



SIL Translation QA Tool

Team Introduction



Subhashree Chowdhury



Ryan Egbert



Poorna Narasimhan

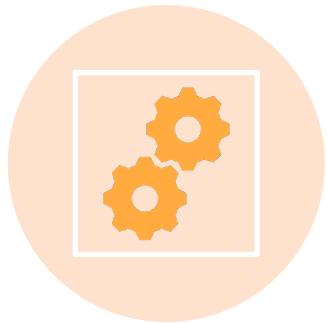


Suyash Sukthankar



Business Problem

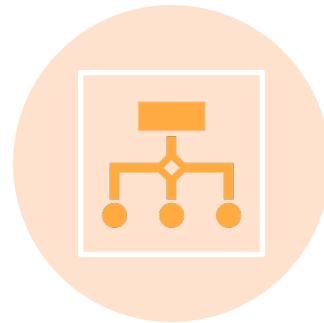
Business Problem



No quality checking application presently available



In cases such as medical translations quality assessment is critical

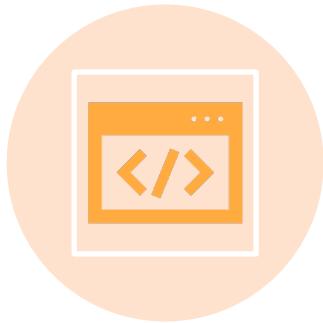


Without a quality assessment no benchmark can be made for the future

Business Benefits



One stop destination for all quality assurance testing and records



Easy to use for both technical and non-technical users



Wide range of metrics that give an overall quality assessment

Business to Analytics Problem Framing

Determine if the translation is accurate and is comprehensive.

Determine the effect of the GUI based translation tool.



What is accuracy of the translations?



Calculate the different scores for the translation like comprehensibility, readability etc.



Is there any part of the text that has a worse score as compared to others?

Real time user interaction wherein the user can ask questions to comprehend the translation in a better fashion



Business Specifications for SIL Translation QA Tool

Business Specifications

As a user:

- Quality assessment using a range of scores
- Ability to assess smaller sections within larger texts
- Usage of different metrics – Similarity, Comprehensibility, Readability for quality check
- Ability to take ‘reference’ and ‘translations’ through text input and files
- Flexibility to select and view only chosen metrics

Similarity	Indicates similarity of semantics between texts
Comprehensibility	Allows question answer interactions to comprehend translations
Readability	Provides difficulty level of reading translated text

Business Specifications

As a consultant:

- Comparison of translations across different metrics
- Ability to see visual charts and dashboard for metrics
- Ability to provide additional input within the post-processed view for metrics (Eg. Comprehensibility)

As an admin:

- Ability to see information on aggregated scores across multiple metrics
- Visually identify outliers in translations in one or more metrics

