

# **Ferramental para Ensino e Aprendizado de Análise de Eficácia de Sistemas de Recuperação da Informação**

**Wanderson Bezerra da Silva<sup>1</sup>; Rodrigo Tripodi Calumby<sup>2</sup>**

1. Bolsista PIBIT/CNPq, Graduando em Engenharia da Computação, UEFS, e-mail: bsilva.wanderson@gmail.com

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

**PALAVRAS-CHAVE:** Eficácia; Aprendizado; Recuperação da Informação.

## **INTRODUÇÃO**

Dadas as necessidades dos usuários e a grande quantidade de informação disponível, muitos métodos têm sido propostos com o objetivo de gerar resultados cada vez melhores em diversas atividades, tais como recuperação, classificação e recomendação de informação. Neste contexto, um fator de grande importância são as métricas de avaliação dos métodos e o modo como são comparadas suas eficácias. (Ferri et. al. 2009; Yilmaz & Robertson, 2010). Assim, a correta avaliação dos resultados e a comparação da qualidade de diferentes sistemas têm sido áreas que atraem a atenção da comunidade científica. Em recuperação da informação (RI), este processo normalmente baseia-se em: (i) construção de bases de dados; (ii) definição de tarefas a serem realizadas sobre esta base; (iii) execução de tarefas seguindo diferentes métodos; (iv) avaliação dos resultados usando as métricas escolhidas; e por fim, (v) o comparativo dos métodos.

Muitas vezes um valor numérico obtido através de uma análise utilizando uma métrica de avaliação não é o suficiente para um usuário entender o comportamento dos resultados de um sistema de RI. Por isso, é importante utilizar métodos de análise de falhas, permitindo ao usuário ter um melhor entendimento dos resultados obtidos pelo sistema de RI avaliado. Estes modelos dão ao usuário a possibilidade de analisar a distribuição e comportamento dos itens ao longo da lista de resultados.

Para uma boa avaliação dos resultados, é necessário conhecer as medidas de avaliação, o objetivo de cada uma e as diferenças entre elas. Este trabalho foi realizado com o intuito de dar suporte aos usuários, para que os mesmos adquiram os conhecimentos necessários para fazer a análise, compreensão e comparação dos resultados de um sistema de RI. O trabalho desenvolvido deu continuidade a projetos anteriores que tem como base a AnalyzIR, uma ferramenta que vem sendo desenvolvida na UEFS e tem como objetivo automatizar os processos de análise dos resultados de um sistema de RI.

## **MATERIAL E MÉTODOS**

A AnalyzIR tem por objetivo atender às principais necessidades dos pesquisadores da comunidade de RI, automatizando os processos de análise, de forma simples e flexível, possuindo funcionalidades que carecem nas principais ferramentas de avaliação de RI disponíveis. De modo a agregar informações úteis para ensino e aprendizado acerca da avaliação de sistemas de RI, foi criada uma seção com conteúdo educativo na página

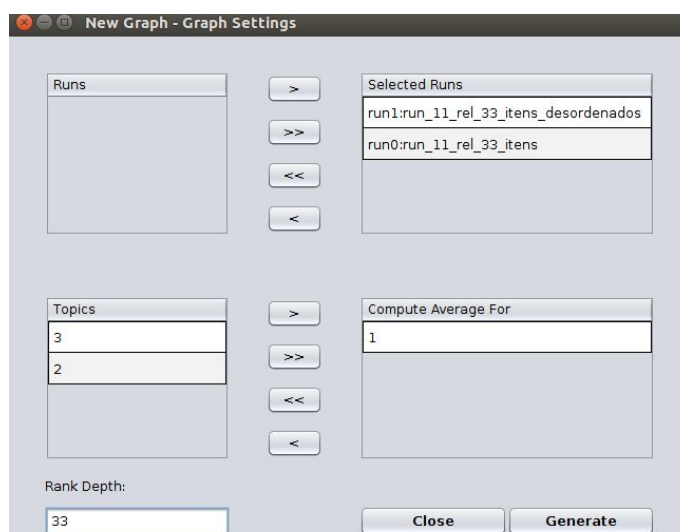
web<sup>1</sup> da ferramenta. Já na própria AnalyzIR, foram incorporadas informações descritivas das medidas de avaliação disponíveis na ferramenta, de modo a prover um guia para entendimento dos resultados obtidos com as medidas de avaliação utilizadas.

