



STILINGUE



**SNLP: por trás do motor de
Processamento de Linguagem
Natural para Português-Brasil
da STILINGUE, um dos mais
consistentes e confiáveis do
mercado**



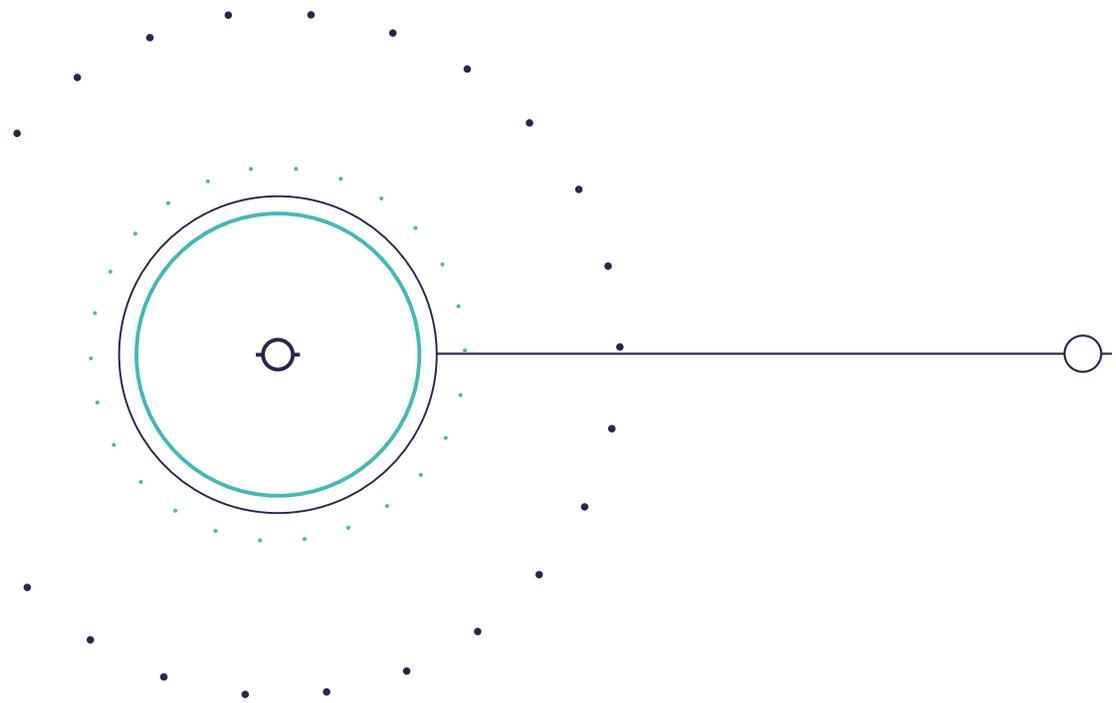
Este documento se propõe a elucidar a respeito do experimento de mensuração de acurácia de análise de sentimento entre os diferentes motores de Processamento de Linguagem Natural (NLPs) para análise do idioma Português-Brasil disponíveis no mercado.

Foram comparados os desempenhos da SNLP, NLP proprietária da STILINGUE, com outras seis tecnologias **de outros players do mercado - cujos nomes serão anonimizados e chamaremos neste estudo de** “Empresa1”, “Empresa2”, “Empresa3”, “Empresa4”, “Empresa5” e “Empresa6”.

Será abordada metodologia por trás do estudo, fases e experimentação, resultados e conclusões do projeto.

Este benchmark se destina a profissionais da área de tecnologia e Inteligência Artificial; a entusiastas sobre o tema, e a gestores e líderes que buscam soluções avançadas para otimização de operações, redução de custos e melhores índices de satisfação do cliente.

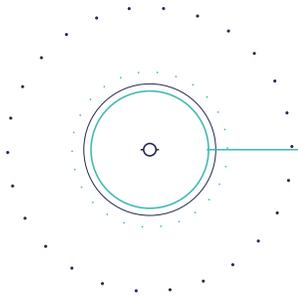
- 03** Objetivo
- 04** Contexto
- 05** Metodologia
- 06** Experimentos
- 23** Conclusões



Objetivo

Realizar uma análise comparativa do desempenho de acurácia de qualificação de sentimento (positivo, neutro ou negativo) entre a SNLP da STILINGUE e NLPs de outras APIs com foco no Português-Brasil.