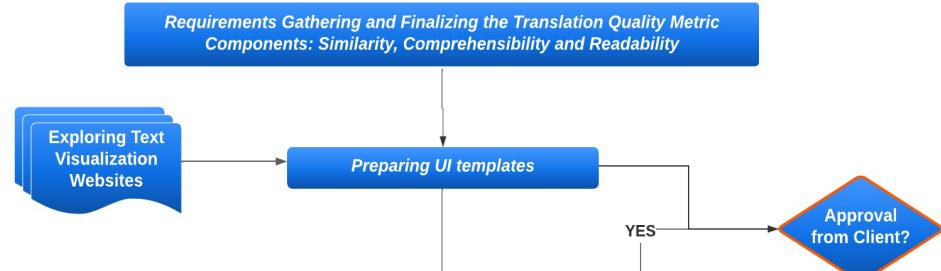




Methodology

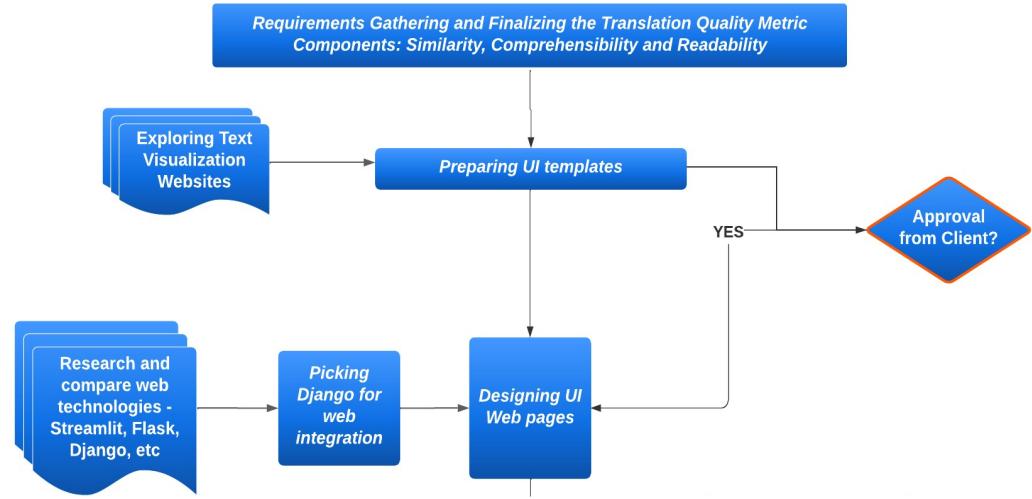
Methodology

- Deciding translation quality metrics
- Preparing UI designs



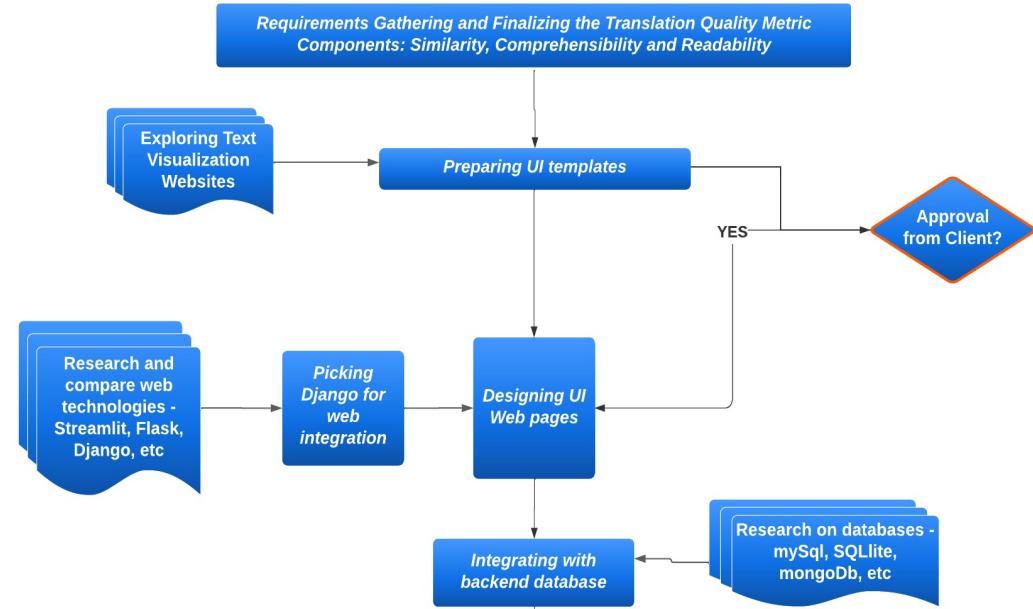
Methodology

- Deciding translation quality metrics
- Preparing UI designs
- Building interactive web-pages



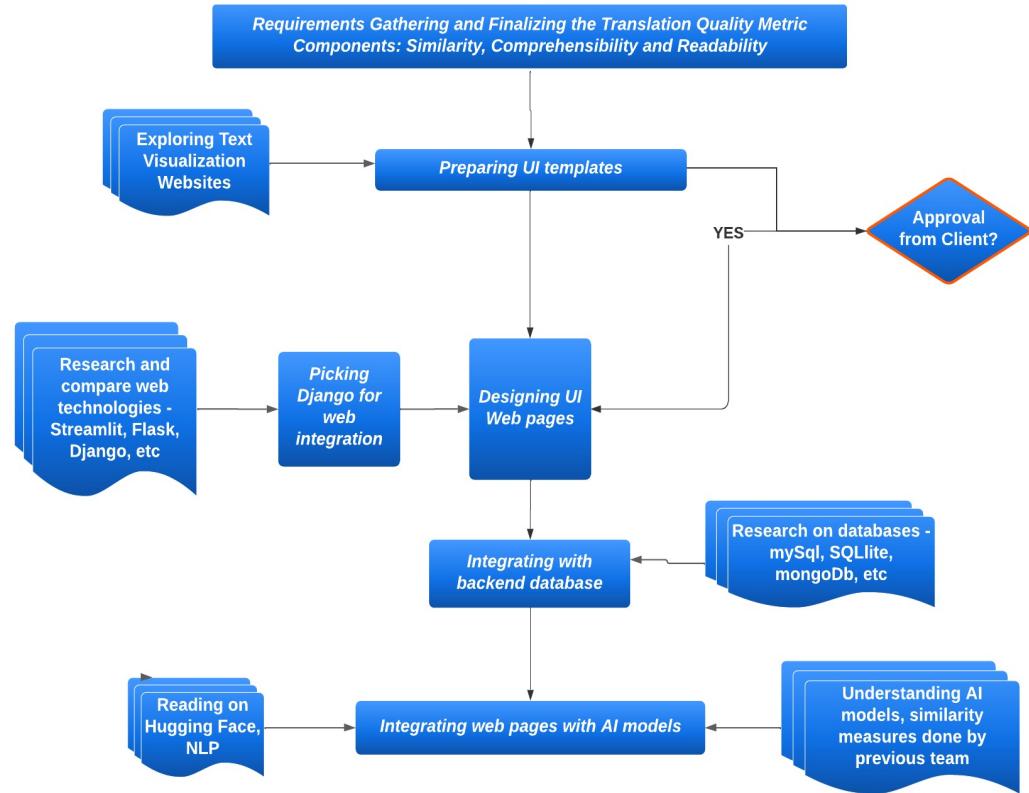
Methodology

- Deciding translation quality metrics
- Preparing UI designs
- Building interactive web-pages
- Hosting backend data storage



Methodology

- Deciding translation quality metrics
- Preparing UI designs
- Building interactive web-pages
- Hosting backend data storage
- Implementing AI models

