º	11º	0.340462	4º	5º	11º	0.368514	4º	5º	10º

Tabela A.7: Recursos importados do PortalMEC - 500 usuários - *query* 'molécula' - top 15 resultados de 424

#	Score	Nome do Recurso
1º	3.7604115	Motores moleculares
2º	3.2713356	Molecular Screening
3º	3.2713356	Molécula DNA
4º	3.019493	Molecular Formulas
5º	3.013316	Forma da Molécula
6º	2.7019122	Bases moleculares da fenilcetonúria
7º	2.7019122	Moléculas e luz
8º	2.7019122	Geometria molecular (2) - Trigonal
9º	2.632288	Geometria molecular (4) – Bipiramidal trigonal
10º	2.632288	The Drosophila Molecular Clock Model
11º	2.632288	Biologia molecular e doenças genéticas
12º	2.6170685	Geometria molecular (5) – Octaédrica
13º	2.6170685	Geometria molecular (1) – Linear
14º	2.6170685	Molécula del agua
15º	2.4155946	Polaridade da molécula

Tabela A.8: Recursos importados - 500 usuários - *query* 'molécula' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.193191	5º	3º	1º	0.163011	4º	3º	1º	0.163604	5º	4º	1º
2º	0.383642	1º	1º	2º	0.310124	1º	1º	2º	0.330647	1º	1º	2º
3º	0.197145	13º	11º	4º	0.165272	13º	11º	4º	0.164627	14º	13º	4º
4º	0.2028	16º	15º	5º	0.169254	15º	15º	5º	0.167091	16º	15º	5º
5º	0.436347	2º	2º	3º	0.355331	2º	2º	3º	0.383195	2º	2º	3º
6º	0.195126	22º	21º	13º	0.16741	21º	21º	13º	0.168518	22º	21º	13º
7º	0.323131	10º	8º	8º	0.263241	10º	8º	8º	0.280978	10º	9º	8º
8º	0.203315	19º	19º	12º	0.173045	20º	19º	12º	0.174035	20º	19º	12º
9º	0.375293	6º	6º	9º	0.301743	7º	7º	10º	0.322303	7º	7º	10º
10º	0.35976	7º	7º	10º	0.303929	6º	6º	9º	0.32386	6º	6º	9º
11º	0.395683	4º	5º	7º	0.329577	5º	5º	7º	0.347519	4º	5º	7º
12º	0.426801	3º	4º	6º	0.347032	3º	4º	6º	0.375184	3º	3º	6º
13º	0.213184	20º	20º	17º	0.184882	19º	20º	14º	0.185274	21º	20º	16º
14º	0.294865	14º	14º	11º	0.244458	14º	14º	11º	0.257419	13º	14º	11º
15º	0.407136	9º	10º	15º	0.331764	9º	10º	16º	0.357951	9º	10º	15º

Tabela A.9: Recursos importados do PortalMEC - 1000 usuários - *query* 'dna' - top 15 resultados de 2826

#	Score	Nome do Recurso
1º	6.026632	DNA
2º	3.4740183	DNA Replication
3º	3.4740183	DNA Transcrição
4º	3.1023014	Stretching DNA
5º	3.1023014	DNA 01
6º	3.1023014	Molécula DNA
7º	3.0942023	DNA: Replicação
8º	3.013316	DNA orbit animated small
9º	3.013316	Do DNA à proteína
10º	2.9690452	DNA: transcrição
11º	2.9690452	DNA Replicação
12º	2.9690452	DNA recombinante
13º	2.9478023	Cell Transcription and Translation
14º	2.7792146	Extração de DNA
15º	2.7792146	DNA simple 2

Tabela A.10: Recursos importados - 1000 usuários - *query* 'dna' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.425612	1º	1º	1º	0.348263	1º	1º	1º	0.378507	1º	1º	1º
2º	0.334863	2º	2º	2º	0.2763	2º	2º	2º	0.294265	2º	2º	2º
3º	0.224362	9º	9º	3º	0.19463	9º	8º	3º	0.193731	9º	9º	3º
4º	0.210216	19º	17º	11º	0.177418	18º	17º	12º	0.176395	20º	18º	12º
5º	0.317705	6º	5º	4º	0.262724	6º	6º	4º	0.279637	6º	6º	4º
6º	0.206504	20º	18º	12º	0.178006	17º	16º	11º	0.179327	19º	17º	11º
7º	0.193224	21º	21º	13º	0.164625	21º	20º	13º	0.165751	23º	21º	13º
8º	0.274219	11º	10º	10º	0.232714	11º	10º	10º	0.242326	11º	10º	10º
9º	0.338229	5º	6º	6º	0.304953	4º	4º	6º	0.321015	4º	4º	6º
10º	0.402017	3º	3º	5º	0.335574	3º	3º	5º	0.358883	3º	3º	5º
11º	0.329402	8º	8º	8º	0.270195	8º	9º	9º	0.288856	8º	8º	8º
12º	0.359078	4º	4º	7º	0.294758	5º	5º	7º	0.3132	5º	5º	7º
13º	0.343048	7º	7º	9º	0.285381	7º	7º	8º	0.299681	7º	7º	9º
14º	0.205351	25º	25º	16º	0.176634	25º	25º	16º	0.177138	25º	25º	16º
15º	0.236274	24º	22º	15º	0.203855	22º	22º	14º	0.198174	24º	24º	15º

Tabela A.11: Recursos importados do PortalMEC - 1000 usuários - *query* 'molécula' - top 15 resultados de 424

#	Score	Nome do Recurso
1º	3.7604115	Motores moleculares
2º	3.2713356	Molecular Screening
3º	3.2713356	Molécula DNA
4º	3.019493	Molecular Formulas
5º	3.013316	Forma da Molécula
6º	2.7019122	Bases moleculares da fenilcetonúria
7º	2.7019122	Geometria molecular (2) - Trigonal
8º	2.7019122	Moléculas e luz
9º	2.632288	Geometria molecular (4) – Bipiramidal trigonal
10º	2.632288	The Drosophila Molecular Clock Model
11º	2.632288	Biologia molecular e doenças genéticas
12º	2.6170685	Geometria molecular (5) – Octaédrica
13º	2.6170685	Molécula del agua
14º	2.6170685	Geometria molecular (1) – Linear
15º	2.4155946	Polaridade da molécula

