

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

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

1. Bolsista PROBIC/UEFS, Graduando em Engenharia da Computação, UEFS, e-mail: [bsilva.wanderson@gmail.com](mailto:bsilva.wanderson@gmail.com)
2. Orientador, Departamento de Ciências Exatas, UEFS, e-mail: [rtcalumby@ecomp.uefs.br](mailto:rtcalumby@ecomp.uefs.br)

**PALAVRAS-CHAVE:** Eficácia, Aprendizado, Recuperação da Informação

## **INTRODUÇÃO**

Dadas as necessidades dos usuários e a grande quantidade de informação disponível, uma enorme quantidade de métodos tem sido proposta 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 as 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; (iii) avaliação dos resultados usando as métricas escolhidas; e por fim, (iv) o comparativo dos métodos.

Atualmente, para efetuar o comparativo de sistemas usando métricas de eficácia, pesquisadores necessitam utilizar um conjunto de ferramentas para mensurar eficácia, realizar análise estatística destes valores e construir modelos visuais para apresentação destes dados. Uma outra grande dificuldade está relacionada à visualização dos resultados da análise de eficácia que, em geral, demanda a construção manual de modelos comparativos.

Este trabalho foi realizado com o intuito de facilitar a análise, compreensão e comparação dos resultados de um sistema, utilizando modelos onde o usuário não precisa se preocupar com os parâmetros utilizados na avaliação. Também foi desenvolvido uma nova forma de análise e visualização dos resultados.

## **MATERIAIS E MÉTODOS**

Em geral, o ensino de RI não difere de outras disciplinas e seus conceitos introdutórios são de fácil entendimento. Com o objetivo de auxiliar usuários pouco experientes nas suas primeiras atividades de avaliação de sistemas de recuperação da informação, foram criados modelos pré-definidos de análise de resultados e construção de comparativos visuais. Diferente da criação dos gráficos convencionais presentes na ferramenta AnalyzIR, esses comparativos predefinidos (Predefs) possuem parâmetros já inseridos diretamente à ferramenta, ou seja, o usuário não precisa se preocupar com o arranjo de medidas e com as consultas que serão utilizadas para avaliação. Os tipos de predefs adicionados à ferramenta são: *Basic Precision*, *Basic Curve*, *Gain* e *Rank Top*.

O predef *Basic Precision* inclui medidas derivadas da Precision, sendo elas: GM\_MAP, MAP e Recip\_rank. O objetivo deste Predef é avaliar o sistema usando as mais conhecidas medidas baseadas em Precision.

O *Gain* possui as medidas baseadas em ganho. Estas medidas levam em consideração os múltiplos níveis de relevância dos elementos e sua posição no ranking. Este predef contém três medidas: G, binG e NDCG. A medida binG utiliza níveis de relevância binária, já as medidas G e NDCG utilizam múltiplos níveis de relevância. Estas medidas tentam computar se os itens relevantes estão no topo do ranking, pois itens relevantes que aparecem no final do ranking tem contribuição reduzida.

O *RankTop* é um predef com o intuito de avaliar apenas o topo do ranking. Este Predef possui medidas nas quais o usuário pode especificar a profundidade do ranking a ser avaliado, sendo elas: Precision, Recall, MAP e NDCG. O predef RankTop possui as principais medidas presentes nos demais predefs, porém considera apenas o topo do ranking.

O *Basic Curve* utiliza gráficos de curva para a análise do comportamento dos sistemas. Utiliza as médias mais simples para o entendimento do usuário iniciante na área de RI, sendo elas: Precision, Recall e a curva de Precision e Recall. O usuário precisa apenas informar a profundidade do ranking e o predef cria uma curva que começa no ponto 5 e vai até a profundidade estabelecida, criando pontos de curva a cada 5 posições no ranking.