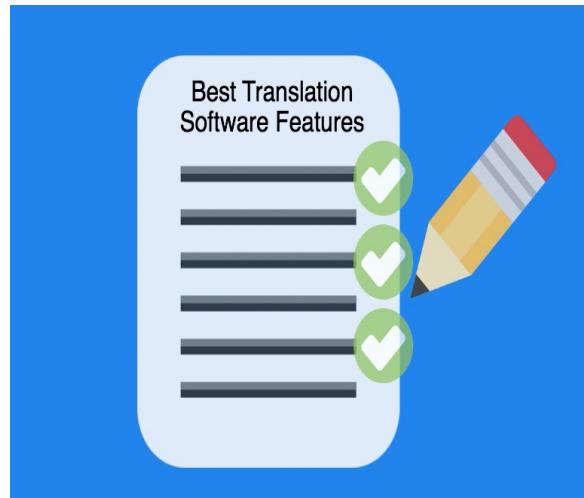




Performance Measures for SIL Translation QA Tool

Performance Measures

- How easily is the user able to see lowest, highest and other scores for selected text?
- How quickly do the texts get highlighted for different range of scores?
- How easily do the user gets to see the overall presentation of dashboards, scores and highlighted translation texts once he selects the range of reference texts?
- How quickly is the user able to provide real-time additional inputs and get the answers?





User Interface

Home Page



SIL Translation Quality Tool

Upload Translations

About Quality Metrics ▾

About the Project

howdydoody ▾

Welcome!

This translation quality checking tool aims to aid translators, consultants, and administrators in creating high-quality and comprehensive translations.

Any feedback is welcome. Please use the contact information found below to let the developers know.

Register

Translation Tool

The tool that was built utilizes AI models to determine the traits of a translation. These traits include semantic similarity, comprehensibility, and readability. Semantic similarity can be described as how close the different words and phrases in a source text are to those of the translated text. Comprehensibility uses question and answer pairs to ensure that information is being captured in a translated text. Readability determines the complexity of a sentence. A high readability score would indicate a piece of

Upload Page



SIL Translation Quality Tool

Upload Translations

About Quality Metrics ▾

About the Project

howdydoody ▾

Upload a File

Choose file Browse

Manually enter translations

Semantic Similarity Comprehensibility Readability Semantic Domains

Source text goes here

Translated text goes here

Upload Page

Two different ways to upload translations

1. Via a file
 - i. Specify source and translated text
 - ii. Ask questions and provide expected answers for comprehensibility

source	translated	question	answer
Ondanks wat je leraar je misschien heeft verteld, is er een verkeerde manier om een „Ää, Äälasso te hanteren.	Despite what your teacher may have told you, there is a wrong way to wield a lasso.	Is there a wrong way to wield a lasso?	Yes
Ze hadden dringend een andere drummer nodig aangezien de huidige alleen bongo's kon spelen.	They desperately needed another drummer since the current one only knew how to play bongos.		
Je weet niet zeker of je hem wel of niet moet vertrouwen, maar erg dankbaar dat je een coltrui droeg.	You're unsure whether or not to trust him, but very thankful that you wore a turtle neck.		
Als je van tonijn en tomatensaus houdt, probeer de twee dan te combineren, het is echter niet zo erg als het klinkt.	If you like tuna and tomato sauce, try combining the two, it is really not as bad as it sounds.		
Hij dronk het leven voordat hij het uitspuugde.	He drank life before spitting it out.		
Op dat moment leerde hij dat er bepaalde delen van het lichaam zijn die je	It was at that moment that he learned there are certain parts of the body that you		

Upload Page

Two different ways to upload translations

2. Manual Entry
 - i. Specify source and translated text
 - ii. Specify quality metrics to be used

The screenshot shows the SIL Translation Quality Tool interface. On the left, there is a file upload form titled "Upload a File" with "Choose file" and "Submit" buttons. On the right, there is a section titled "Manually enter translations" with three checked checkboxes: "Semantic Similarity", "Comprehensibility", and "Readability". Below these checkboxes is a list of French poems followed by their English translations. At the bottom of the page is a "Submit" button.

French Poem	English Translation
dire ce que je pensais. Enfoui profondément dans la neige, il espérait que ses batteries étaient fraîches dans sa balise d'avalanche.	I never knew what hardship looked like until it started raining bowling balls.
Il fouilla dans l'obscurité à la recherche de l'interrupteur, mais quand il le trouva finalement, quelqu'un était déjà là.	I was starting to worry that my pet turtle could tell what I was thinking.
C'est une carte au trésor attrayante que je ne peux pas lire.	Buried deep in the snow, he hoped his batteries were fresh in his avalanche beacon.
J'avais l'habitude de pratiquer le tissage avec des spaghetti trois heures par jour mais j'ai arrêté parce que je ne voulais pas mourir seule.	He fumbled in the darkness looking for the light switch, but when he finally found it there was someone already there.
Vous avez parfaitement le droit d'être en colère, mais cela ne vous donne pas le droit d'être méchant.	That is an appealing treasure map that I can't read.
La mémoire que nous avions l'habitude de partager n'est plus cohérente.	I used to practice weaving with spaghetti three hours a day but stopped because I didn't want to die alone.
Je serais ravi si la mer était pleine de jus de concombre.	You have every right to be angry, but that doesn't give you the right to be mean.
Malgré ce que votre professeur vous a peut-être dit, il existe une mauvaise façon de manier un lasso.	The memory we used to share is no longer coherent.
C'est la rareté qui a alimenté sa créativité.	I would be delighted if the sea were full of cucumber juice.
Soyez prudent avec ce couteau à beurre.	Despite what your teacher may have told you, there is a wrong way to wield a lasso.
Toutes les personnes qui errent ne sont pas perdues.	It was the scarcity that fueled his creativity.