Tabela A.12: Recursos importados - 1000 usuários - *query* 'molécula' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.258187	3º	3º	1º	0.226792	2º	2º	1º	0.235876	3º	2º	1º
2º	0.440441	1º	1º	2º	0.364584	1º	1º	2º	0.394302	1º	1º	2º
3º	0.206504	12º	12º	5º	0.178006	12º	12º	4º	0.179327	13º	12º	5º
4º	0.374104	4º	4º	4º	0.311296	4º	4º	5º	0.33163	4º	4º	4º
5º	0.43073	2º	2º	3º	0.349474	3º	3º	3º	0.376337	2º	3º	3º
6º	0.388462	7º	7º	7º	0.323345	7º	7º	7º	0.338556	7º	7º	7º
7º	0.198201	25º	24º	14º	0.170241	24º	23º	14º	0.170751	25º	24º	14º
8º	0.40015	6º	6º	6º	0.333652	5º	5º	6º	0.354448	6º	5º	6º
9º	0.377899	9º	9º	9º	0.309047	9º	9º	9º	0.332513	9º	9º	9º
10º	0.429056	5º	5º	8º	0.348217	6º	6º	8º	0.37618	5º	6º	8º
11º	0.345442	11º	11º	12º	0.299171	11º	11º	11º	0.309556	11º	11º	12º
12º	0.386118	8º	8º	10º	0.316944	8º	8º	10º	0.338923	8º	8º	10º
13º	0.362	10º	10º	11º	0.304444	10º	10º	12º	0.326253	10º	10º	11º
14º	0.293332	18º	16º	13º	0.242499	17º	17º	13º	0.25384	18º	17º	13º
15º	0.371798	14º	13º	15º	0.312705	14º	13º	15º	0.328394	14º	14º	15º

Tabela A.13: Recursos importados do PortalMEC - 5000 usuários - *query* 'dna' - top 15 resultados de 2826

#	Score	Nome do Recurso
1º	6.026632	DNA
2º	3.4740183	DNA Replication
3º	3.4740183	DNA Transcrição
4º	3.1023014	Stretching DNA
5º	3.1023014	DNA 01
6º	3.1023014	Molécula DNA
7º	3.0942023	DNA: Replicação
8º	3.013316	DNA orbit animated small
9º	3.013316	Do DNA à proteína
10º	2.9690452	DNA: transcrição
11º	2.9690452	DNA Replicação
12º	2.9690452	DNA recombinante
13º	2.9478023	Cell Transcription and Translation
14º	2.7792146	Extração de DNA
15º	2.7792146	DNA simple 2

Tabela A.14: Recursos importados - 5000 usuários - *query* 'dna' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.379379	1º	1º	1º	0.310194	1º	1º	1º	0.334608	1º	1º	1º
2º	0.288179	4º	3º	3º	0.238755	4º	3º	3º	0.250938	4º	3º	3º
3º	0.328358	2º	2º	2º	0.270895	2º	2º	2º	0.28533	2º	2º	2º
4º	0.366615	3º	4º	4º	0.30123	3º	4º	4º	0.322474	3º	4º	4º
5º	0.344913	7º	7º	6º	0.284548	7º	7º	6º	0.302163	7º	7º	6º
6º	0.312076	12º	11º	8º	0.260128	11º	11º	7º	0.274246	12º	11º	8º
7º	0.358449	5º	5º	5º	0.294692	6º	6º	5º	0.314158	6º	5º	5º
8º	0.37414	6º	6º	7º	0.313308	5º	5º	8º	0.332802	5º	6º	7º
9º	0.348072	9º	9º	9º	0.290035	9º	9º	9º	0.306632	9º	9º	9º
10º	0.364298	8º	8º	10º	0.30317	8º	8º	10º	0.323837	8º	8º	10º
11º	0.356292	10º	10º	11º	0.291937	10º	10º	11º	0.311411	10º	10º	11º
12º	0.300178	14º	14º	13º	0.24518	15º	14º	13º	0.25925	15º	15º	13º
13º	0.347757	11º	12º	12º	0.285686	12º	12º	12º	0.303952	11º	12º	12º
14º	0.341981	15º	15º	15º	0.282109	14º	15º	15º	0.302155	14º	14º	15º
15º	0.363294	13º	13º	14º	0.307155	13º	13º	14º	0.326098	13º	13º	14º

Tabela A.15: Recursos importados do PortalMEC - 5000 usuários - *query* 'molécula' - top 15 resultados de 424

#	Score	Nome do Recurso
1º	3.7604115	Motores moleculares
2º	3.2713356	Molecular Screening
3º	3.2713356	Molécula DNA
4º	3.019493	Molecular Formulas
5º	3.013316	Forma da Molécula
6º	2.7019122	Bases moleculares da fenilcetonúria
7º	2.7019122	Moléculas e luz
8º	2.7019122	Geometria molecular (2) - Trigonal
9º	2.632288	The Drosophila Molecular Clock Model
10º	2.632288	Geometria molecular (4) – Bipiramidal trigonal
11º	2.632288	Biologia molecular e doenças genéticas
12º	2.6170685	Geometria molecular (5) – Octaédrica
13º	2.6170685	Geometria molecular (1) – Linear
14º	2.6170685	Molécula del agua
15º	2.4155946	Polaridade da molécula

