



UNIVERSIDADE DE BRASÍLIA

INSTITUTO DE LETRAS

DEPARTAMENTO DE LÍNGUAS ESTRANGEIRAS E TRADUÇÃO

LÍNGUAS ESTRANGEIRAS APLICADAS AO MULTILINGUISTO E À SOCIEDADE DA
INFORMAÇÃO

BÁRBARA COSTA

**AVALIAÇÃO DE SUMARIZADORES AUTOMÁTICOS DE TEXTOS EM INGLÊS A
PARTIR DA FERRAMENTA COMPUTACIONAL ROUGE**

BRASÍLIA

Maio de 2020

Bárbara Costa

**AVALIAÇÃO DE SUMARIZADORES AUTOMÁTICOS DE TEXTOS EM INGLÊS A
PARTIR DA FERRAMENTA COMPUTACIONAL ROUGE**

Trabalho de Conclusão de Curso do Bacharelado
em Línguas Estrangeiras Aplicadas ao
Multilinguismo e à Sociedade da Informação
(LEA-MSI) da Universidade de Brasília

Orientador: Prof. Dr. Francisco Cláudio Sampaio
de Menezes

BRASÍLIA

RESUMO

A sumarização de textos, no presente trabalho definida como a produção de textos resumidos, é uma atividade frequente e necessária no cotidiano e, principalmente, no ambiente acadêmico, haja vista a velocidade e volume de informações e conteúdos que são disponibilizados, razão pela qual surgiu a necessidade de automatizar esse processo. Assim, a Sumarização Automática (SA) trata da produção automática de sumários a partir de um ou mais textos-fonte e é uma subárea de pesquisa de Processamento de Língua Natural (PARDO, 2008, p.2). Dessa forma, a partir dos estudos feitos no projeto de extensão MOBILANG na linha de pesquisa “Línguas em contato na web”, será oferecida uma avaliação e comparação sobre a qualidade de dois dos sumarizadores para textos acadêmicos gratuitos e mais populares disponíveis na web: *Tools4Noobs* e SMMRY. A avaliação será realizada por meio do avaliador automático de sumários ROUGE 2.0 (Recall-Oriented Understudy for Gisting Evaluation), ferramenta implementada e explorada no projeto.

Palavras-chave: Sumarização automática; Sumários de referência humana; Avaliação automática de sumários; Sumarizadores em língua inglesa;

ABSTRACT

The summarization of texts is a frequent and necessary activity in daily life and, mainly, in the academic environment, given the speed and volume of information made available. Therefore, there was a need to automate that process. Automatic Summarization deals with the automatic production of summaries from one or more source texts and is a research subarea of Natural Language Processing (PARDO, 2008, p.2). Thus, based on the studies carried out in the MOBILANG extension project in the research line called Línguas em contato na web (languages in contact on the web), the work proposes an evaluation and comparison on the quality of two of the summarizers most popular for academic texts available on the web for free: Tools4Noobs and SMMRY. The evaluation will be guided by the automatic summary evaluator ROUGE 2.0 (recall-oriented understudy for gisting evaluation), a tool implemented and explored during the project.

Keywords: Automatic summarization; Human reference summaries; Automatic evaluation of summaries; English-language summarizers;

Nenhum homem é melhor do que uma máquina e nenhuma máquina é melhor do que um homem com uma máquina.” Paul Tudor Jones.

1. INTRODUÇÃO

Os sumários são extremamente úteis e frequentes na sociedade contemporânea. Pode-se citar como sumários escritos, por exemplo, notícias em jornais, artigos de revistas, resumo de textos científicos, entre muitos outros. O sumário é um resumo, cujo objetivo é captar a ideia principal do autor e passá-la em poucas linhas para o leitor (GUELPELI, 2012, p.40).

Os estudos sobre sumarização datam dos anos 50, e a ciência do processamento de linguagem natural, cujo campo de estudo realiza uma aproximação entre a linguística e a computação, tem se debruçado sobre a pesquisa nos estudos de sumarização automática e teve grandes avanços, principalmente no início do novo milênio, na sociedade da informação, com a difusão da internet.