Upload Page

SIL Translation Quality Tool Upload Translations About Quality Metrics ▾ About the Project sampleuser ▾

Upload a File

test_input_questions3.csv Browse

Submit

Manually enter translations

Semantic Similarity Comprehensibility Readability

Source text goes here

Translated text goes here

Processing...

Similarity:
C C

Comprehensibility:
C C

Readability:
C

Submit

Results Page

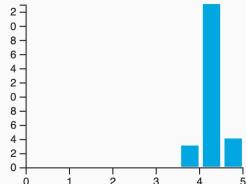


SIL Translation Quality Tool

[Upload Translations](#)[About Quality Metrics](#) ▾[About the Project](#)

sampleuser ▾

Translation ID - 27535

[◀ To Main Results](#)[Similarity](#) [Comprehensibility](#) [Readability](#) [Reset](#)

भीड़ चिल्लाती है और अधिक memes के लिए चिल्लाती है।
वे विमान में उड़ते हुए पियानो बजा रहे हैं।
सिर्फ इसलिए कि पानी लाल है इसका मतलब यह नहीं है कि आप
इसे नहीं पी सकते।
अपनी सुनहरी मछली को साइकिल दिलाकर उसकी शारीरिक
फिटनेस में सुधार करें।
उस समय वह संगीत नहीं सुन रहा था, वह एक अनुभव जी रहा था।

आइसक्रीम ट्रक हम सभी के लिए बुरी यादें वापस लाते हैं।

जैसे ती चर्चा में गाते। किंगा चर्चे किंगी के घोल घोल में

The crowd yells and screams for more memes.
They're playing the piano while flying in the plane.
Just because the water is red doesn't mean you can't
drink it.

Improve your goldfish's physical fitness by getting him a

bicycle.

At that moment he wasn't listening to music, he was
living an experience.

The ice-cream trucks bring back bad memories for all of
us.

As he entered the church he could hear the soft voices of

Similarity Score

Score:

Results Page



SIL Translation Quality Tool

Upload Translations

About Quality Metrics ▾

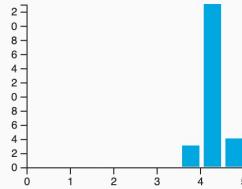
About the Project

sampleuser ▾

Translation ID - 27535

[← To Main Results](#)

Similarity Comprehensibility Readability Reset



इसे नहीं पा सकते।

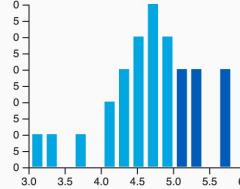
अपनी सुनहरी मछली को साइकिल दिलाकर उसकी शारीरिक फिटनेस में सुधार करें।

उस समय वह संगीत नहीं सुन रहा था, वह एक अनुभव जी रहा था।

आइसक्रीम ट्रक हम सभी के लिए बुरी यादें वापस लाते हैं।

जैसे ही उन्होंने चर्च में प्रवेश किया, उन्हें किसी के सेल फोन में फुसफुसाते हुए नरम आवाज सुनाई दी।

मुझे टोस्टेड चीज़ और टूना सैंडविच खाना बहुत पसंद है।
अंत में



drink it.

Improve your goldfish's physical fitness by getting him a bicycle.

At that moment he wasn't listening to music, he was living an experience.

The ice-cream trucks bring back bad memories for all of us.

As he entered the church he could hear the soft voice of someone whispering into a cell phone.

I love eating toasted cheese and tuna sandwiches.

At last

Similarity Score

Score: 4.2

Readability Score

Score: 5.7

Results Page



SIL Translation Quality Tool

Upload Translations

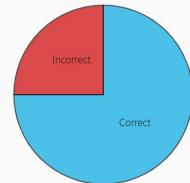
About Quality Metrics ▾

About the Project

sampleuser ▾

Translation ID - 27535

◀ To Main Results



संख्या ६।

वह हमेशा रात में अपना धूप का चश्मा पहनते थे।

चार्ल्स ने फ्रेंच फ्राइज़ को यह जानते हुए खा लिया कि वे उनका आखिरी भोजन होगा।

टॉम को पाई का एक छोटा सा टुकड़ा मिला।

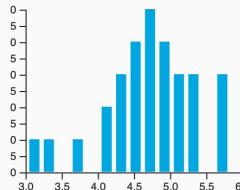
दो सीटें खाली थीं।

यादगार अवसर के लिए अपने डिजाइनर काउबॉय हैट को स्कूबा गियर के साथ पेयर करें।

हैरोल्ड को विश्वास था कि किसी को भी उसके जासूसी कबूतर पर शक नहीं होगा।

मैंने पूरे अफ्रीका का प्रमण किया है और अभी भी मुझे वह गूनहीं

Similarity Comprehensibility Readability Reset



was a try-over state.

