Language collider

Introduction

Language is the oldest, most advanced, and flexible distributed technology of human communication. You cannot teach a language, you can only learn it. Learning a language is like a journey. Each word, like a lighthouse, illuminates the space of the noosphere (a multidimensional space of ordered knowledge with meaning) and the mind, like a ship, navigates through it better and better. People play with words throughout their lives, in the sense that they choose which words to use, which to forget, which to modify, and how to use them. This is my favorite game

Terminology

Language Collider - a project aimed at creating a community of people studying different languages and willing to exchange accumulated knowledge with other members

Babelfish - a bot that provides an interface to the Language Collider services

User - a project member

Term - a word/phrase/text

LanguageService - f(x), f - ChatGPT prompt, x - term
List of services: Definition, Etymology, Usage Examples, Quotes, Synonyms, Antonyms, Homonyms, What, Conjugate, Proofread, Translation

LanguageService - f(x), where f is the ChatGPT prompt and x is the term. List of services: Definition, Etymology, Usage Examples, Quotes, Synonyms, Antonyms, Homonyms, What, Conjugate, Proofread, Translation

LanguageServiceResult - the result of executing LanguageService

LanguageServiceResultModification - a modification of LanguageServiceResult. It includes: continue, regenerate, translate, add_context, fix_error, suggest_correction, discard_correction, apply_correction

TermConversation - a term and all its LanguageServiceResults

Tag - Each term can have 1+ tags. By default, all user terms have the tag "default". By default, the tags for all term services are "default" and the service name. Additional tags should indicate a topic that allows the grouping of other terms and term services by meaning. The limit for the number of tags is 100(?).
Possible tags:
- Users (filtering by unique user ID and the default tag).
- Locale (the locale field in all generated/corrected content). Is the input language deterministic?
- Service type (filtering by service type - definitions, etymology, quotes, etc.).
- Topics (drive license, traveling, small talk, ortho, rude talk, etc.).

Dictionary - sets of Terms, LanguageServiceResults, LanguageServiceResultModifications, TermConversation, or their combinations corresponding to the tag vector.

Term Feed - Content feed from terms and language services corresponding to the tag vector.

Reminder - Notification from the project via web/teams/telegram/discord, containing text that allows remembering a term or term service. Reminders should not be sent during non-working hours. The user sets the limit for reminders per day. There should be a minimum of 10 minutes

between reminders. A reminder includes the ability to cancel the reminder ("learned"). A reminder includes a button to pause reminders for 1 hour/day/week/month.

Use Cases

User Registration

Omniauth registration through Microsoft, Telegram, Discord, Google.

Language services

As a user I want to be able to request different services for terms.

For example, I should be able to request definition for term "stress" on English language. As result of this request I should receive generated response with something looking like definition of stress. This language service has "Definition" service type.

Here is full list of possible services:
- Definition - meaning of term
- Translation - translation of term/text/another language service result
- Etymology - version of historical information about term
- Usage Examples - few examples of possible context
- Quotes - some very popular quotes. Can be very wrong, because ChatGPT is Language model, not reference indexer. But for learning purposes it is valid.
- Synonyms
- Antonyms
- Homonyms
- What - Return specified word by provided description. It is common problem if you forget word
- Conjugate - verb conjugation
- Proofread - fix grammar mistakes in term/text

Language service result modifications

There are several additional operations, which can be performed on language service results.
- Continue - sometimes output is too huge to be placed in single response body.
- Regenerate - OpenAI output is not straightforward. That's feature included in default chat.openai.com
- Translate - As a language learner I should be able to translate service to different languages.
- Add context/Suggest correction - As a language learner I should be able to add context or suggest correction to improve language service result.
- New words appears sometime. In such case OpenAI will provide response that term is not defined yet. It is also possible, that user is not agree and don't want to waste time and also competent enough to fix language service result by himself. So: As a user, I want be able to manually fix language service result to provide correct version, add some information or completely rewrite result.

Tags

As a user I want to add/remove tags to TermConversation. LanguageServiceResults should contain locale tag. As a user I should be able to filter term_conversation by tags. Provided filtered collection can be called Dictionary.

Terms feed

As a user I should be able to navigate through history of my and other users term_conversations.
As a user I should be able to subscribe to tags.
As a user I should be able to subscribe to other users.
As a user I should be able to rate term conversations.
As a user I should be able to provide suggestion for correction for other users term conversations.
As a user I should be able to provide comments to term conversations.
As a user I should be able to response to comments in term conversations.

Reminders

As a user I want to receive reminders with specified language service result on specified language several times, so I will remember it. As a user I should be able to specify hours in my timezone for reminders to avoid my sleeping time. As a user I should not receive reminders more than once in 10 minutes. As a user I should be able to cancel specified reminders if I learned them. As a user I should be able to pause reminders if im busy(possible interval for pause: is same as reminders schedule. Here is schedule for reminders:
- 1 hour
- 5 hours
- 1 day
- 3 days
- 5 days
- 7 days
- 14 days
- 25 days
- 8 weeks
- 16 weeks

Rankings

- Total term count
- Last 24 hours term count
- Total tag count
- Last 24 hours tag count
- Language Service Result 0-5
- Language Service Result Modification likes - dislikes
- TermConversation likes - dislikes
- Dictionary likes - dislikes

Games

- Dungeons and dragons POC
- Select correct word by definition POC
- Select correct translation POC
- Select synonyme
- Select antonym

Problems solved by product

Language learning
Research of interesting words
Self-learning community mechanism
Collaborative dictionaries as source of truth for knowledge base purposes

Roadmap

Code
- telegram MVP on rails(babel_fish_bot)
- Design and mockups
- Rails swagger api
- Omniauth integration
- Babelfish revamp and integration to other messengers
- Chrome plugin/Firefox add-on
- Frontend
- IOS/Android Application
- DAO

Presentations
- { 1v1 }
- CWS
- Aligntech
- Public

Team

Thank you for reading!
If you want to collaborate on this project and be published in that section, please tell me.

Developers:
Frontend:
- Ekaterina Grishina
Backend:
- Andrey Titov
- Kirill Dorovskii

Language teachers:
English:
- Samvel Galstyan

We need:
- Feedback
- Ideas
- Design
- Developers
- QA
- Language Teachers
- Users