Adicionalmente, o modelo de análise visual da distribuição dos resultados implementado é o Failure Analysis (análise de falhas). Segundo (Angeline et. al. 2012), o FailureAnalysis, é um modelo de classificação que nos permite compreender o que acontece quando se desloca documentos com diferentes graus de relevância em uma lista ranqueada, por exemplo, colocar um documento pouco relevante antes de um documento altamente relevante para a consulta. Seguindo este modelo, foi implementada a função Relative Position (R\_Pos), que indica o quanto um documento foi deslocado em relação à sua posição ideal.

## RESULTADOS E DISCUSSÃO

Como mostrado na Figura 1, para utilizar os predefs basta ir ao menu *Graphs*, selecionar a opção *New Predef* e selecionar o predef desejado.

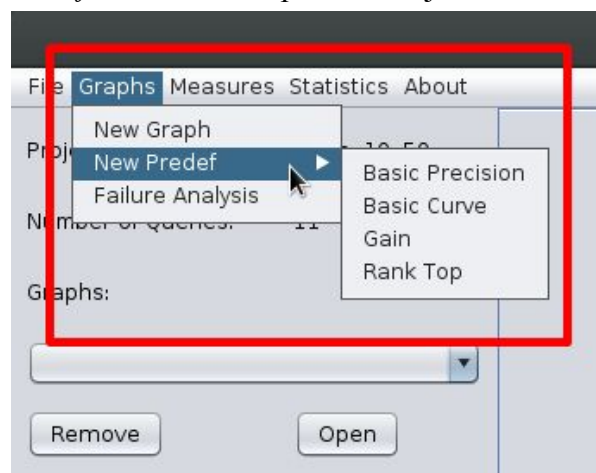
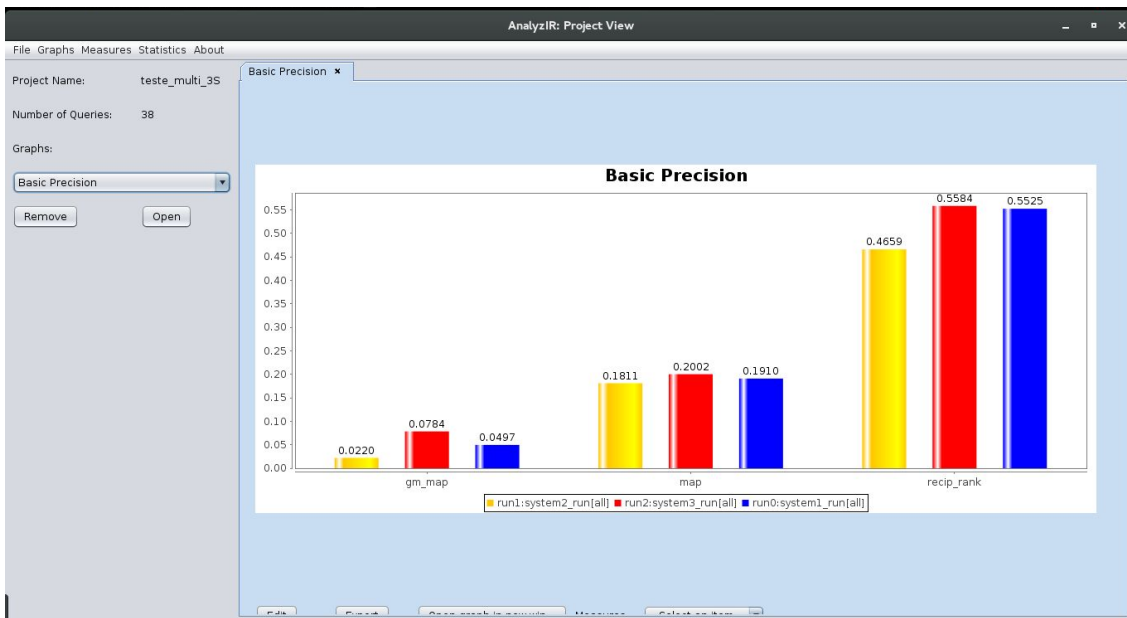


Figura 1 - Menu Graph presente na ferramenta AnalyzIR

A Figura 2 mostra três sistemas sendo avaliados pelo predef *Basic Precision*. O sistema 3 (vermelho) é o melhor avaliado, tendo o melhor resultado em todas as medidas do predef. Em seguida vem o sistema 1 (azul), que consegue obter um desempenho bastante próximo do sistema 3. Por último, fica o sistema 2 (amarelo), que consegue um resultado inferior aos demais sistemas em todas as medidas.