Na perspectiva de Pardo (2008, p.6), os sumários produzidos pela sumarização automática (SA) podem ser classificados quanto à função, audiência e formação. Quanto à formação, têm-se os sumários *abstracts* e extratos. Ainda segundo esse mesmo autor, relativo ao nível linguístico utilizado pelo sumário, temos duas abordagens: a abordagem superficial e a abordagem profunda. E, caso tenhamos as duas abordagens em um mesmo sumário, trata-se da abordagem híbrida.

Vale destacar que o interesse por tecnologias linguísticas e a participação nos estudos de SA, na linha de pesquisa “Línguas em contato na web” (coordenada pelo prof. Cláudio Menezes) do grupo de pesquisa Mobilidades e contatos de línguas (MOBILANG), nessa frente de pesquisa, foram os propulsores para a elaboração do presente artigo.

Dessa forma, busca-se comparar e analisar dois sumarizadores bastante populares da língua inglesa com o objetivo de analisar os sumários automáticos disponíveis na web voltados para resumos de textos acadêmicos em inglês, de atestar a qualidade desses resumos de forma computadorizada, bem como de explorar o uso da ferramenta ROUGE. O primeiro a ser analisado denominado SMMRY, possui diversos recursos e configurações avançadas, como o uso de URL ou upload de arquivos. A segunda ferramenta de resumo será o *Tools4Noobs* que também possui uma ampla gama de configurações. O *software* funciona de forma similar com upload de textos ou inserção de URL que deseja sumarizar.

Tendo em vista a pertinência da sumarização automática, um dos maiores desafios da área é gerar sumários de fato informativos em relação ao texto-fonte. Nesse estudo, os dois

sumarizadores produzirão resumos a partir de textos acadêmicos, selecionados na área de Língua e Literatura que serão submetidos à análise pelo *software* de avaliação automática ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*). Tal ferramenta avalia o sumário automático a partir do cálculo da quantidade de palavras similares em relação ao sumário ideal (LIN e HOVY, 2004, p.1) e é uma das mais utilizadas e amplamente divulgadas para a finalidade proposta.

2. SUMARIZADORES AUTOMÁTICOS DE TEXTOS

Um sumário é um resumo que tem o objetivo de captar a ideia principal de determinado autor e representá-la em poucas linhas (HUTCHIN, 1987 apud GUELPELI; BERNARDINI; GARCIA, 2008, p.1). Os sumarizadores selecionados no presente trabalho utilizam o método extrativo de sumarização e os textos obtidos podem ser classificados como extratos informativos, genéricos e monodocumentos. Para (Pardo, 2008, p.5), a sumarização automática monodocumento tradicional produz o sumário de um único texto-fonte, já os extratos podem ser entendidos como uma composição de segmentos relevantes de um texto. Os sumários extrativos têm como parâmetros clássicos de informatividade as medidas de *precision* (precisão), *recall* (cobertura) e *F-measure* (média). As medidas podem ser explicadas da seguinte forma:

“o Recall é número de sentenças do sumário automático presentes no sumário de referência / número de sentenças do sumário automático; o *Precision*: número de sentenças do sumário automático presentes no sumário de referência / número de sentenças do sumário de referência; o *F-Measure*: $((P \times R) / (P + R)) \times 2$, o *F-Measure* combina as métricas de *Recall* e de *Precision*. O resultado F-Measure é um indicativo de que, quanto mais próximo de 1, melhor é o sumário e resultados mais próximos de 0, demonstram que os sumários são ruins.” (GUELPELI; BERNARDINI; GARCIA, 2008, p.2)¹

Ainda quanto à classificação, podemos dizer que ambos seguem uma abordagem superficial, ou seja, faz pouco ou nenhum uso de conhecimento linguístico para produzir sumários e quando usa conhecimento, este se restringe, em geral, aos níveis morfossintático e sintático (PARDO, 2008, p.4).