He always wore his sunglasses at night.

Charles ate the french fries knowing they would be his last meal.

Tom got a small piece of pie.

Two seats were vacant.

Pair your designer cowboy hat with scuba gear for a memorable occasion.

Harrold felt confident that nobody would ever suspect his spy pigeon.

I've traveled all around Africa and still haven't found the

Comprehensibility

Q: When did he wear his sunglasses

A: always wore his sunglasses at night

Expected: at night

Add Question

Readability Score

Results Page

▼ TO EDIT RESULTS

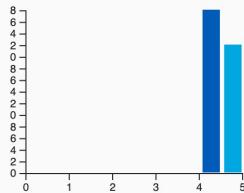
Similarity

Comprehensibility

Readability

Semantic Domains

Reset



Regarder l'équipe de softball masculin gériatrique a rappelé des souvenirs d'enfants de 3 ans jouant au t-ball.

Il y a un message pour vous si vous levez les yeux.

Des mots aléatoires devant d'autres mots aléatoires créent une phrase aléatoire.

La lumière qui brûle deux fois plus brille brûle deux fois moins longtemps.

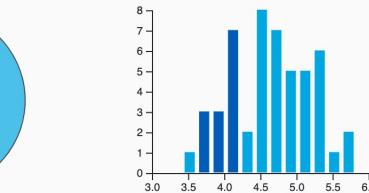
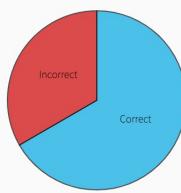
Je suis une fournaise vivante.

Il n'est pas difficile de faire le poirier si vous vous tenez simplement sur vos mains.

Sa confiance aurait été admirable s'il n'y avait pas eu sa bêtise.

La vieille pomme se délecte de son autorité.

Les meilleurs amis sont comme de vieilles tomates et des lacets.



Watching the geriatric men's softball team brought back memories of 3 yr olds playing t-ball.

There's a message for you if you look up.

Random words in front of other random words create a random sentence.

The light that burns twice as bright burns half as long,

I'm a living furnace.

It isn't difficult to do a handstand if you just stand on your hands.

His confidence would have bee admirable if it wasn't for his stupidity.

The old apple revels in its authority.

Best friends are like old tomatoes and shoelaces.

Similarity Score

Score: 4.3

Comprehensibility

Q: Where should you look to find the message?
A: if you look up
Expected: up

Add Question

Readability Score

Score: 3.67

User Translations Page



SIL Translation Quality Tool

[Upload Translations](#)[About Quality Metrics](#)[About the Project](#)

sampleuser ▾

Translations

Translation ID - 27535
भीड़ चिल्लाती है और अधिक ...
Average Similarity Score: 4.27
Correct Questions: 3 / 4
Average Readability Score: 4.67

Translation ID - 91312
Волнение сменилось страхо...
Average Similarity Score: 4.24
Correct Questions: 0 / 0
Average Readability Score: 4.62

Translation ID - 80820
ເຂົາພໍາຍາຍາມທີ່ຈະຮັກຊາສນາມທຸກ...
Average Similarity Score: 3.91
Correct Questions: 1 / 2
Average Readability Score: 4.56

Translation ID - 89589
Separation anxiety is wha...
Average Similarity Score: 4.19
Correct Questions: 0 / 0
Average Readability Score: 4.86

Translation ID - 13191
भीड़ चिल्लाती है और अधिक ...
Average Similarity Score: 4.27
Correct Questions: 3 / 4
Average Readability Score: 4.67

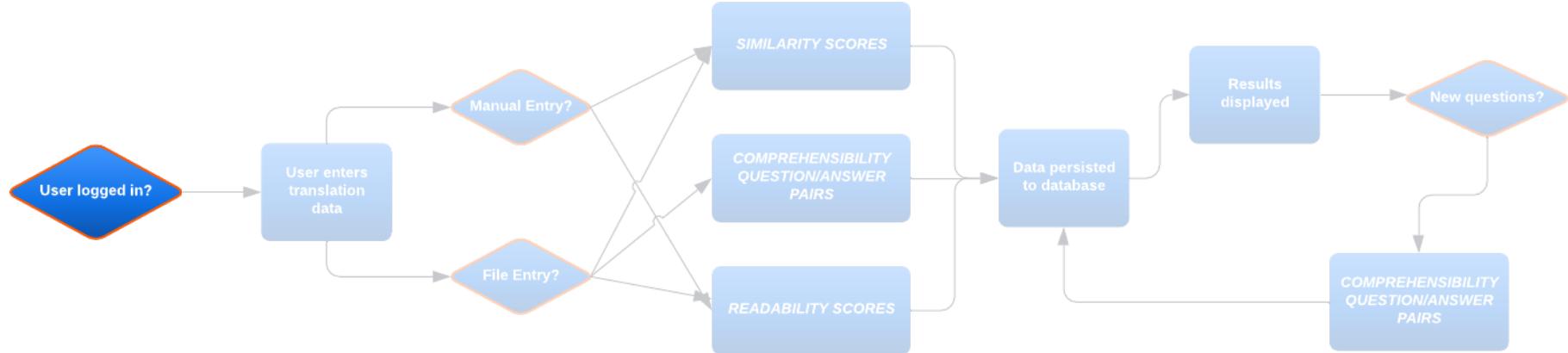
Translation ID - 90579
Ondanks wat je leraar je ...
Average Similarity Score: 4.36
Correct Questions: 1 / 3
Average Readability Score: 4.27



