



UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Autorizada pelo Decreto Federal nº 77.496 de 27/04/76

Recredenciamento pelo Decreto nº 17.228 de 25/11/2016



PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
COORDENAÇÃO DE INICIAÇÃO CIENTÍFICA

XXIV SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2020

Ferramenta de Análise de Eficácia Baseada em Diversidade para Sistemas de Recuperação de Informação

Raul do Carmo Peixoto¹ e Rodrigo Tripodi Calumby²

¹Bolsista PIBIC/CNPq, Graduando em Engenharia de Computação, UEFS, e-mail: raul.p.carmo@gmail.com

²Orientador, Departamento de Ciências Exatas, UEFS, e-mail: rtcalumby@uefs.br

PALAVRAS-CHAVE: Recuperação da Informação; Diversidade; Medidas de Eficácia.

INTRODUÇÃO

Sistemas de Recuperação de Informação (RI) visam a obtenção de dados requisitados por um usuário a partir de uma consulta em uma base de dados. O resultado geralmente é uma lista de itens que satisfaçam a requisição feita, os quais são classificados com base em sua relevância em relação à consulta. Recuperar e classificar documentos baseados apenas em sua relevância leva a dois problemas: Redundância e ambiguidade (Rodrygo et al., 2015). O problema da redundância surge uma vez que documentos com informações similares podem ser igualmente relevantes para uma consulta. A ambiguidade decorre da incapacidade de saber quais são as intenções do usuário por trás da consulta.

A RI baseada em diversidade visa atenuar os problemas citados, tendo o objetivo de trazer resultados que satisfaçam as possíveis interpretações de uma consulta, sendo diverso e, consequentemente, reduzindo a ambiguidade. A partir disso, surge a necessidade de analisar os resultados e quantificar a eficácia dos sistemas segundo a diversidade. Várias medidas e *frameworks* foram criados para análise de eficácia: Alpha (Clarke et al., 2008), IA (Agrawal et al., 2009), D e D# (Sakai et al., 2010) e Subtopic-Recall (Zhai et al., 2015). O objetivo principal deste trabalho foi projetar, desenvolver e integrar estes *frameworks* e medidas de análise de diversidade à ferramenta base, AnalyzIR. A AnalyzIR está sendo desenvolvida na UEFS e já oferece análise de relevância e significância estatística, sendo uma plataforma integrada que combina computo de medidas de eficácia, representações gráficas e comparação de resultados.

METODOLOGIA

Neste trabalho, foi realizado o estudo e catalogação dos *frameworks* e medidas de diversidade disponíveis na literatura. Posteriormente, o código e arquitetura da ferramenta foram

analisadas para modelagem das novas funcionalidades, mantendo o código conciso e aberto a futuras mudanças. Os *frameworks* foram implementados utilizando as medidas de relevância já disponíveis na ferramenta. Para implementar medidas exclusivamente de diversidade, como Subtopic-Recall, D e D#, utilizou-se a ferramenta NTCIREval (Sakai, 2011) integrado ao AnalyzIR por um wrapper. Para testar as funcionalidades, foi utilizado um conjunto de dados sintéticos.

RESULTADOS E DISCUSSÃO

Para fazer a análise de algum sistema, o usuário deve criar um *Projeto de Diversidade* (Figura 1). Para isso, o usuário deve fornecer um nome para o projeto, um diretório de trabalho onde os dados serão armazenados e utilizados durante os cálculos, um arquivo *Qrels* (*ground-truth*), um arquivo *Cluster Qrels* (*ground-truth de diversidade*), um arquivo opcional *IProb* (probabilidades de cada intenção) e os sistemas que serão analisados (*Runs*).

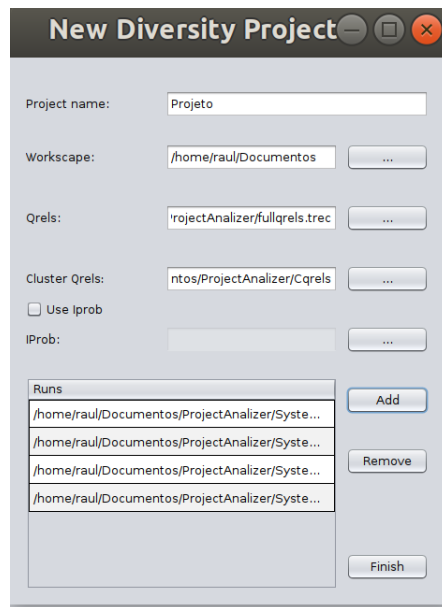


Figura 1: Tela de Criação do Projeto de Diversidade Preenchida.