O SMMRY é um sumarizador online criado para resumir artigos e textos acadêmicos. Para fornecer as informações necessárias de maneira eficiente, o programa classifica as sentenças

¹(GUELPELI; BERNARDINI; GARCIA, 2008, p.2)

usando o algoritmo principal em um processo de sete etapas que faz a associação de palavras com suas contrapartes gramaticais (por exemplo, "carro" e "carros"), calcula a ocorrência de cada palavra no texto, atribui pontos a cada palavra, dependendo de sua popularidade, detecta quais palavras podem figurar no final de uma frase (por exemplo, "Sr." Não pode terminar uma frase), divide o texto em frases individuais, classifica as sentenças pela soma dos pontos de suas palavras e, por fim, faz o retorno das sentenças mais bem classificadas de acordo com os pontos alcançados. Além disso, esse sumário reorganiza o resumo para focar um tópico; por meio da seleção de uma palavra-chave, remove frases de transição e faz também a remoção de cláusulas desnecessárias. O usuário tem a liberdade de regular a taxa de compressão dos sumários indicando como parâmetros o número de linhas e número de caracteres do sumário.

O sumário *Tools4Noobs*² por sua vez, faz um processamento em menos fases: extrai as sentenças do texto fornecido, identifica as palavras-chave e a relevância de cada palavra e posteriormente identifica e seleciona as frases com palavras-chave mais importantes. A relevância é um número (não uma porcentagem) calculado adicionando a relevância de cada palavra-chave detectada. Vale destacar que o processo de extração das sentenças segue alguns parâmetros, como o Limiar, que é o valor usado para limitar as frases com base em sua relevância, sendo a relevância determinada pelo número ou palavras relevantes na sentença. Acrescenta-se que o número de linhas que devem ser retornadas no resumo pode ser definido pelo usuário. Essas linhas serão classificadas pela relevância (as primeiras linhas serão as mais relevantes). Considera-se também o comprimento mínimo da frase sendo que por definição são 50 caracteres e comprimento mínimo da palavra que por padrão deve conter no mínimo 4 caracteres.

Nesse programa, há opções adicionais de mostrar as palavras mais relevantes do texto resumido. Se essa opção for escolhida serão apresentadas as palavras-chave mais relevantes, bem como a relevância dessas palavras (número de ocorrências em todo o texto). Se esta opção estiver selecionada, uma saída adicional será gerada com o mesmo resumo, mas com as palavras-chave destacadas, quanto mais relevante for cada palavra-chave, maior será a aparência.

² Disponível em: <https://www.tools4noobs.com/summarize/>

3. AVALIAÇÃO DE SISTEMAS DE SUMARIZAÇÃO

Tão importante quanto a produção automática de sumário é a avaliação da sua qualidade e informatividade. Para uma avaliação de qualidade de sumário, são realizadas conferências no âmbito internacional dedicadas a esse tema, como a TAC (*Text Analysis Conference*) que ocorre anualmente. Segundo (Pardo, 2008, p.8), a qualidade de um sumário diz respeito ao seu grau de legibilidade, gramaticalidade e fluência, dentre outros quesitos, enquanto a informatividade refere-se ao volume de informação apresentado pelo sumário.

Nesse sentido, a ferramenta computacional de avaliação automática adotada neste trabalho é uma das mais difundidas e usadas nesse campo de estudo para avaliação de conteúdo: o ROUGE, ou *Recall-Oriented Understudy for Gisting Evaluation* que é um método para determinar automaticamente a qualidade de um resumo comparando-o a outro conjunto de resumos (ideais), frequentemente criados por humanos (LIN e HOVY, 2003, p.2). Dessa forma, permite-se comparar diretamente um sumário automático e um sumário humano por meio de n-gramas, ou seja, esse sistema calcula a sequência de conjunto de palavras em comum entre os dois sumários. Quanto mais n-gramas em comum houver, maior a nota (entre 0 e 1) atribuída ao sumário automático (PARDO, 2008, p.8).