Tabela A.16: Recursos importados - 5000 usuários - *query* 'molécula' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.329572	1º	1º	1º	0.271253	1º	1º	1º	0.287908	1º	1º	1º
2º	0.348724	2º	2º	2º	0.296486	2º	2º	2º	0.312932	2º	2º	2º
3º	0.312076	3º	3º	3º	0.260128	3º	3º	3º	0.274246	3º	3º	3º
4º	0.317246	5º	5º	5º	0.261297	5º	5º	5º	0.275094	5º	5º	5º
5º	0.3572	4º	4º	4º	0.292614	4º	4º	4º	0.311833	4º	4º	4º
6º	0.344227	8º	8º	6º	0.283718	8º	8º	6º	0.303411	8º	8º	6º
7º	0.328525	10º	10º	7º	0.274392	10º	10º	7º	0.291283	10º	10º	7º
8º	0.309871	13º	12º	10º	0.255647	13º	12º	10º	0.271111	13º	12º	10º
9º	0.368351	7º	7º	9º	0.308835	7º	7º	9º	0.326691	7º	7º	9º
10º	0.341801	11º	11º	12º	0.279902	11º	11º	12º	0.298023	11º	11º	12º
11º	0.352275	9º	9º	11º	0.297595	9º	9º	11º	0.314087	9º	9º	11º
12º	0.330672	12º	13º	13º	0.273122	12º	13º	13º	0.290502	12º	13º	13º
13º	0.316978	14º	14º	14º	0.260442	14º	14º	14º	0.27663	14º	14º	14º
14º	0.389505	6º	6º	8º	0.325372	6º	6º	8º	0.346725	6º	6º	8º
15º	0.353599	17º	17º	17º	0.289497	17º	17º	17º	0.309443	17º	17º	17º

Tabela A.17: Recursos importados do PortalMEC - 10000 usuários - *query* 'dna' - top 15 resultados de 2826

#	Score	Nome do Recurso
1º	6.026632	DNA
2º	3.4740183	DNA Replication
3º	3.4740183	DNA Transcrição
4º	3.1023014	DNA 01
5º	3.1023014	Stretching DNA
6º	3.1023014	Molécula DNA
7º	3.0942023	DNA: Replicação
8º	3.013316	DNA orbit animated small
9º	3.013316	Do DNA à proteína
10º	2.9690452	DNA: transcrição
11º	2.9690452	DNA Replicação
12º	2.9690452	DNA recombinante
13º	2.9478023	Cell Transcription and Translation
14º	2.7792146	Extração de DNA
15º	2.7792146	DNA simple 2

Tabela A.18: Recursos importados - 10000 usuários - *query* 'dna' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.331761	1º	1º	1º	0.273323	1º	1º	1º	0.290445	1º	1º	1º
2º	0.339569	3º	3º	3º	0.279005	3º	3º	3º	0.297911	3º	3º	3º
3º	0.361392	2º	2º	2º	0.299979	2º	2º	2º	0.320494	2º	2º	2º
4º	0.343206	6º	5º	4º	0.282156	6º	6º	4º	0.30103	6º	6º	4º
5º	0.333456	7º	7º	5º	0.274483	7º	7º	5º	0.291874	7º	7º	5º
6º	0.313702	10º	10º	6º	0.259358	10º	10º	6º	0.274788	10º	10º	7º
7º	0.307355	12º	11º	9º	0.25536	11º	11º	9º	0.269145	12º	11º	9º
8º	0.370813	4º	4º	7º	0.313586	4º	4º	7º	0.333697	4º	4º	6º
9º	0.364236	5º	6º	8º	0.308268	5º	5º	8º	0.32822	5º	5º	8º
10º	0.35786	8º	8º	10º	0.298768	8º	8º	10º	0.317252	8º	8º	10º
11º	0.355421	9º	9º	11º	0.294243	9º	9º	11º	0.314464	9º	9º	11º
12º	0.335046	11º	12º	12º	0.276089	12º	12º	12º	0.293854	11º	12º	12º
13º	0.299691	14º	13º	13º	0.249893	14º	14º	13º	0.263446	14º	14º	13º
14º	0.329465	15º	15º	15º	0.274427	15º	15º	15º	0.290829	15º	15º	15º
15º	0.342661	13º	14º	14º	0.288178	13º	13º	14º	0.305351	13º	13º	14º

Tabela A.19: Recursos importados do PortalMEC - 10000 usuários - *query* 'molécula' - top 15 resultados de 424

#	Score	Nome do Recurso
1º	3.7604115	Motores moleculares
2º	3.2713356	Molecular Screening
3º	3.2713356	Molécula DNA
4º	3.019493	Molecular Formulas
5º	3.013316	Forma da Molécula
6º	2.7019122	Geometria molecular (2) - Trigonal
7º	2.7019122	Moléculas e luz
8º	2.7019122	Bases moleculares da fenilcetonúria
9º	2.632288	Geometria molecular (4) – Bipiramidal trigonal
10º	2.632288	The Drosophila Molecular Clock Model
11º	2.632288	Biologia molecular e doenças genéticas
12º	2.6170685	Geometria molecular (5) – Octaédrica
13º	2.6170685	Geometria molecular (1) – Linear
14º	2.6170685	Molécula del agua
15º	2.4155946	Geometria molecular (3) - Tetraédrica

Tabela A.20: Recursos importados - 10000 usuários - *query* 'molécula' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.354577	1º	1º	1º	0.294146	1º	1º	1º	0.314137	1º	1º	1º
2º	0.367391	2º	2º	2º	0.309962	2º	2º	2º	0.329167	2º	2º	2º
3º	0.313702	3º	3º	3º	0.259358	3º	3º	3º	0.274788	3º	3º	3º
4º	0.351513	4º	4º	4º	0.289007	4º	4º	4º	0.309193	4º	4º	4º
5º	0.325978	5º	5º	5º	0.271006	5º	5º	5º	0.286838	5º	5º	5º
6º	0.344847	6º	6º	6º	0.285731	6º	6º	6º	0.304806	6º	6º	6º
7º	0.313995	13º	12º	8º	0.260794	12º	12º	8º	0.275221	13º	12º	8º
8º	0.332265	10º	8º	7º	0.275537	11º	10º	7º	0.292535	10º	9º	7º
9º	0.308507	14º	14º	14º	0.25793	14º	14º	14º	0.271481	14º	14º	14º
10º	0.344006	11º	11º	12º	0.290705	10º	11º	10º	0.30699	11º	11º	11º
11º	0.351188	9º	10º	9º	0.293629	9º	9º	9º	0.309825	9º	10º	9º
12º	0.361324	7º	7º	10º	0.299348	7º	7º	11º	0.320236	7º	7º	10º
13º	0.335029	12º	13º	13º	0.277977	13º	13º	13º	0.295405	12º	13º	13º
14º	0.357679	8º	9º	11º	0.298893	8º	8º	12º	0.317746	8º	8º	12º
15º	0.319166	20º	19º	17º	0.263951	20º	19º	17º	0.279861	20º	20º	17º