CONTEXTO

Processamento de Linguagem Natural (ou *Natural Language Processing*) é o nome dado ao estudo e aplicação de tecnologias no entendimento e interpretação às linguagens humanas, por meio dos idiomas falados pelas diversas culturas.

Dessa forma, um diferencial dos sistemas de NLP é a garantia de uma melhor compreensão para além do sentido literal de cada palavra:

- Contexto da conversa
- Interpretação de texto
- Resumo de conteúdos
- Significados sintáticos e semânticos
- Análise de sentimentos

Existem diversas aplicações que integram NLP ao sistema para garantir uma comunicação cada vez mais eficiente entre máquinas e usuários: ferramentas de social listening, chatbots, assistentes virtuais, sistemas de correção automática, filtros de spam em e-mails, plataformas de buscas online, etc...

Em comum a estas aplicações, estão os objetivos de otimizar operações, reduzir custos, ampliar satisfação do cliente, estreitar relacionamento, facilitar tarefas, entre outros.

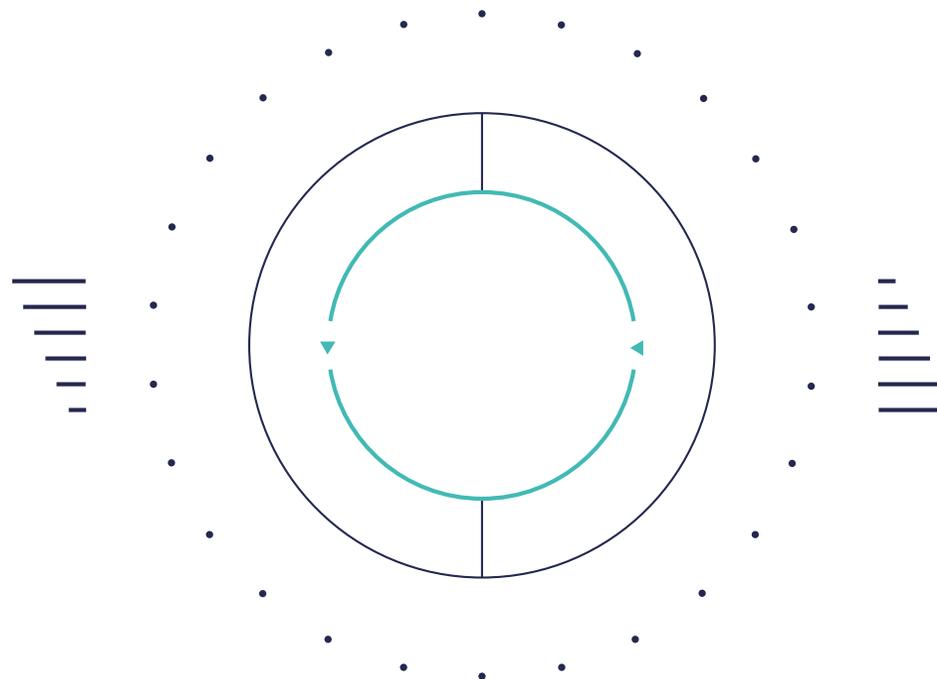


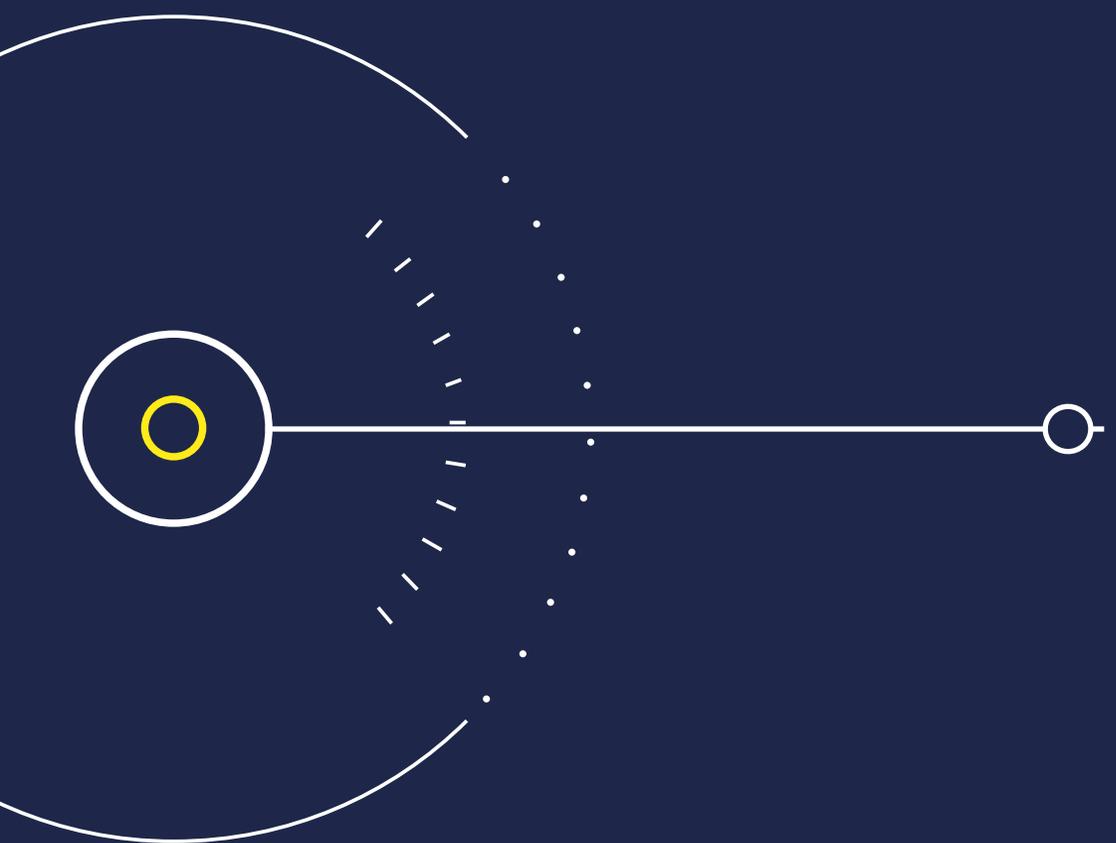
METODOLOGIA