3.1 Implementação do pacote ROUGE 2.0

O ROUGE 2.0 é uma versão em Java do ROUGE original em Pearl. Segundo (Ganesan, 2018):

“Essa versão aprimorada e atualizada, permite a captura da sobreposição semântica por meio do uso de um dicionário de sinônimos e também permite a avaliação de tópicos específicos ou subconjunto de conteúdo. Além disso, a versão Java simplifica todo o sistema e a convenção de nomenclatura do resumo de referência e a formatação, permitindo que os pesquisadores se concentrem na avaliação. Uma vez que este pacote é independente de plataforma, ele permitiria um alcance mais amplo a todos os pesquisadores e desenvolvedores da indústria.” (GANESAN, 2018, p.4, tradução livre)³

³(GANESAN, 2018, p.4, tradução livre)

Nesse artigo, a autora disponibiliza o pacote completo desse avaliador, juntamente com o código-fonte. A partir desse material que foi implantado a versão para a UnB do ROUGE 2.0 que é atualizada e mantida pelo estudante Talles Silva, integrante do grupo de pesquisa, e pode ser acessada a partir deste endereço eletrônico: <https://infinite-dawn-29318.herokuapp.com/login>.

4. SUMÁRIOS DE REFERÊNCIA

A seleção de amostra textual foi feita baseada em artigos na área de Linguística Aplicada, os textos foram coletados da biblioteca eletrônica russa *Cyberleninka*⁴. Essa biblioteca digital fornece, de forma legal, artigos revisados gratuitos de várias revistas científicas. Foram selecionados dez artigos no campo de humanidades, na área de Línguas e Literatura. Esses textos foram convertidos para .txt e arquivados, sendo os *abstracts* dessas produções acadêmicas usados como resumos de referência humana e o corpo do texto usados como fonte para produção dos resumos extrativos nos sites de sumarização selecionados. Para (Martins et al. 2001, p.7), supõe-se que o *authoring summarizing (abstract)* é escrito pelo próprio autor do trabalho que está sendo relatado: por conhecer o assunto, ele recorre a informações variadas, em relação à fonte.

A fim de tornar justa a comparação entre os sites, os sumarizadores foram utilizados com as taxas de compressão iguais no valor de 85%, ou seja, os resumos extrativos correspondiam a 15% do tamanho dos artigos originais, a escolha desse padrão tornou os sumários próximos em número de linhas do tamanho dos *abstracts* de referência. A taxa de compressão equivale ao tamanho da redução do extrato em relação ao texto-fonte, (Rino e Pardo, 2003, p.10) a define pela fórmula geral (TC = Taxa de Compressão): $TC = 1 - (\text{tamanho do extrato} / \text{tamanho do texto-fonte})$.

5. AVALIAÇÃO ROUGE 2.0

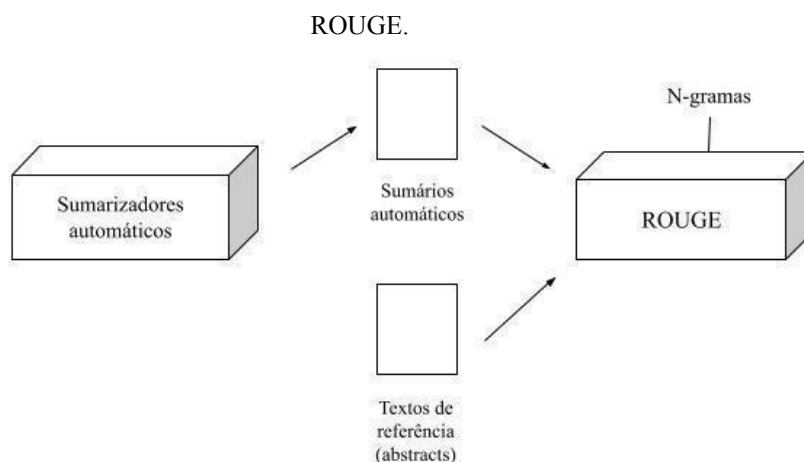
O produto final da seleção textual foram trinta textos em língua inglesa, sendo 10 *abstracts* dos artigos, 10 textos do sumarizador SMMRY e 10 textos do sumarizador

⁴ Disponível em: <https://cyberleninka.org/>

Tools4Noobs que foram submetidos ao ROUGE 2. A análise foi feita individualmente para cada sumariizador, um par de texto por vez (*abstract* e extrato de SA). Dessa forma, o programa gerou as métricas para a análise de cada resumo extrativo do artigo, tendo como referência o *abstract* deste mesmo texto.