Com o projeto criado, o usuário pode criar gráficos com as medidas disponíveis. Para criar um gráfico, primeiramente, deve-se escolher um *framework de diversidade* (Figura 2). O próximo passo é selecionar os dados a serem utilizados na criação do gráfico (Figura 3). As *Runs* são os sistemas que serão avaliados; *Topic* é uma execução (consulta) nesse sistema; Por fim, seleciona-se as medidas de diversidade do *framework* escolhido. Caso o usuário escolha o *framework* D# ou *Alpha*, também deverá fornecer os valores de *Gamma* e *Alpha*.

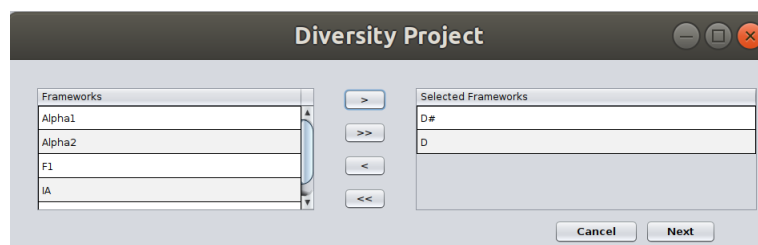


Figura 2: Tela de Escolha do *framework*.

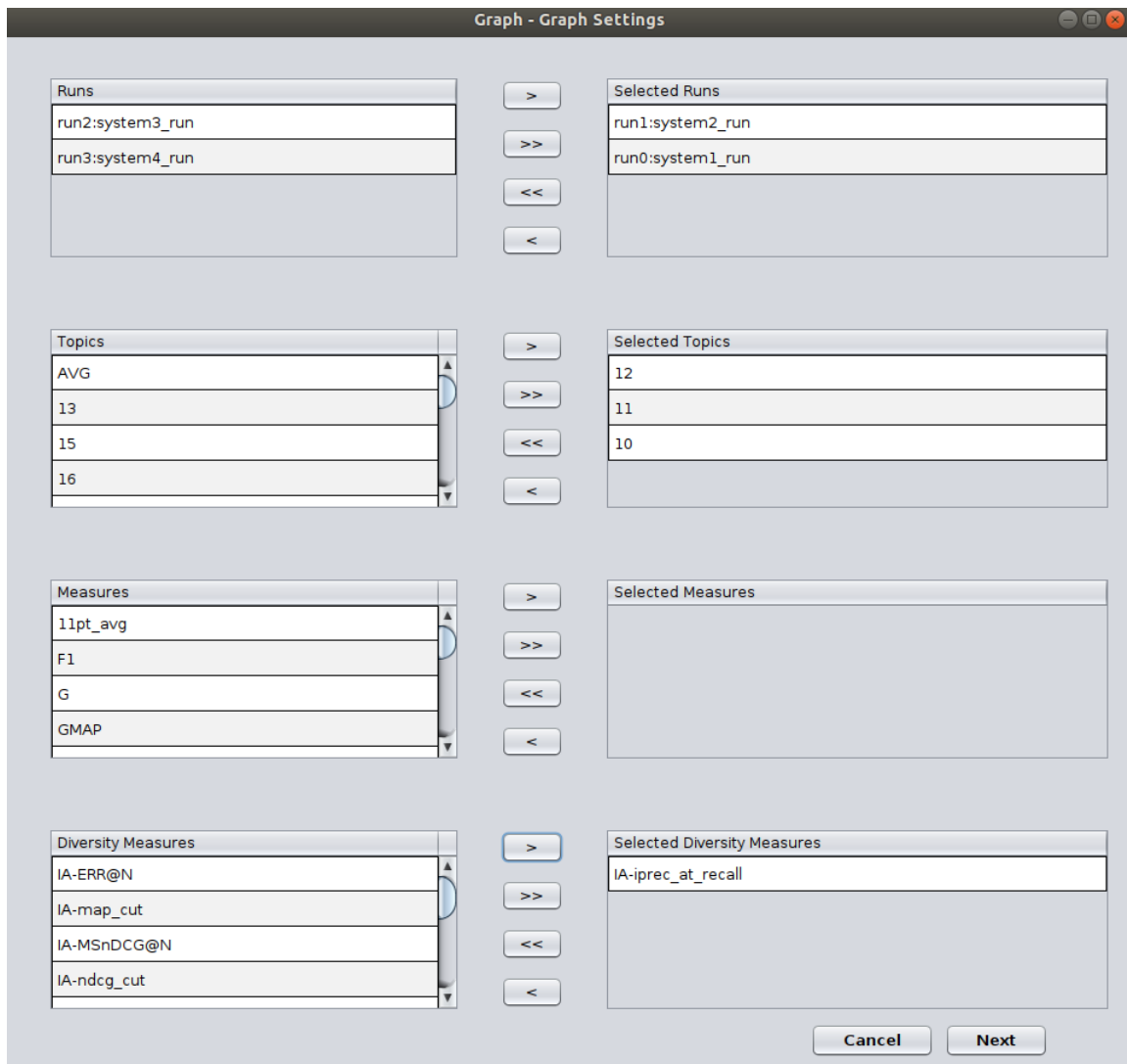


Figura 3: Tela de construção do gráfico.