Na avaliação de RI, os itens presentes no resultado podem ser classificados como relevantes ou irrelevantes. Em sistemas que apresentaram resultados com múltiplos níveis de relevância, o resultado indica se um item é relevante e o quão relevante ele é. Para facilitar esta análise e melhorar a compreensão dos resultados, foi adicionado à AnalyzIR o F\_Rel (Find Relevants). O F\_Rel é um método de análise de falhas que permite ao usuário identificar o quão relevante é um item e como esses itens estão distribuídos ao longo de um ranking.

Para auxiliar no uso da AnalyzIR, foram criados vídeos tutoriais da ferramenta, demonstrando suas principais funcionalidades.

## RESULTADOS E/OU DISCUSSÃO

A Figura 1 mostra tela de configuração do gráfico do F\_Rel, onde devem ser definidos quais sistemas e consultas serão utilizadas para a avaliação e a profundidade do ranking. Após inserir as informações necessárias e clicar no botão *Generate*, o gráfico do F\_Rel será exibido. Este modelo de análise de falhas permite uma visualização completa da lista dos resultados obtidos, permitindo ao usuário identificar onde estão distribuídos os itens e qual o nível de relevância de cada um, como mostra a Figura 2. Um esquema com diferentes tons de azul foi utilizado para diferenciar os itens mais relevantes dos poucos relevantes, onde, quanto mais escuro for a tonalidade, mais relevante é o item por ela identificado. São utilizados até 5 tons de azul, então, caso haja mais que 5 níveis de relevância, valores próximos podem ser identificados com a mesma tonalidade. Os itens irrelevantes são identificados pela cor cinza.



**Figura 1 - Tela de configuração do gráfico do F\_Rel.**

A Figura 2 apresenta o gráfico do F\_Rel para os resultados de dois sistemas com 12 níveis de relevância. Com este modelo pode-se ver claramente como os itens relevantes estão distribuídos. O resultado representado pela barra da direita indica um melhor

<sup>1</sup> <https://sites.google.com/site/analyzir/learning-resources>

resultado, pois os itens mais relevantes estão no topo do ranking e os menos relevantes estão na base. No sistema representado pela barra da esquerda há itens com um nível de relevância mais baixo acima de itens com um nível de relevância mais alto.

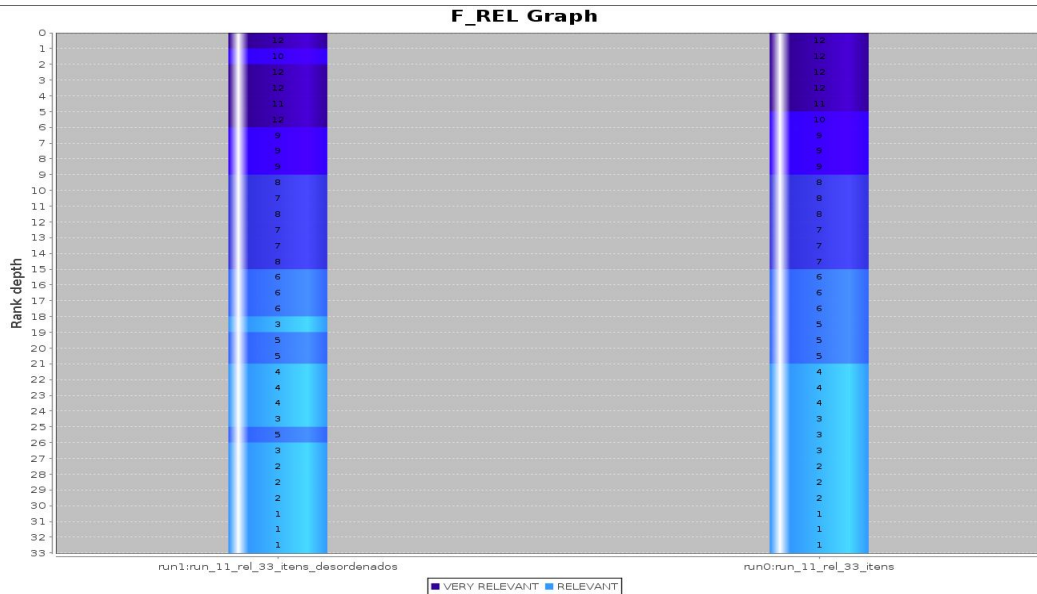


Figura 2 - Tela de visualização do F\_Rel.

Conhecer as medidas de avaliação é um passo muito importante para quem deseja aprender avaliação de sistemas de RI. Com base nesta informação, foi adicionada à AnalyzIR uma funcionalidade que contém descrições para todas as medidas de avaliação presentes na ferramenta. Para acessar esta funcionalidade basta acessar o menu Help e selecionar a opção Measure Description. Em seguida será exibida a janela representada na Figura 3.