Tabela A.21: Recursos gerados pelo populador - 120 usuários - *query* 'sint' - top 15 resultados de 398

#	Score	Nome do Recurso
1º	1.4389775	Ut aut sint occaecati.
2º	1.4389775	Vel sunt sint perspiciatis.
3º	1.2609439	Molestiae quod aperiam sint deserunt.
4º	1.2591053	Eos corrupti aperiam quos sint.
5º	1.1928575	Magnam occaecati ratione sint.
6º	1.134607	Eum perferendis aut velit quia minima sint.
7º	1.134607	Voluptatem est sint vel quas delectus.
8º	1.134607	Cupiditate sint tempore sit consequuntur a dignissimos.
9º	1.134607	Et enim magni non possimus ratione sint.
10º	1.0808091	Autem assumenda sint non quam possimus.
11º	1.0808091	Et accusamus qui quia recusandae sint quia.
12º	1.0808091	Aut sint facere doloremque est hic eius.
13º	1.0792332	Sint incidunt consequatur dolores beatae id.
14º	1.0622671	Culpa rerum deleniti sint.
15º	1.054347	Voluptatem odio sint sint adipisci officia pariatur totam.

Tabela A.22: Recursos gerados - 120 usuários - *query* 'sint' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.506481	1º	1º	1º	0.493303	1º	1º	1º	0.51404	1º	1º	1º
2º	0.191055	7º	7º	2º	0.155383	9º	6º	2º	0.169796	8º	7º	2º
3º	0.109892	29º	28º	7º	0.101869	26º	25º	6º	0.099067	27º	27º	7º
4º	0.304999	4º	4º	3º	0.251214	4º	4º	3º	0.26813	4º	4º	4º
5º	0.291708	6º	6º	5º	0.230558	8º	8º	5º	0.265232	6º	6º	5º
6º	0.491695	2º	2º	4º	0.47345	2º	2º	4º	0.5282	2º	2º	3º
7º	0.298083	9º	8º	8º	0.237568	10º	10º	8º	0.269683	10º	9º	8º
8º	0.287955	11º	10º	9º	0.233949	11º	11º	9º	0.2508	11º	11º	9º
9º	0.124636	32º	30º	14º	0.113903	31º	29º	13º	0.110553	32º	30º	14º
10º	0.231177	17º	16º	13º	0.177789	20º	19º	14º	0.205904	17º	17º	13º
11º	0.162989	27º	26º	15º	0.133921	27º	27º	15º	0.151873	26º	25º	15º
12º	0.150849	28º	29º	16º	0.130173	28º	28º	16º	0.131421	30º	28º	16º
13º	0.103456	37º	36º	17º	0.092175	36º	35º	17º	0.088062	37º	37º	17º
14º	0.104088	38º	38º	18º	0.092813	37º	36º	18º	0.088597	38º	38º	18º
15º	0.520451	3º	3º	6º	0.474748	3º	3º	7º	0.507875	3º	3º	6º

Tabela A.23: Recursos gerados pelo populador - 120 usuários - *query* 'velit' - top 15 resultados de 306

#	Score	Nome do Recurso
1º	1.6412256	Impedit iste velit voluptates.
2º	1.6412256	Inventore asperiores labore velit.
3º	1.4360723	Velit non totam ipsa veniam.
4º	1.2485772	Velit quis mollitia dolorem.
5º	1.2309191	Voluptatem velit eos inventore doloremque quos nisi.
6º	1.1147532	Doloremque soluta sed velit.
7º	1.0925051	Velit incidunt minima magni non.
8º	1.0785924	Perferendis fugit ut omnis velit.
9º	1.025766	Veniam accusantium a nesciunt repudiandae sequi aut beatae velit.
10º	0.9579561	Qui impedit eveniet velit.
11º	0.9579561	Nam blanditiis velit ratione.
12º	0.93643296	Repellendus velit aut omnis rerum ea.
13º	0.93643296	Nulla velit mollitia debitis id accusantium veritatis.
14º	0.93643296	Dolorem ut velit itaque ducimus exercitationem.
15º	0.93643296	Accusantium adipisci tenetur et velit et quo.

Tabela A.24: Recursos gerados - 120 usuários - *query* 'velit' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.302685	1º	1º	1º	0.267365	1º	1º	1º	0.283417	1º	1º	1º
2º	0.104088	13º	11º	2º	0.092813	11º	10º	2º	0.088597	15º	13º	2º
3º	0.104088	19º	18º	4º	0.092813	20º	18º	4º	0.088597	21º	20º	4º
4º	0.415664	2º	2º	3º	0.373261	2º	2º	3º	0.383986	2º	2º	3º
5º	0.291772	4º	4º	5º	0.239149	4º	4º	5º	0.269463	4º	4º	5º
6º	0.214781	16º	13º	7º	0.183512	16º	15º	8º	0.189785	16º	15º	8º
7º	0.293853	6º	5º	6º	0.258455	7º	5º	6º	0.267996	7º	7º	6º
8º	0.114443	39º	39º	10º	0.105054	38º	36º	9º	0.101025	39º	39º	9º
9º	0.109687	43º	40º	17º	0.097907	42º	40º	16º	0.092865	44º	42º	17º
10º	0.356197	8º	7º	9º	0.299005	9º	9º	11º	0.315257	9º	9º	11º
11º	0.286901	17º	17º	14º	0.242515	17º	17º	14º	0.253503	17º	17º	14º
12º	0.601486	3º	3º	8º	0.593253	3º	3º	7º	0.6191	3º	3º	7º
13º	0.317055	14º	15º	15º	0.278267	13º	12º	15º	0.285325	14º	14º	16º
14º	0.357194	10º	10º	12º	0.307481	10º	11º	12º	0.323865	10º	10º	13º
15º	0.225387	22º	22º	19º	0.201082	22º	22º	19º	0.203824	23º	22º	19º