Para essa análise, utilizamos como referência o modelo e estrutura de benchmark realizado pela Monkeylearn para comparativo de sentimento entre APIs.

Ele distribui o processo de experimentação em 4 etapas:

- (1) coleta;
- (2) rodar chamadas nas APIs;
- (3) análise de desempenho;
- (4) comparativo de resultados.





Esperimento

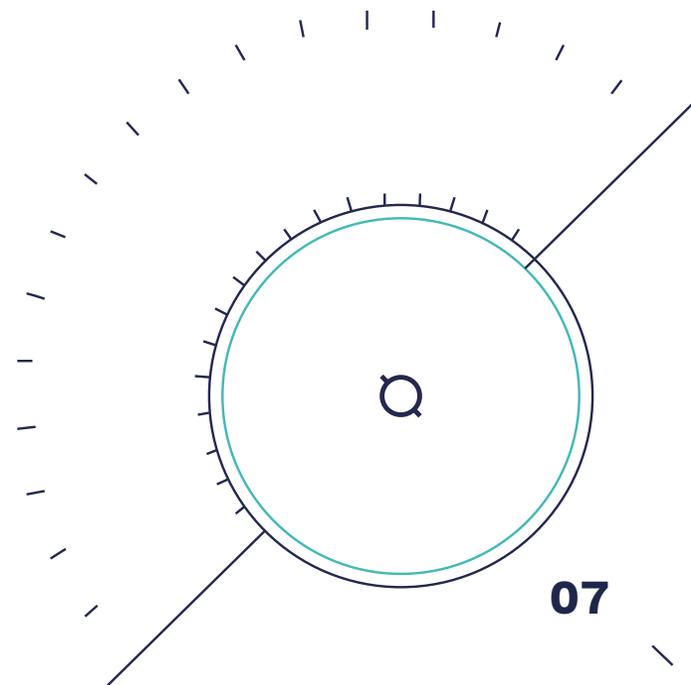


Fase 1

Uma amostra aleatória de 4500 posts foi retirada de um conjunto de mais de 20 mil posts coletados por uma ferramenta externa.

Os posts foram pré-processados para eliminar repetições e posts sem conteúdo de texto.

A amostragem se organizou em 6 áreas balanceadas, cada uma com 750 postagens: Educação, LGBTQ+, Moda, Beleza, Turismo e Práticas Alimentares.





Todas as postagens foram anotadas manualmente por nossa equipe de machine teaching usando três etiquetas de anotação de sentimento (Positivo, Negativo e Neutro).

Em paralelo, os mesmos dados foram anotados automaticamente por 6 sistemas distintos: Empresa 1, Empresa 2, Empresa 3, Empresa 4, Empresa 5, Empresa 6 e STILINGUE (SNLP).