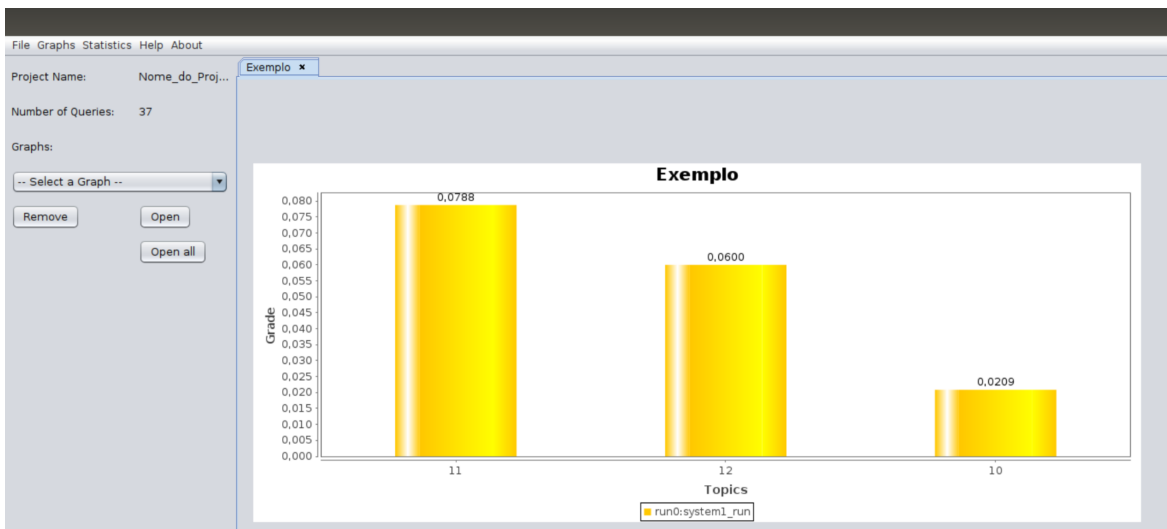


Figura 4: Exemplo de resultado com o sistema *run1*, os tópicos 10, 11 e 12 e a medida *IA-P@N*.

A Figura 4 apresenta um exemplo de gráfico utilizando o sistema "run1", as consultas 10, 11 e 12 e a medida $IA-P@N$.

Por fim, também foi incluída medida para computo da média harmônica, chamada F1 (Yang and Liu, 1999) como mostra a Equação 1. O F1 combina a nota de duas medidas quaisquer, sendo estas de relevância ou diversidade, tal que m_1 e m_2 são os resultados das duas medidas. Importante frisar que esta opção está presente tanto no projeto de relevância como no de diversidade. É possível combinar uma medida de relevância com uma de diversidade.

$$F1 = \frac{2m_1m_2}{m_1 + m_2} \quad (1)$$

CONSIDERAÇÕES FINAIS

Com a adição do projeto de diversidade e cálculo de medidas de diversidade, a AnalyzIR tornou-se ainda mais robusta, oferecendo uma plataforma integrada para análise de diversidade juntamente com as outras funcionalidades já disponíveis. Sendo assim, tem-se uma ferramenta importante para pesquisadores e profissionais da área, devido à sua intuitividade de uso e funcionalidades integradas. Além disso, houve a produção de documentação contendo descrições detalhadas e exemplos ilustrativas sobre cada *framework* integrado à ferramenta.

REFERÊNCIAS

- AGRAWAL, R., GOLLAPUDI, S., HALVERSON, A.; IEONG, S. 2009. Diversifying search results. In Proceedings of the second ACM international conference on web search and data mining, p.5–14. ACM.
- CLARKE, C.L., KOLLA, M., CORMACK, G.V., VECHTOMOVA, O., ASHKAN, A., BÜTTCHER, S.; MACKINNON, I. 2008. Novelty and diversity in information retrieval evaluation. In Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, p.659–666. ACM.
- RODRYGO, L., MACDONALD, C.; OUNIS, I. 2015. Search result diversification. Foundations and Trends in Information Retrieval, 9(1):1–90
- SAKAI, T. 2011. Ntcireval: A generic toolkit for information access evaluation. In Proceedings of the forum on information technology, volume 2, p.23–30.
- SAKAI, T., CRASWELL, N., SONG, R., ROBERTSON, S., DOU, Z.; LIN, C.-Y. 2010. Simple evaluation metrics for diversified search results. In EVIA@ NTCIR, p.42–50.
- YANG, Y.; LIU, X. 1999. A re-examination of text categorization methods. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, p.42–49.
- ZHAI, C., COHEN, W.W.; LAFFERTY, J. 2015. Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In ACM SIGIR Forum, volume 49, p.2–9. ACM.