Tabela A.25: Recursos gerados pelo populador - 500 usuários - *query* 'sint' - top 15 resultados de 935

#	Score	Nome do Recurso
1º	1.3831264	Eaque sint pariatur sint.
2º	1.2416571	Sint sint vero facilis cupiditate ut.
3º	1.1706456	Quibusdam itaque id sint.
4º	1.1706456	Sequi tempora corporis sint.
5º	1.1706456	Est veniam sint libero.
6º	1.1706456	Quibusdam et occaecati sint.
7º	1.1281147	Omnis sint dolore rerum.
8º	1.114141	Impedit qui ratione sint.
9º	0.98710036	Velit sint error ut nesciunt.
10º	0.97801805	Sint et nihil aliquam.
11º	0.97801805	Mollitia adipisci perspiciatis sint.
12º	0.97487336	Nobis autem aut ut sint.
13º	0.9221362	Sint sed aspernatur exercitationem suscipit.
14º	0.9221362	Adipisci illo assumenda corporis sint.
15º	0.9221362	Libero repellendus sint temporibus voluptatem.

Tabela A.26: Recursos gerados - 500 usuários - *query* 'sint' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.297524	2º	2º	1º	0.255307	3º	2º	1º	0.270232	3º	2º	1º
2º	0.228938	9º	9º	3º	0.196036	9º	9º	3º	0.200995	10º	9º	3º
3º	0.46565	1º	1º	2º	0.44865	1º	1º	2º	0.462099	1º	1º	2º
4º	0.342675	4º	4º	4º	0.300613	4º	4º	4º	0.317721	4º	4º	4º
5º	0.307792	5º	5º	5º	0.263227	5º	5º	5º	0.276245	5º	5º	5º
6º	0.277629	8º	7º	6º	0.23199	8º	8º	6º	0.24732	8º	8º	6º
7º	0.307254	6º	6º	7º	0.261893	7º	6º	7º	0.280727	7º	6º	7º
8º	0.274113	11º	10º	8º	0.227717	11º	10º	9º	0.242206	11º	10º	9º
9º	0.26377	22º	20º	11º	0.220062	22º	21º	11º	0.230406	22º	22º	11º
10º	0.562642	3º	3º	9º	0.558707	2º	3º	8º	0.577864	2º	3º	8º
11º	0.178791	56º	49º	14º	0.159244	54º	49º	14º	0.160707	57º	50º	14º
12º	0.400354	7º	8º	10º	0.35732	6º	7º	10º	0.375967	6º	7º	10º
13º	0.361253	13º	13º	12º	0.32363	12º	12º	12º	0.334051	14º	12º	12º
14º	0.280451	28º	26º	17º	0.249604	24º	22º	15º	0.25835	26º	24º	16º
15º	0.286837	25º	24º	15º	0.228438	36º	29º	16º	0.25855	25º	23º	15º

Tabela A.27: Recursos gerados pelo populador - 500 usuários - *query* 'velit' - top 15 resultados de 721

#	Score	Nome do Recurso
1º	1.3269383	Accusamus velit ut quia non.
2º	1.1960292	Repellendus magnam velit voluptates.
3º	1.1960292	Et eveniet velit alias.
4º	1.1373756	Explicabo velit illum quae cupiditate officiis natus.
5º	1.1373756	Velit ut consequatur et quia sapiente.
6º	1.1373756	Velit accusamus rem laudantium nisi dolor.
7º	1.1373756	Corrupti saepe velit quas perspiciatis quis animi.
8º	1.1373756	Fugit velit rerum ratione corporis quisquam.
9º	1.1177979	Amet velit dignissimos velit deleniti rerum.
10º	1.05387	Officiis velit error deserunt.
11º	1.0466002	Velit inventore reprehenderit alias.
12º	1.0466002	Totam quo sunt velit.
13º	1.0466002	Perspiciatis aut ducimus velit.
14º	1.0466002	Velit facere qui ut.
15º	1.0465255	Ut officiis velit quaerat illo.

Tabela A.28: Recursos gerados - 500 usuários - *query* 'velit' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.479783	1º	1º	1º	0.427219	1º	1º	1º	0.464303	1º	1º	1º
2º	0.325871	4º	4º	4º	0.27749	5º	5º	2º	0.297224	4º	4º	2º
3º	0.103456	73º	66º	8º	0.092175	69º	62º	8º	0.088062	76º	72º	8º
4º	0.460179	2º	2º	2º	0.35719	2º	2º	3º	0.389518	2º	2º	3º
5º	0.453364	3º	3º	3º	0.354124	3º	3º	4º	0.386062	3º	3º	4º
6º	0.358989	5º	5º	5º	0.314819	4º	4º	5º	0.324924	5º	5º	5º
7º	0.309709	6º	6º	6º	0.281188	7º	6º	6º	0.29538	6º	6º	6º
8º	0.257922	11º	9º	7º	0.221236	14º	11º	7º	0.229472	14º	10º	7º
9º	0.173576	41º	37º	10º	0.151972	40º	39º	10º	0.151352	43º	40º	10º
10º	0.250046	24º	18º	13º	0.222887	22º	20º	11º	0.225892	23º	21º	13º
11º	0.243882	26º	25º	14º	0.20797	27º	24º	14º	0.218522	28º	24º	14º
12º	0.140216	70º	64º	17º	0.133271	60º	53º	16º	0.128474	69º	61º	17º
13º	0.103456	88º	84º	20º	0.092175	87º	80º	19º	0.088062	91º	87º	19º
14º	0.103456	89º	85º	21º	0.092175	88º	81º	20º	0.088062	92º	88º	20º
15º	0.2784	15º	13º	11º	0.233544	18º	16º	12º	0.246875	17º	15º	11º