Em todos os casos, quando os sistemas de anotação incluíam outras categorias de anotação de sentimento (como *mixed*) ou então retornavam um número real ao invés de uma categoria, estabelecemos regras de transposição para o sistema de três etiquetas de forma a aumentar as taxas de acerto desses sistemas.



Fase 2



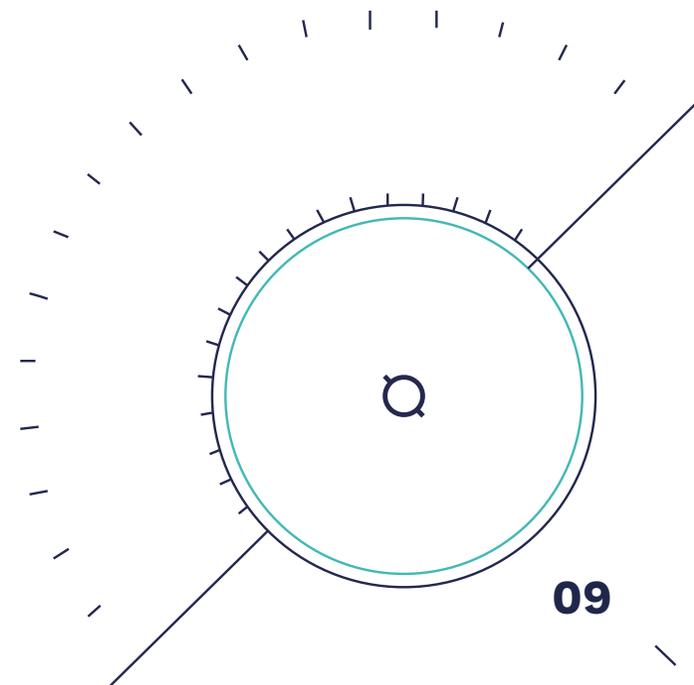
Por fim, a anotação manual foi considerada o gabarito da tarefa de anotação de sentimento. É importante salientar que a anotação foi feita por anotadores experientes de acordo com metodologia que direciona e busca padronizar a execução da anotação.

Uma outra dimensão incluída nas análises foi a de dificuldade da anotação de sentimento. Os posts foram divididos em 3 categorias de dificuldade de anotação (fácil, médio e difícil) a partir de uma matriz de confusão da anotação automática e posterior revisão manual. Essas três categorias indicam o quanto o sentimento percebido é explícito ou implícito, a ideia é que posts fáceis sejam mais explícitos e posts difíceis mais implícitos.

Fácil - Entre 0 e 2 anotadores erraram

Médio - Entre 3 e 4 anotadores erram

Difícil - Entre 5 e 7 anotadores erram



Fase 2

Os 4500 dados estão organizados da seguinte forma pelas categorias de dificuldade:

Dificuldade	Quantidade de posts
Difícil	1470
Médio	2340
Fácil	690



Fase 2



A partir disso, seguem os resultados quantitativos obtidos e depois algumas análises consolidadas pelo Machine Teacher responsável.

Distribuição de labels, quantidade de posts, positivo, negativo e neutro nos datasets:

	Positivo	Neutro	Negativo	
Educação	85	271	394	750
Beleza	252	377	121	750
Moda	281	391	78	750
Práticas alimentares	378	195	177	750
LGBTQ+	204	297	249	750
Turismo	251	358	141	750
	1451	1889	1160	4500





3.1 Visão quantitativa – desempenho geral de acurácia

O quanto o motor de NLP acertou na classificação de sentimento dos conteúdos?

Desempenho geral	Acertos	Erros	% de acertos
SNLP	2729	1771	60,64%
Empresa 1	2652	1848	58,93%
Empresa 5	2528	1972	56,18%
Empresa 6	2286	2214	50,80%
Empresa 3	2012	2488	44,71%
Empresa 2	1939	2561	43,09%
Empresa 4	1495	3005	33,22%





3.2 Visão quantitativa – desempenho por área

O quanto o motor de NLP acertou na classificação de sentimento dos conteúdos de acordo com as áreas analisadas?