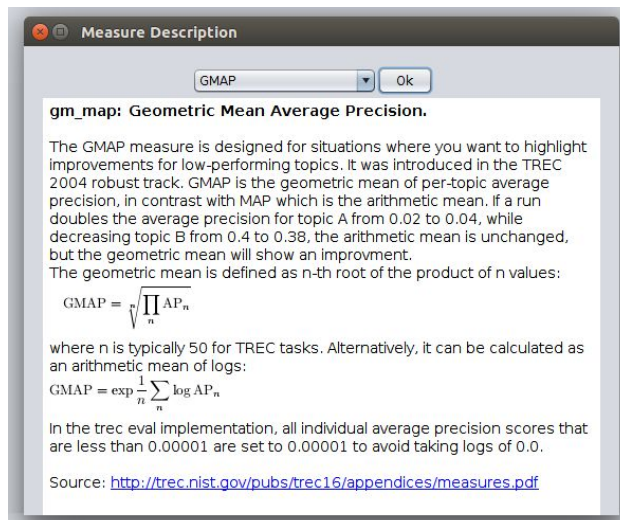


Figura 3 - Descrição da medida GMAP.

A criação da seção com conteúdo educativo na página web da ferramenta AnalyzIR agrega informações úteis para ensino e aprendizado acerca da avaliação de sistemas de RI. Estas informações incluem artigos de revisão de literatura acerca do tema, artigos relacionados à avaliação de sistemas de RI, links para ferramentas úteis, livros, páginas web e vídeos sobre avaliação de sistemas de RI. A seção pode ser acessada pelo link

<https://sites.google.com/site/analyzir/learning-resources> e está dividida em informações textuais (artigos, livros, entre outros). A seção já possui materiais sobre ensino e aprendizado de RI, avaliação de RI, além disso, também conta com artigos sobre outras ferramentas para RI e avaliação de RI.

Para demonstrar as principais funcionalidades da AnalyZIR, 9 vídeos foram produzidos, possuindo os respectivos conteúdos:

1. Criação de um novo projeto;
2. Importar um projeto;
3. Criação de gráficos com parâmetros predefinidos: Etapas para criar um predef;
4. Usar o guia das descrições das medidas;
5. Etapas e parâmetros necessários para criar uma gráfico de barra;
6. Etapas e parâmetros necessários para criar uma gráfico de curva;
7. Edição e exportação de gráficos;
8. Etapas e parâmetros necessários para criar uma gráfico do F\_Rel;
9. Etapas e parâmetros necessários para criar uma gráfico do R\_Pos;

Estes vídeos serão disponibilizados junto ao manual da ferramenta na sua página web.

## CONSIDERAÇÕES FINAIS

Para uma boa análise e compreensão dos resultados de um sistema de RI, é essencial ter um domínio das técnicas e das métricas utilizadas na avaliação. O conteúdo disponível na seção educacional do site da AnalyZIR fornece ao usuário os conhecimentos básicos necessários para entender como se dá o processo de avaliação dos sistemas de RI e como interpretar os resultados obtidos. O guia de descrição das medidas presentes na AnalyZIR é de grande importância tanto para usuários iniciantes quanto para usuário que já possuem um certo nível experiência. O usuário iniciante pode usar o guia para conhecer as principais medidas de avaliação, e ter mais segurança ao avaliar os resultados de um sistema de RI, enquanto um usuário mais experiente pode expandir seus conhecimentos utilizando estas descrições.

De forma geral, o ferramental desenvolvido procura tornar práticas, atividades de ensino e aprendizado acerca do tema, não apenas na UEFS, mas também em outras instituições de ensino, pesquisa e empresas da área, fornecendo em um mesmo ambiente os métodos necessários para fazer a avaliação dos sistemas de RI e os conteúdos a eles associados.

## REFERÊNCIAS

- FERRI, C;HERNÁNDEZ-ORALLO, J; MODROIU, R. (2009). An experimental comparison of performance measures for classification. *Pattern Recognition Letters*, 30(1), 27-38.
- YILMAZ, E;ROBERTSON, S. (2010). On the choice of effectiveness measures for learning to rank. *Information Retrieval*, 13(3), 271-290.
- ANGELINI, M; FERRO, N; SANTUCCI, G; SILVELLO, G. 2012. Visual Interactive Failure Analysis: Supporting Users in Information Retrieval Evaluation. Proc. *4th Symposium on Information Interaction in Context*. ACM Press, New York, USA