Tabela A.29: Recursos gerados pelo populador - 1000 usuários - *query* 'sint' - top 15 resultados de 1641

#	Score	Nome do Recurso
1º	1.0982507	Sit nihil voluptatem sint eos sint.
2º	1.0354408	Pariatur nihil sint nisi.
3º	1.0354408	Sint voluptas et modi.
4º	1.0354408	Incidunt sed sint aut.
5º	0.9505286	Sint accusantium beatae minus.
6º	0.9505286	Nisi quo sint accusamus.
7º	0.9505286	Sint eos nihil necessitatibus.
8º	0.9505286	Sint iure quibusdam culpa.
9º	0.9505286	Sint consecetur sequi quisquam.
10º	0.9505286	Hic quod sint molestiae.
11º	0.9505286	Est vitae incidunt sint.
12º	0.94290537	Sit adipisci sint dicta.
13º	0.94290537	Aspernatur possimus porro sint.
14º	0.94290537	Eos sint quis laboriosam.
15º	0.94290537	Sunt consequuntur sint quis.

Tabela A.30: Recursos gerados - 1000 usuários - *query* 'sint' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.28939	6º	5º	1º	0.241011	7º	6º	1º	0.256854	7º	6º	1º
2º	0.362026	1º	1º	2º	0.327948	3º	1º	2º	0.341019	3º	1º	2º
3º	0.333785	5º	4º	3º	0.28107	6º	5º	3º	0.302344	5º	5º	3º
4º	0.153609	81º	65º	8º	0.132756	84º	70º	7º	0.134383	91º	74º	8º
5º	0.304609	15º	13º	12º	0.258959	16º	16º	12º	0.274246	16º	15º	11º
6º	0.306773	12º	12º	10º	0.253054	20º	18º	13º	0.261337	23º	21º	13º
7º	0.356347	8º	8º	5º	0.313485	8º	8º	6º	0.329836	8º	8º	6º
8º	0.259152	33º	28º	18º	0.218137	39º	31º	17º	0.231173	37º	32º	17º
9º	0.314065	11º	11º	9º	0.263867	15º	14º	10º	0.276711	15º	14º	10º
10º	0.211994	56º	52º	25º	0.182527	62º	54º	24º	0.189261	64º	55º	25º
11º	0.241699	42º	36º	22º	0.200128	52º	46º	21º	0.212512	52º	44º	23º
12º	0.417126	3º	3º	4º	0.382391	4º	4º	4º	0.396971	4º	4º	4º
13º	0.307047	17º	15º	13º	0.279785	14º	10º	11º	0.288087	14º	10º	12º
14º	0.374378	7º	7º	6º	0.312536	9º	9º	8º	0.33079	9º	9º	7º
15º	0.322841	10º	10º	11º	0.26218	17º	17º	14º	0.276934	17º	17º	14º

Tabela A.31: Recursos gerados pelo populador - 1000 usuários - *query* 'velit' - top 15 resultados de 1179

#	Score	Nome do Recurso
1º	1.0763311	Velit dolorem quis labore.
2º	1.0763311	Velit maxime enim ducimus.
3º	1.0763311	Qui aperiam facilis velit.
4º	1.0763311	Quia velit dolorum sequi.
5º	1.0763311	Illo quisquam velit enim.
6º	1.0763311	Velit quidem iure adipisci.
7º	1.0763311	Velit sequi provident sint.
8º	1.0354408	Officiis velit quae dolorum.
9º	1.0354408	Velit est provident voluptates.
10º	1.0354408	Accusamus dicta velit ex.
11º	1.0354408	Dolores velit ea totam.
12º	1.019747	Ab aut velit consectetur.
13º	1.019747	Debitis non velit sed.
14º	1.019747	Cupiditate velit omnis quos.
15º	0.9871445	Quisquam accusamus velit quam.

Tabela A.32: Recursos gerados - 1000 usuários - *query* 'velit' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.389559	3º	3º	1º	0.335513	5º	3º	2º	0.35085	5º	3º	2º
2º	0.346345	7º	5º	4º	0.312045	7º	7º	3º	0.325195	8º	7º	3º
3º	0.234432	25º	20º	9º	0.206541	28º	25º	9º	0.213407	29º	26º	9º
4º	0.297205	10º	10º	6º	0.259742	12º	12º	6º	0.268036	13º	11º	6º
5º	0.269315	15º	14º	7º	0.233839	18º	16º	7º	0.238971	18º	16º	8º
6º	0.148022	86º	76º	16º	0.113579	109º	93º	17º	0.110398	119º	111º	17º
7º	0.103456	149º	129º	20º	0.092175	141º	125º	18º	0.088062	160º	146º	20º
8º	0.460837	1º	1º	2º	0.445485	1º	1º	1º	0.45069	1º	1º	1º
9º	0.43702	2º	2º	3º	0.390266	2º	2º	4º	0.405034	2º	2º	4º
10º	0.2624	19º	18º	11º	0.226329	27º	23º	13º	0.238651	27º	21º	13º
11º	0.232963	33º	31º	15º	0.201394	36º	32º	14º	0.208654	40º	32º	14º
12º	0.365047	8º	8º	8º	0.347193	8º	8º	8º	0.363159	7º	8º	7º
13º	0.428407	4º	4º	5º	0.378283	4º	4º	5º	0.395474	4º	4º	5º
14º	0.330555	11º	11º	10º	0.298106	11º	10º	10º	0.305811	11º	10º	10º
15º	0.28148	24º	22º	18º	0.257675	21º	19º	19º	0.273196	20º	18º	18º