Desempenho por área	LGBTQ+	Moda	Beleza	Turismo	Educação	Práticas alimentares
SNLP	57,87%	60,40%	54,67%	57,20%	78,80%	54,93%
Empresa 5	55,60%	51,87%	43,87%	50,93%	72,40%	62,40%
Empresa 1	52,93%	62,93%	62,67%	62,67%	64,93%	47,47%
Empresa 3	44,13%	42,13%	41,87%	46,40%	40,00%	53,73%
Empresa 6	40,40%	59,73%	43,47%	50,00%	68,00%	43,20%
Empresa 2	39,73%	55,60%	31,07%	33,33%	59,33%	39,47%
Empresa 4	26,53%	34,67%	28,27%	29,60%	45,73%	34,53%





3.3 Visão quantitativa – desempenho por polaridade

O quanto cada anotador acerta dependendo da categoria de sentimento?

Porcentagem de acertos por categoria de sentimento	Positivo	Neutro	Negativo
SNLP	69,12%	48,65%	69,57%
Empresa 5	80,77%	30,60%	67,07%
Empresa 1	46,52%	85,39%	31,47%
Empresa 3	92,14%	14,93%	33,88%
Empresa 6	56,51%	55,32%	60,17%
Empresa 2	22,33%	30,23%	90,00%
Empresa 4	64,92%	32,82%	24,31%





3.4 Visão quantitativa

Os anotadores têm algum viés em relação à categoria de sentimento?

Viéses de posts etiquetados por categoria de sentimento	Positivo	Neutro	Negativo
SNLP	7,67%	-10,49%	2,82%
Empresa 5	18,76%	-23,69%	4,93%
Empresa 1	-10,00%	25,62%	-15,62%
Empresa 3	40,16%	-26,27%	-13,89%
Empresa 6	0,09%	0,11%	-0,20%
Empresa 2	-19,31%	-21,22%	40,53%
Empresa 4	23,73%	-7,47%	-16,27%





3.5 Visão quantitativa

Desempenho por nível de dificuldade dos posts

Desempenho por dificuldade	Fácil	Médio	Difícil
SNLP	98,26%	76,50%	17,76%
Empresa 5	99,13%	68,63%	16,19%
Empresa 1	89,57%	56,45%	48,50%
Empresa 3	91,30%	49,57%	15,10%
Empresa 6	97,54%	63,85%	8,10%
Empresa 2	67,54%	45,68%	27,48%
Empresa 4	84,49%	35,64%	5,31%





Resultados

Em comparativo de desempenho por área de análise, o SNLP esteve sempre entre as 2 empresas com mais assertividade na classificação;

SNLP registrou uma média de acurácia de classificação por categoria de sentimento de 62,45%, 2,97 pontos percentuais acima do segundo colocado (Empresa 5 - 59,48%);

A Empresa 3 apresentou forte viés para posts etiquetados na categoria positiva (40,16%), ou seja, atribuiu sentimento positivo para 40% de posts a mais do que os anotadores humanos, e a Empresa 2 na categoria negativa (40,53%). Já a Empresa 1, registrou viés de 25,62% para posts etiquetados na categoria neutra;

A SNLP foi a segunda tecnologia com menores índices de enviesamento de classificação por categoria de sentimento, perdendo apenas para a Empresa 6;

A Empresa 5 apresentou maior índice de assertividade em conteúdos qualificados como fáceis (99,13%). Enquanto, a Empresa 1 computou 48,5% nos difíceis;

A SNLP esteve entre os 3 melhores resultados nas 3 categorias de dificuldade: segunda colocada entre os fáceis (98,26%), primeira colocada entre os médios (76,5%) e terceira colocada entre os difíceis (17,76%);



Exemplos de conteúdos analisados

Fase 4

POSTS POR ÁREA



Práticas Alimentares

Follow

*Olha só a @XXX com o nosso Kit Imunidade! 🍯
Nutrir o corpo e a alma é parte de uma rotina diária de cuidados que precisam andar juntos! 🙏 Essa união ganha muito mais importância agora que a gente tá entrando na fase mais fria do ano. Fazendo aquele gás a mais na imunidade ser ESSENCIAL! 😊
E aí? Bora se aquecer e se proteger?!*

❤️ 1.4k 💬 800



Moda

Follow

Você vende moda evangélica e ainda não tem os produtos da [marca] 😞? Vendas somente no atacado. Chame no direct!!
#rosarrara #modaprimaverao #rendaextra #rendaextraemcasa #modaexecutiva #modaevangélica #atacadoderoupas #lojaderoupas #roupaprarevenda #boutique #revendedora #revendedoras #modaevangélica #modafemina #roupasevangelicas #estilo #modagospel

❤️ 1.4k 💬 800



*Dados de engajamento são meramente ilustrativos



Fase 4



Beleza

Follow



Meu Deus preciso comprar maquiagem

1.4k 800



Educação

Follow



Quando vão disponibilizar algo para os alunos que foram aprovados no Fies levem os documentos até a faculdade? Vcs adiaram o prazo para 24 de Abril, e agora está aparecendo que o prazo foi até 16 março! Por favor, conNeuro. Novo tema: Nota explicativa: A estigmatização dos portadores de HIV - Link do tema: #TemasdeRedação #Redação #ENEM #ENEM2020 #Corrijame #RedaçãoENEM certem isso!