Na figura 1, é possível ver o esquema simplificado do processo de avaliação do sumário extrativo no ROUGE-2.

Figura 1 — Fonte: adaptação de Esquema simplificado da avaliação de sumários utilizando-se a ferramenta



Fonte: Adaptação de Delgado, Vianna e Guelpele, 2010, p.6.

6. RESULTADOS

Foram obtidas as seguintes métricas no programa ROUGE 2.0 para os sumários extrativos obtidos pelo SMMRY (tabela 1):

Tabela 1— Resultados obtidos da avaliação ROUGE 2.0 dos resumos extrativos do sumariizador SMMRY.

	Recall	Precision	F-Score
Texto 1	0,38835	0,24242	0,29851
Texto 2	0,47929	0,29562	0,36569
Texto 3	0,42581	0,22917	0,29797
Texto 4	0,64078	0,37714	0,47482
Texto 5	0,95567	0,02388	0,04660

Texto 6	0,55789	0,18276	0,27532
Texto 7	0,40000	0,19293	0,26030
Texto 8	0,36000	0,11960	0,17955
Texto 9	0,91667	0,02280	0,22222
Texto 10	0,39073	0,29073	0,33456

Fonte: de autoria própria.

Quanto ao sumariador *Tools4Noobs*, sob as mesmas condições foram obtidas as seguintes medidas relativas ao sumários extrativos obtidos (tabela 2).

Tabela 2 — Resultados obtidos da avaliação ROUGE 2.0 dos resumos extrativos do sumariador *Tools4Noobs*.

	Recall	Precision	F-score
Texto 1	0,4077	0,13084	0,19811
Texto 2	0,76923	0,23050	0,35471
Texto 3	0,53548	0,20854	0,30018
Texto 4	0,50485	0,27660	0,35739
Texto 5	0,58621	0,52889	0,55607
Texto 6	0,51579	0,11369	0,18631
Texto 7	0,48667	0,14747	0,22636
Texto 8	0,40000	0,08386	0,13865
Texto 9	0,91667	0,02280	0,04450
Texto 10	0,38095	0,31655	0,34578

Fonte: de autoria própria.

Ao comparar os valores do F-score das tabelas e fazendo uma média simples da soma total desses valores para cada grupo de textos é possível perceber que o sumariador *SMMRY* obteve um sutil melhor resultado final em relação ao *Tools4Noobs* summarize, ou seja, os extratos são mais próximos da pontuação 1, o que indica que é um melhor sumário pelos parâmetros analisados.

Observa-se que já era previsível uma média de pontuações baixas uma vez que o sistema ROUGE espera que os resumos recuperem exatamente o conteúdo de referência (LLORET; PLAZA; AKER, 2017, p.8), entretanto é possível expressar uma mesma ideia de diferentes maneiras, lançando mão de sinônimos ou diferentes expressões, por exemplo.

7. CONCLUSÕES

A atividade de sumarização automática e avaliação de sumários é uma tarefa complexa. Os estudos desenvolvidos na linha de pesquisa “Línguas em contato na web” do projeto de extensão MOBILANG nessa área, bem como a exploração da ferramenta computacional ROUGE 2.0 se mostraram promissores uma vez que o pacote ROUGE 2.0 implementado teve um ótimo desempenho e permitiu efetivar a comparação e análise inicial proposta entre os sumarizadores de texto SMMRY e *Tools4Noobs*.