Tabela A.33: Recursos gerados pelo populador - 5000 usuários - *query* 'sint' - top 15 resultados de 8270

#	Score	Nome do Recurso
1º	1.1309981	Sint maiores sint vel.
2º	0.893759	Qui sint sint quibusdam expedita et.
3º	0.89373934	Sint sint adipisci eum molestias ea quia.
4º	0.89373934	Porro nisi sint sint quas cumque veritatis.
5º	0.89373934	Et sint consequatur sint est architecto.
6º	0.8794402	Sint nisi sint reiciendis voluptatem aut.
7º	0.8794402	Qui ut harum sint sint cum.
8º	0.85630965	Sint aut sint inventore quaerat dolor eum.
9º	0.8482486	Sint corrupti non sint est possimus et.
10º	0.8482486	Ullam sint consequatur reiciendis ipsa sit sint.
11º	0.8426441	Officia ut aut sint.
12º	0.8426441	Dolorum ea sint impedit.
13º	0.8426441	Aliquid veniam sint harum.
14º	0.8426441	Perspiciatis commodi sint placeat.
15º	0.8426441	Amet atque sint ipsa.

Tabela A.34: Recursos gerados - 5000 usuários - *query* 'sint' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.26266	2º	1º	1º	0.223955	5º	2º	1º	0.232943	5º	2º	1º
2º	0.259562	76º	66º	12º	0.223215	81º	72º	10º	0.237129	82º	69º	11º
3º	0.248705	87º	78º	15º	0.207855	102º	89º	16º	0.221703	103º	91º	17º
4º	0.263035	70º	60º	11º	0.202533	115º	98º	20º	0.212725	118º	103º	21º
5º	0.323879	17º	15º	2º	0.272172	33º	25º	5º	0.288741	30º	20º	5º
6º	0.202155	192º	170º	44º	0.160757	253º	215º	49º	0.184761	204º	177º	43º
7º	0.298905	39º	31º	9º	0.232106	80º	70º	21º	0.269681	56º	51º	12º
8º	0.335659	25º	23º	16º	0.298462	30º	27º	14º	0.315542	29º	24º	16º
9º	0.321288	40º	32º	25º	0.277645	52º	47º	26º	0.297276	48º	42º	25º
10º	0.254448	121º	106º	56º	0.210389	153º	131º	62º	0.22138	154º	135º	65º
11º	0.333153	33º	30º	27º	0.267459	61º	59º	36º	0.278438	65º	60º	36º
12º	0.350826	22º	21º	21º	0.331327	18º	17º	15º	0.34629	19º	17º	18º
13º	0.354151	19º	18º	19º	0.326722	19º	18º	19º	0.342674	21º	19º	19º
14º	0.297917	69º	64º	40º	0.280395	54º	49º	33º	0.29226	57º	54º	33º
15º	0.321677	44º	38º	31º	0.289402	46º	39º	28º	0.301105	49º	43º	29º

Tabela A.35: Recursos gerados pelo populador - 5000 usuários - *query* 'velit' - top 15 resultados de 6050

#	Score	Nome do Recurso
1º	1.0789543	Sunt nam velit velit.
2º	1.0188178	Repellendus quia velit velit reprehenderit.
3º	0.9282018	Velit voluptas quisquam velit magnam harum.
4º	0.9194535	Voluptatibus sunt et officia velit velit.
5º	0.9194535	Velit sint voluptate fuga voluptatem velit.
6º	0.87511706	Aperiam architecto velit debitis.
7º	0.87511706	Velit ducimus voluptates harum.
8º	0.87511706	Autem modi velit dolorem.
9º	0.87511706	Eaque ab velit sequi.
10º	0.87511706	Quisquam ut velit et.
11º	0.87511706	Nisi velit et aut.
12º	0.87511706	Recusandae vero perferendis velit.
13º	0.87511706	Sit velit qui aut.
14º	0.87511706	Id minus repellendus velit.
15º	0.87511706	Dolores pariatur vel velit.

Tabela A.36: Recursos gerados - 5000 usuários - *query* 'velit' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.338405	3º	3º	1º	0.297533	3º	3º	2º	0.311262	4º	3º	2º
2º	0.462534	1º	1º	2º	0.44167	1º	1º	1º	0.457117	1º	1º	1º
3º	0.350086	14º	11º	5º	0.313725	16º	14º	4º	0.332387	16º	12º	5º
4º	0.284822	47º	44º	12º	0.239136	63º	49º	14º	0.258121	55º	44º	14º
5º	0.195881	182º	157º	40º	0.164166	208º	178º	31º	0.170313	216º	185º	39º
6º	0.29054	71º	64º	45º	0.262269	66º	58º	34º	0.276953	69º	57º	37º
7º	0.286486	77º	69º	48º	0.255829	74º	66º	40º	0.269986	73º	63º	38º
8º	0.306868	54º	50º	36º	0.270767	54º	46º	29º	0.284319	57º	49º	33º
9º	0.341539	28º	24º	19º	0.304441	35º	32º	20º	0.322031	31º	26º	20º
10º	0.565921	2º	2º	3º	0.570833	2º	2º	3º	0.600106	2º	2º	3º
11º	0.424158	8º	7º	8º	0.374142	10º	10º	9º	0.389573	11º	10º	9º
12º	0.285156	81º	71º	50º	0.25592	73º	65º	39º	0.268963	74º	66º	40º
13º	0.283715	84º	74º	51º	0.243119	94º	86º	49º	0.257817	91º	83º	49º
14º	0.382409	15º	15º	13º	0.362323	15º	13º	12º	0.378069	15º	15º	11º
15º	0.374764	18º	18º	15º	0.3513	17º	16º	13º	0.370145	17º	16º	13º

Tabela A.37: Recursos gerados pelo populador - 10000 usuários - *query* 'sint' - top 15 resultados de 15625