Backend Functionality

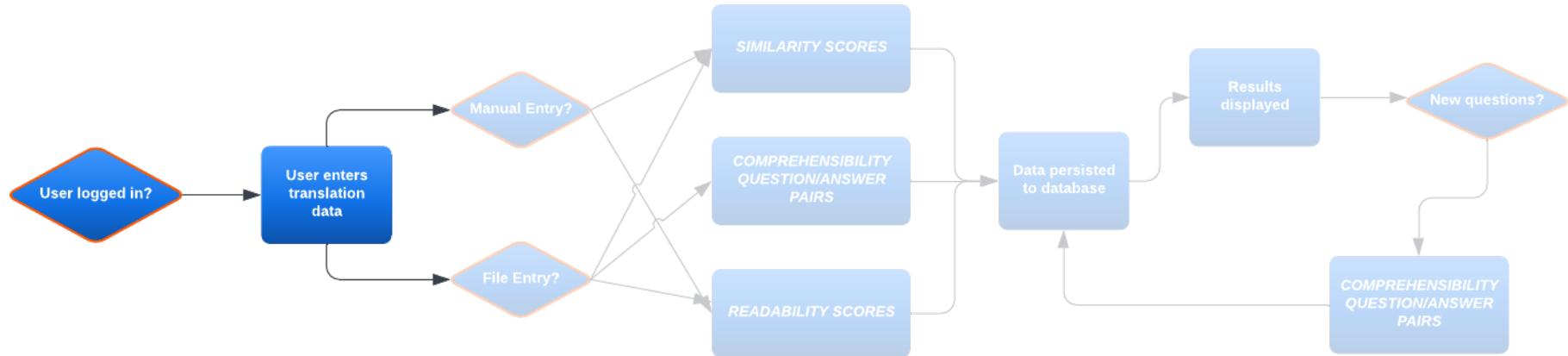
Information Flowchart

User must be logged in to access translations



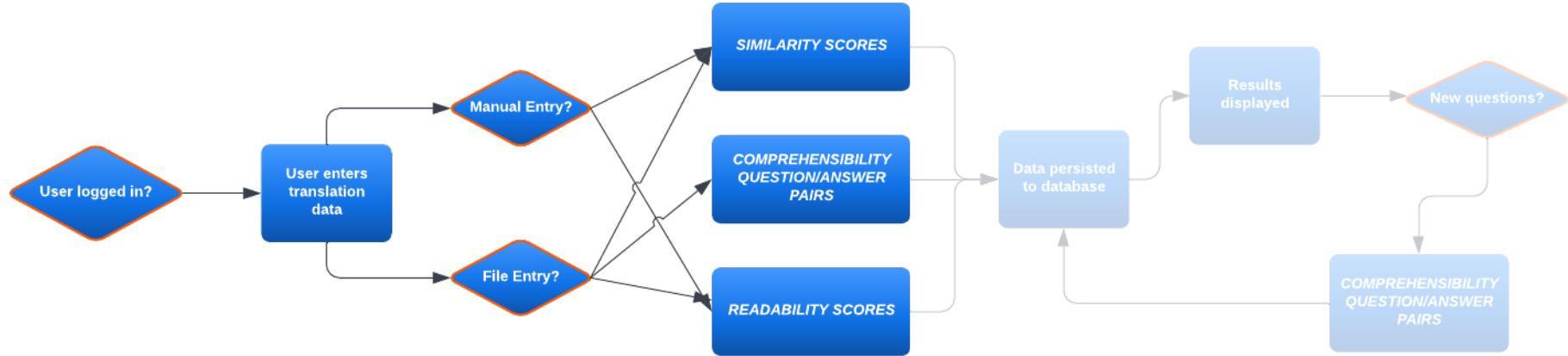
Information Flowchart

User enters source and translation text via a file upload or manual entry



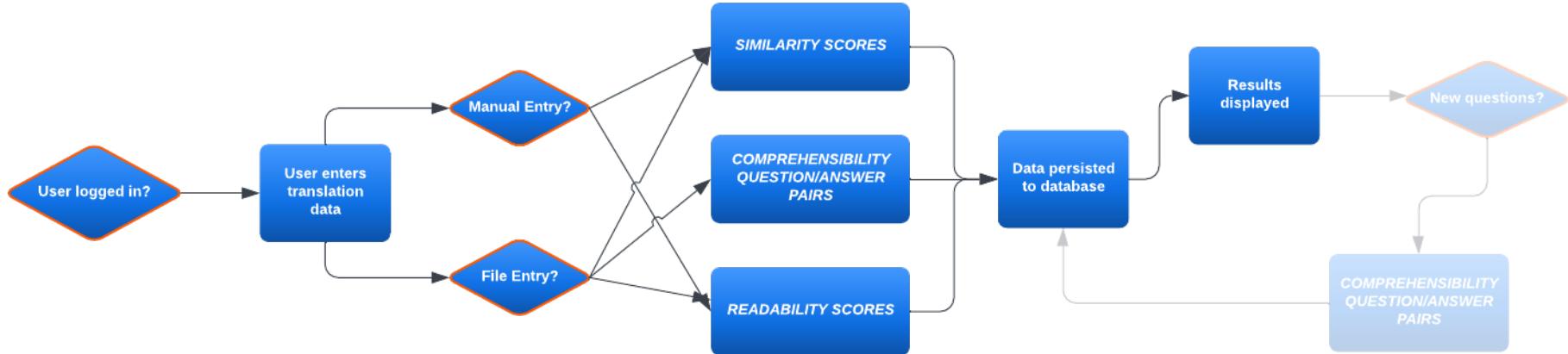
Information Flowchart

Depending on the method of upload, different metrics will be calculated



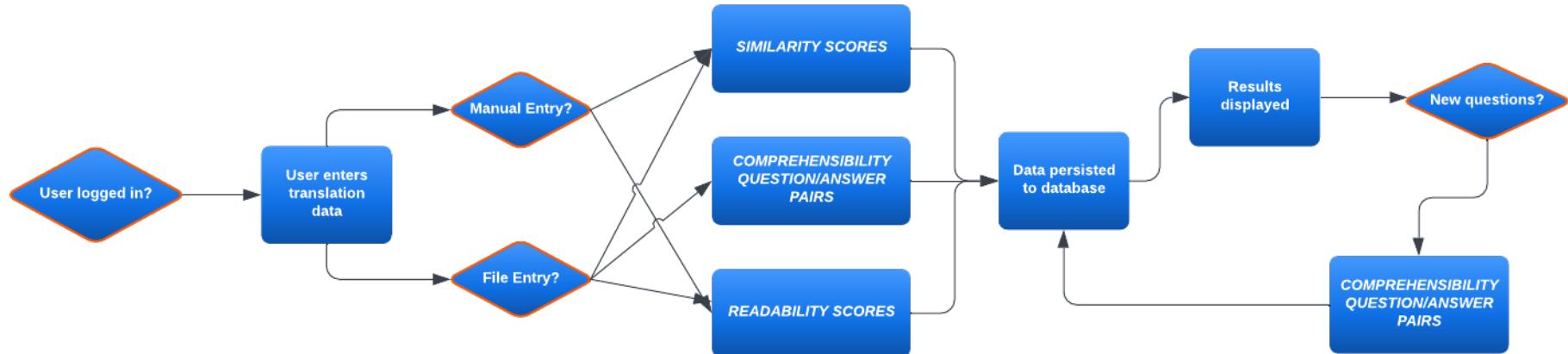
Information Flowchart

Data is persisted to the database, and results are displayed to the webpage



Information Flowchart

If a user enters additional questions, those questions will be persisted to the database and then the results are updated





Conclusion

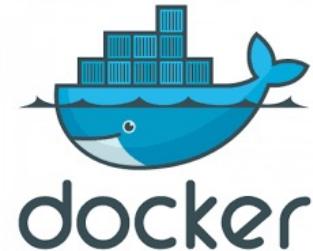
Future Scope

- Taking in voice based user inputs instead of csv using speech recognition
- Converting the translated output to speech using a text to speech tool
- Exposing an API for translations
- Allowing the user to select source and expected translation languages to provide a score based on tool made translations
- Building a robust model for readability scoring.



Takeaways and Conclusion

- Exploring the most suitable tech stack
- New Tech stack
- Full stack end to end product development
- Version control
- Code reviews
- Project management via Smartsheet
- Docker and containerization
- Poster making
- Working on a research paper





THANK YOU



Appendix



Covid-19 Translated Data

Converting JSON Data into CSV

```
"allTitles": "{\"en\":\"Avoid COVID-19 (Africa) Poster\",\"lgr\":\n\"author\": null,\n\"bookInstanceId\": \"0ad38776-391b-48cd-adca-6f77d77a0c39\",\n\"bookLineage\": \"17848ae4-d76f-4566-9275-b01a0fb16cf4,3d60fcecc-f53\n\"bookSamples\": {\n    \"0\": {\n        \"arg\": \"\",\n        \"eng\": \"Avoid COVID-19 (Africa) Poster\",\n        \"lgr\": \"Lukaotem COVID-19\"\n    },\n    \"1\": {\n        \"eng\": \"\",\n        \"lgr\": \"\"\n    },\n    \"2\": {\n        \"eng\": \"\",\n        \"lgr\": \"\"\n    },\n    \"3\": {\n        \"eng\": \"\",
```