1.4k 800



LGBTQ+

Follow



Essa é uma vitória para a luta por mais igualdade no nosso país! O STF decidiu que é inconstitucional proibir doação de sangue por homens gays. A importante decisão vem em um momento em que vemos estoques baixos em várias partes do Brasil! Seguimos na luta 🙌

1.4k 800



Turismo

Follow



@XXX Mario Henrique, sabemos que você precisa de ajuda. Por isso, acesse Minhas Viagens para alterar ou cancelar sua reserva. Se você já fez o pedido, por favor aguarde nossa resposta por e-mail. E se tiver mais perguntas, veja mais informações em <https://t.co/XBVKtv3PqN>

1.4k 800





Fase 4



Positivo

Follow



Podem me dar maquiagem de presente, eu adoro

#Fuvest #Unesp #BarroBranco

1.4k 800



Negativo

Follow



@XXX @XXX Essa é a pior, na própria reportagem de hoje dava pra notar que ela escureceu a maquiagem pra estar o mais escura possível pro debate. Que vergonha heim!!

1.4k 800





Positivo

Follow



Já que vocês não estão respeitando a "quarentena" pelo menos ajude as pessoas que estão fazendo de tudo pra tirar um dinheiro nessa crise. Comprem bolo de pote das amigas, roupas, maquiagem. AJUDEM. É o mínimo que podemos fazer

1.4k 800



Negativo

Follow



RT @XXX: Uma moda de maquiagem que todos os dias eu rezo para abandonarmos é essa porra de afinar o nariz

1.4k 800



Neutro

Follow



Sou fã de kamaitachi, já ouviu cabelos arco íris?

1.4k 800





Positivo

Follow



*tentei fazer challenge mas eh mto difícil parabéns blogueiras.
ps: não foi de maquiagem* <https://t.co/DeQ65pRGhp>*

1.4k 800



Negativo

Follow



*eu fiquei 2 HORAS GRAVANDO TIKTOK, fiz uma maquiagem maravilhosa.
o vídeo ficou uma bosta. ai eu pensei, pra recompensar vou tentar tirar
umas fotos aqui né.
MEU DEUS AS FOTOS FICARAM HORROROSAS*

1.4k 800



Neutro

Follow



*Live trata dos cuidados com a mente e o corpo durante
pandemia - <https://t.co/mW7NN0GHU5>*

1.4k 800





CONCLUSÕES

SNLP é o motor com desempenho mais consistente nas diferentes análises;

Comprovação da eficiência e diferencial da SNLP frente às outras APIs focadas em PT-BR com uma média de acurácia de classificação por categoria de sentimento de 62,45%, 2,97 pontos percentuais acima do segundo colocado (Empresa 5 - 59,48%);

Em conteúdos com classificação nível médio de dificuldade, a SNLP registrou um desempenho significativamente superior ao dos demais motores, com 8% a mais do segundo colocado;

Evidência de qualidade em NLP feita no Brasil desenvolvida pela STILINGUE e com especialidade no entendimento e interpretação da linguagem local, como: gírias, palavras e expressões próprias, internetês, ortografia, pontuação, polaridade por categoria gramatical, entre outros recursos;

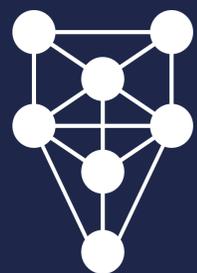




Nossos Diferenciais

Diferencial em dedicação de equipe de especialistas para aperfeiçoamento contínuo da SNLP: mais de 14 profissionais especialistas, sendo 1 Doutor, 4 Mestres e 2 Bacharéis em Ciência da Computação; 4 profissionais de Letras especialistas em Machine Teaching, 2 Engenheiros de Controle e Automação e 1 Doutorando em Bio Informática.





STILINGUE



www.stilingue.com.br