#	Score	Nome do Recurso
1º	1.0009131	Sint facere amet quia sint.
2º	0.9839411	Tempore sint dolor sint aliquam.
3º	0.85792553	Sint sint doloribus expedita tempore dolorum vel.
4º	0.8543059	Sint modi doloremque sint voluptas sed et.
5º	0.84337807	Voluptas dolorem dolore sint numquam sint.
6º	0.84337807	Culpa veritatis sint sint id sed aut.
7º	0.8320823	Cumque sint rerum eveniet voluptate amet sint.
8º	0.8296015	Fuga quis ut ut sint magni sint.
9º	0.8296015	Corrupti fuga sint sunt quia sed sint.
10º	0.8296015	Numquam sint sint blanditiis aut iure quaerat.
11º	0.80886	Aspernatur dicta sint aperiam.
12º	0.80886	Mollitia modi minus sint.
13º	0.80886	Sint voluptatem molestiae cupiditate.
14º	0.80886	Dignissimos vitae atque sint.
15º	0.80886	Porro et voluptatem sint.

Tabela A.38: Recursos gerados - 10000 usuários - *query* 'sint' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.21977	52º	33º	2º	0.173959	94º	67º	2º	0.18078	109º	74º	2º
2º	0.321055	1º	1º	1º	0.291987	4º	1º	1º	0.303107	3º	1º	1º
3º	0.207276	278º	230º	39º	0.17879	297º	249º	37º	0.186435	312º	258º	42º
4º	0.276996	89º	74º	10º	0.232072	110º	92º	17º	0.247407	111º	88º	18º
5º	0.334755	28º	24º	3º	0.29531	35º	35º	5º	0.313061	33º	31º	5º
6º	0.154358	684º	580º	126º	0.134915	682º	596º	107º	0.136494	756º	649º	123º
7º	0.352066	23º	21º	5º	0.267697	72º	60º	24º	0.314017	39º	36º	15º
8º	0.264737	144º	128º	45º	0.229947	149º	132º	50º	0.240717	153º	140º	51º
9º	0.29238	93º	80º	28º	0.23451	131º	118º	44º	0.249267	134º	122º	46º
10º	0.193812	418º	359º	119º	0.16341	474º	412º	112º	0.170724	488º	421º	117º
11º	0.366908	27º	26º	23º	0.338664	30º	26º	22º	0.359547	27º	22º	20º
12º	0.379007	20º	17º	17º	0.354585	18º	18º	16º	0.369721	19º	17º	16º
13º	0.393827	15º	12º	12º	0.370772	13º	11º	10º	0.390986	13º	11º	10º
14º	0.381987	18º	16º	16º	0.358029	17º	16º	14º	0.365484	20º	19º	17º
15º	0.387896	16º	15º	15º	0.361923	16º	14º	11º	0.377751	16º	15º	14º

Tabela A.39: Recursos gerados pelo populador - 10000 usuários - *query* 'velit' - top 15 resultados de 11642

#	Score	Nome do Recurso
1º	1.1227992	Id velit dicta velit.
2º	1.0825349	Distinctio architecto velit velit.
3º	0.9824493	Velit quos tempore velit odit.
4º	0.94721806	Velit velit earum quia sint.
5º	0.94721806	Provident vel quam velit velit.
6º	0.84209937	Molestias velit est ea nostrum error velit.
7º	0.84209937	Velit est eius in velit quibusdam aliquid.
8º	0.8179799	Sit totam velit eos velit officia delectus.
9º	0.8119012	Quia et sunt aliquam velit velit quod.
10º	0.8119012	Velit sunt nobis velit ea pariatur.
11º	0.8095774	Velit velit officia libero qui perferendis aliquid.
12º	0.8095774	Suscipit velit voluptatem sunt velit nihil similique.
13º	0.8095774	Est velit magnam animi rem velit.
14º	0.79776967	Velit impedit numquam animi praesentium laudantium velit.
15º	0.7939389	Velit quas nemo excepturi.

Tabela A.40: Recursos gerados - 10000 usuários - *query* 'velit' - Comparativos

#	Peso 1				Peso 2				Peso 3			
	Rank	m1	m2	m3	Rank	m1	m2	m3	Rank	m1	m2	m3
1º	0.323919	2º	1º	1º	0.265627	3º	2º	1º	0.288816	2º	2º	1º
2º	0.299998	3º	3º	2º	0.273298	5º	3º	2º	0.28464	5º	3º	2º
3º	0.43775	1º	2º	3º	0.421918	1º	1º	3º	0.443551	1º	1º	3º
4º	0.322977	8º	7º	4º	0.261251	24º	13º	4º	0.284901	20º	9º	4º
5º	0.114109	747º	595º	6º	0.103101	701º	571º	7º	0.098449	877º	726º	7º
6º	0.225246	229º	197º	27º	0.19498	239º	211º	29º	0.205198	246º	212º	34º
7º	0.229332	211º	186º	23º	0.192916	252º	217º	34º	0.203618	252º	214º	36º
8º	0.281065	116º	93º	33º	0.228822	174º	148º	44º	0.247155	155º	137º	44º
9º	0.359382	32º	26º	12º	0.333765	30º	29º	12º	0.348794	32º	29º	11º
10º	0.267317	159º	136º	50º	0.20504	262º	227º	67º	0.24102	188º	166º	54º
11º	0.373058	26º	23º	11º	0.33566	31º	30º	13º	0.350671	33º	30º	13º
12º	0.297902	86º	77º	40º	0.259694	104º	85º	42º	0.277685	91º	79º	39º
13º	0.240819	239º	208º	70º	0.201851	279º	250º	72º	0.216106	266º	236º	74º
14º	0.344172	47º	47º	24º	0.318817	45º	44º	23º	0.326629	47º	46º	28º
15º	0.378405	31º	27º	18º	0.347399	35º	32º	18º	0.358597	36º	33º	20º