Todavia, o método de avaliação utilizado sozinho não é suficiente para se obter uma resposta taxativa sobre a qualidade de um resumo. Dessa forma, há a necessidade de estudos posteriores nessa frente de trabalho sobre outros métodos de avaliação a fim de validar e complementar a pesquisa nesse tema, como os métodos da pirâmide (NEKOVA e PASSONNEAU, 2004 apud PARDO, 2008), o dos elementos básicos (HOVY et al., 2006 apud PARDO, 2008, p.8), ou ainda, cálculos a partir de testes estatísticos ANOVA de Friedman e do coeficiente de concordância de Kendall usados nas conferências dedicadas à sumarização automática (DELGADO; VIANNA; GUELPELI, 2010, p.8).

REFERÊNCIAS

- ASSOCIATION “OPEN SCIENCE”. **CyberLeninka Online Library**, 2021. Página inicial. Disponível em: <<https://cyberleninka.org>>. Acesso em: 10 de maio de 2021.
- DELGADO, Carlos Henrique; VIANNA, Caroline Evangelista; GUELPELI, Marcus Vinicius. **Comparando sumários de referência humanos com extratos ideais no processo de avaliação de sumários extrativos**. Centro Tecnológico, Universidade Federal dos Vales do Jequitinhonha, Minas Gerais, 2010.
- ESPINA, A.P; RINO, L.H.M, **Utilização de Métodos Extrativos na Sumarização Automática de Textos**. Relatórios Técnicos (NILC-TR-02-06). NILC – ICMC – USP. São Carlos, Brasil, 2002.
- GANESAN, Kavita. **ROUGE: A Package for Automatic Evaluation of Summaries**. 8p. 2015.

GUELPELI, Marcus Vinicius Carvalho, **Cassiopeia: Um Modelo de Agrupamento de Textos Baseado em Sumarização**. Tese (Doutorado em Computação) - Niterói: Universidade Federal Fluminense, 2. MARTINS, Camilla Brandel, 2012.

GUELPELI, Marcus Vinicius Carvalho; BERNARDINI, Flavia Cristina; GARCIA, Ana Cristina Bicharra. **Todas as Palavras da Sentença como métrica para um Sumarizador Automático**. 5 f. Departamento de Ciência da Computação, Universidade Federal Fluminense - Uff, São Domingos - Niterói - Rj, 2008.

LIN, C-Y. and HOVY, E.H., **Automatic Evaluation of Summaries Using N-gram Cooccurrence Statistics**. In the Proceedings of the Language Technology Conference. Edmonton, Canada, 2003.

LLORET, Elena; PLAZA, Laura; AKER, Ahmet. **The Challenging Task of Summary Evaluation: an overview**. *Lang Resources & Evaluation* 52, p.101–148 2018. Disponível em: <https://doi.org/10.1007/s10579-017-9399-2>. Acesso em: 10 de Maio de 2021.

MARTINS, C.B., PARDO, T. A. S., ESPINA A., e RINO, L. H. M. "**Introdução à Sumarização Automática**". Relatório Técnico do Departamento de Computação da UFSCar RT-DC 002/2001. UFSCar. 2001.

SMMRY. **SMMRY**, 2021. Summarize (página inicial). Disponível em: <<https://smmry.com/>>. Acesso em: 10 de Maio de 2021.

PARDO, T.A.S. **Sumarização Automática: Principais Conceitos e Sistemas para o Português Brasileiro**. Relatórios Técnicos (NILC-TR-08-04). NILC – ICMC – USP São Carlos, SP, Brasil, 2008.

PARDO, T.A.S. RINO, L.H.M., **TeMário: Um Corpus para Sumarização Automática de Textos**. Relatórios Técnicos (NILC-TR-03-09). NILC – ICMC – USP. São Carlos, Brasil, 2009.

TOOLS4NOOBS. **Tools4Noobs summarize**, online summarize tool (free summarizing), 2020. Disponível em: <<https://www.tools4noobs.com/summarize/>>. Acesso em: 10 de Maio de 2021.

