

# Introdução

---

Analizamos 3 dos maiores fornecedores de Machine Translation no mercado: Microsoft, Google e Amazon. Todos possuem planos que cobram por palavra ou caractere traduzido, sendo o mais comum o caractere. O ticket médio varia para todos, porém para traduções com alta densidade de dados todos solicitam negociação. Antes de prosseguirmos para os valores, é preciso entender como a maneira com que esses fornecedores precificam seus serviços afetam nossa estratégia de aplicação dessa tecnologia.

Consideremos a cobrança por caractere. Isso quer dizer que cada um dos serviços mencionados acima cobram um valor X por Y caracteres traduzidos. Por exemplo, Microsoft cobrará US\$40 a cada 1 milhão de caracteres traduzidos (favor consultar o documento com os valores atualizados), e conforme o montante de caracteres traduzidos aumentam, regras de descontos vão sendo aplicadas. Considerando o fluxo de informação no Chat Bot, sabemos que a informação do Bot é constante, ou seja: o bot tem sempre as mesmas falas (cadastradas pela equipe de conteúdo), assim o conteúdo variável é o do usuário que responderá como bem entende. Dito isso, desconsideramos a tradução em tempo real pelos seguintes motivos:

1. Se ligarmos a fala do Bot ao Machine Translation em tempo real, podemos incorrer em duplicação de serviço, isso porque toda vez que o Bot disser algo, essa frase passará pelo machine translation, sofrerá a cobrança por caractere e será traduzida. Como o bot diz sempre as mesmas coisas, estaremos pagando de forma duplicada.
2. Uma vez que o bot esteja traduzido (conforme estratégia que demonstraremos abaixo), não há necessidade de traduzir o conteúdo inserido pelo usuário, uma vez que este deve ser fluente ou nativo no idioma em que escolheu.

Sendo assim, tomamos uma estratégia afim de economizar evitando tradução recorrente ou desnecessária.

## Estratégia

---

Cada Bot possui sua base conversacional cadastrada que define as capacidades "cognitivas" e analíticas do modelo, ou seja: aquilo que ele é capaz de entender e responder. Em nossa análise, concluímos que o melhor caminho é utilizar a API de serviço de Machine Learning para criar o nosso "tradutor de base de dados de bot". A ideia é que tenhamos uma base de dados num idioma X (inglês por exemplo), e possamos usar machine learning para traduzir a base inteira para outro idioma de escolha. Assim, estaremos pagando exatamente pelo número de caracteres que utilizamos realmente, e não pelo número de vezes que um bot disse a mesma frase e esta precisou ser traduzida. Uma vez traduzida a base de dados ela deve passar por uma revisão, de preferência por um nativo, e então ser inserida no novo Bot com o novo idioma.

Esta estratégia visa evitar qualquer uso excessivo, desnecessário ou duplicado das capacidades de Machine Learning. Caso mudanças posteriores sejam feitas na base de conhecimento do bot, reaplicaremos o mesmo motor de Machine Translation para atualizar a tradução corrente.

## A base de dados

---

A base conversacional customizada de um chat bot possui em média 185.871 caracteres. Esse valor pode variar para mais ou para menos, porém nos dá a média necessária para projetarmos os custos com machine learning caso sigamos a estratégia aqui apresentada. Junto deste report temos um documento detalhando os valores para cada uma das alternativas pesquisadas. Em resumo, são os seguintes:

### MICROSOFT

US\$10 para cada milhão de caracteres traduzidos em modo standard, ou US\$40 por milhão de caracteres traduzidos em modo customizado sendo 2 mil caracteres gratuitos por mês.

Source: <https://azure.microsoft.com/en-us/pricing/details/cognitive-services/translator/>

### Google

US\$20 por milhão de caractere traduzido cobrado a partir de 500.000 caracteres até 1 bilhão. Após 1 bilhão de caracteres traduzidos os preços devem ser negociados.

Source: <https://cloud.google.com/translate/pricing>

### Amazon

US\$15.00 por milhão de caracteres em modo standard, e \$60.00 por milhão de caracteres em modo Custom, sendo os 2 primeiros milhões gratuitos por mês nos 12 primeiros meses.

Source: <https://aws.amazon.com/translate/pricing/>

\*\* Para todos os valores acima consideramos primeiramente o modo custom.

## Competencia do Fornecedor

---

Não pudemos efetuar testes profundos, porém todos possuem aproximadamente o mesmo grau de competencia, tendo pequenas variações. Os três serviços apresentados possuem cases e clientes confiáveis que utilizam o serviço, sendo alguns inclusive utilizados por possíveis fornecedores Sitel, como é o caso da InBot que utilizado Google Translate e disse estar contente e sem reclamações por parte dos usuários. Todos os fornecedores também possuem APIs e possibilidade de customização da tradução e base de dados. Dito isso, o norteador da análise permanece sendo o valor e a flexibilidade de implementação.

---

Att

TSC Development Team