A	B	C	D
1	lgr	eng	
2	0 Lukaotem COVID-19	Avoid COVID-19 (Africa) Poster	
3	1		
4	5	Illustrations: →Ari and Elaine Vitikain	
5	8	Adapted by Paul Frank from original, C	
6	11	How to Create Posters with this Temp	
7	12	This template is for creating posters.	
8	15	Type words into the box. If the box is	
9	19 Wasim han blong yu plande taem lon	Wash your hands regularly with soap	
10	22 No tasim ae or nos or maos blong yu.	Avoid touching eyes, nose and mouth.	
11	24 Lukaotem COVID-19	Avoid Covid-19	
12	27 Livim tufala mita melwan yu an nara	Maintain 2 meters distance between	
13	28 Sapos yu go long ples wea plande pip	Wear a mask in public.	
14	31 Folom gud olketa toktok blong dokta :	Stay informed on the latest developm	
15	35 Sapos yu kof, kavem nos an maos bloi	Cover your mouth and nose with your	
16	39 Sapos yu siki, stap long haad nomoa.	I stay home if you feel unwell. If you ha	
17	42 Lukaotem COVID-19 Lukaotem COVI	Avoid COVID-19 (Africa) Poster Avoi	
18			
19			
20			
21			
22			

Converting JSON Data into CSV

- Over 200 usable CSV files to be used for testing
- Each CSV file has at least two languages with text corresponding to the other languages
- Ultimately these files needed to be converted into a usable UTF-8 or UTF-16 encoded CSV organized into the correct format.



Original UI Design High Level
Mock

SIL Translation Quality Check

Upload a CSV

upload

Original text goes here

Translated text goes here

Original lang ▼

Translated lang ▼

submit

SIL Translation Quality Check

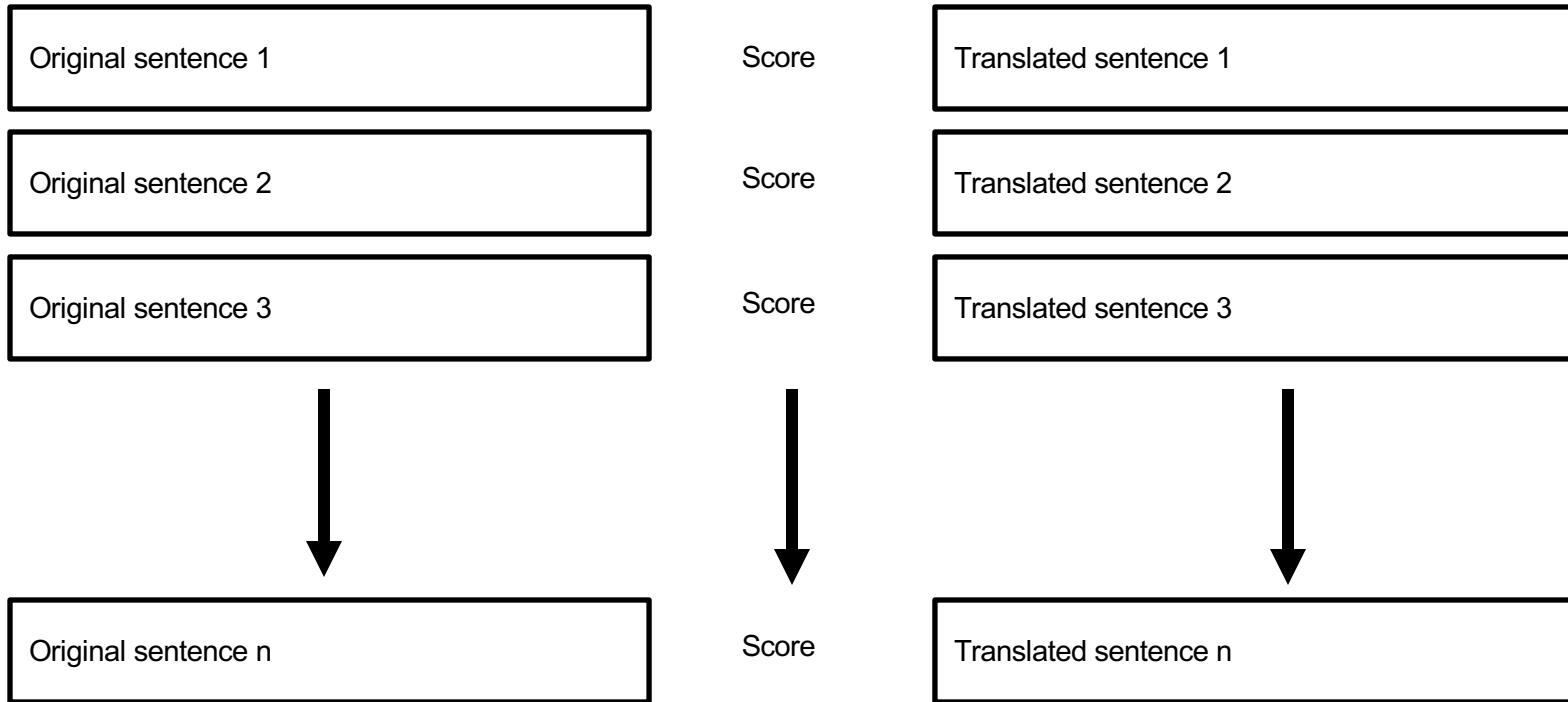
Similarity

Comprehensibility

Readability

Semantic Domains

Similarity



Overall Score

Comprehensibility

Ask a question?

lang ▼

Q: Original language question

lang ▼

A: Translated language question

lang ▼

Q: Original language question

lang ▼

A: Translated language question

lang ▼

Q: Original language question

lang ▼

A: Translated language question

Correct

Slightly Incorrect

Incorrect

Readability

Translated text goes here...

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

High Readability

Medium Readability

Low Readability