**Figura 2 - Predef *Basic Precision***

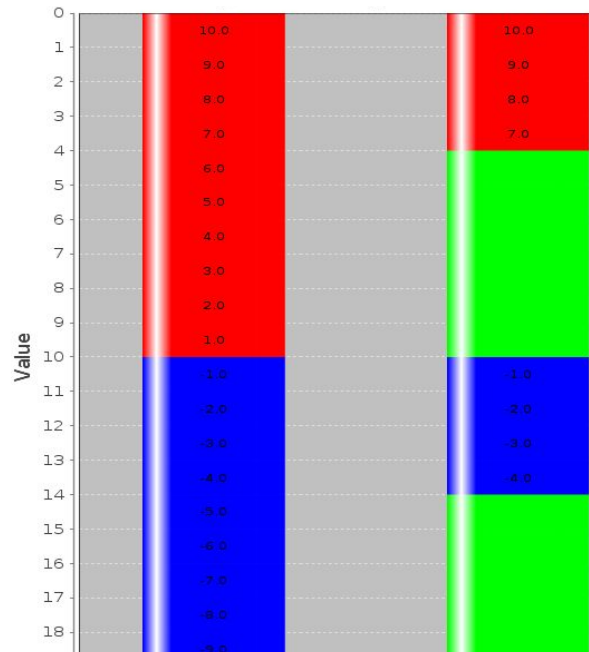
Para utilizar o modelo de análise de falhas na ferramenta AnalyzIR, basta ir ao menu Graphs, e selecionar a opção Failure Analysis, como mostra a Figura 1. Em seguida será exibida a tela para configuração do gráfico do R\_Pos (Figura 3), onde o usuário deve informar quais sistemas serão avaliados, a profundidade do ranking e com quais consultas (topics) será feita a avaliação.



**Figura 3 - Tela de configuração do gráfico do R\_Pos**

Cada posição no gráfico indica um documento no ranking e sua cor indica o respectivo *escore R\_Pos*. Utiliza-se a cor verde para os documentos que estão na sua posição ideal,

azul para os que estão abaixo do ideal e vermelho para os documentos que estão acima da posição ideal. Os valores presentes dentro das células indicam o valor de  $R\_Pos$  do documento, ou seja, se um documento tem  $R\_Pos = 2$ , significa que ele está duas posições acima do intervalo ideal.



**Figura 4 - Gráfico do  $R\_Pos$ .**

A Figura 4 mostra dois sistemas sendo avaliados com o  $R\_Pos$ . Com este método o usuário pode analisar o comportamento do seu sistema, identificando se há uma maior incidência de erros no topo, no final ou ao longo do ranking.

## CONSIDERAÇÕES FINAIS

O modelo de análise de falhas desenvolvido e sua forma de visualização são bastante inovadores, pois, seu objetivo é mostrar para o usuário o comportamento do sistema, diferente dos outros modelos de classificação que geralmente tem como foco apresentar para o usuário os resultados numéricos de eficácia de um sistema. Os predefs são importantes para usuários iniciantes na área de RI e avaliação de RI, afinal, não precisa se preocupar com o arranjo de medidas e com as consultas que serão utilizadas para avaliação.

Os resultados deste trabalho irão contribuir decisivamente com as pesquisas desenvolvidas na UEFS. Este projeto 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

## REFERÊNCIAS

- FERRI, C.; ORALLO, J. H; MODROIU, R. 2009. An experimental comparison of performance measures for classification. Pattern Recogn.
- YILMAZ, E; ROBERTSON, S. 2010. On the choice of effectiveness measures for learning to rank. Inf. Retr..
- ANGELINI, M; FERRO, N; SANTUCCI, G; SILVELLO, G. 2012. Visual Interactive Failure Analysis: Supporting Users in Information Retrieval Evaluation.