DAY – 47, 48

6, 7 October 2025

1. **Dataset** → English base hovega (jaive tusi already likheya).
2. User **jo vi language likhe** (Punjabi, Hindi, English, short spelling), oh → English vich translate hove.
3. Dataset vich matching hove (with short spelling fix).
4. Answer → user di original language vich translate karke wapas ditta jave.

Simplified Correct Code

```
import re
import difflib
from deep_translator import GoogleTranslator
from langdetect import detect

# ===== LOAD DATASET =====
faq_data = {}
with open("dataset.txt", "r", encoding="utf-8") as f:
    for line in f:
        if "=" in line:
            q, a = line.split("=", 1)
            faq_data[q.strip().lower()] = a.strip()

# ===== SHORT FORMS + MISSPELLINGS =====
SHORT_FORM_DICT = {
    "u": "you",
    "ur": "your",
    "r": "are",
    "pls": "please",
    "plz": "please",
```

```
}
MISSPELLINGS = {
    "registartion": "registration",
    "applcation": "application",
    "acount": "account",
}

def normalize_english(text):
    """ Fix short forms & misspellings only for English """
    text = text.lower()
    for k, v in SHORT_FORM_DICT.items():
        text = re.sub(r"\b" + re.escape(k) + r"\b", v, text)
    for k, v in MISSPELLINGS.items():
        text = text.replace(k, v)
    return text.strip()

# ===== MAIN RESPONSE =====
def get_response(user_input):
    try:
        # Detect user language
        lang_code = detect(user_input)
    except:
        lang_code = "en"

    # STEP 1: Translate user question → English (for matching)
    if lang_code != "en":
        input_en = GoogleTranslator(source="auto", target="en").translate(user_input)
    else:
        input_en = user_input

    # STEP 2: Normalize English
    input_en = normalize_english(input_en)

    # STEP 3: Search in dataset
```

```
matches = difflib.get_close_matches(input_en, faq_data.keys(), n=1, cutoff=0.6)
if matches:
    answer = faq_data[matches[0]]
else:
    answer = "Sorry, I don’t have information about that."

# STEP 4: Translate answer back to user’s language
if lang_code != "en":
    answer = GoogleTranslator(source="en", target=lang_code).translate(answer)

return answer
```

Example Run

User → "plz registartion"
→ Normalize → "please registration"
→ Match dataset → "tell me about new registration"
→ Answer → "To use the application, first click the sign-up button..."
User → "ਕਿਰਪਾ ਕਰਕੇ ਰਜਿਸਟ੍ਰੇਸ਼ਨ"
→ Translate → "please registration"
→ Match dataset → same as above
→ Answer translated back → Punjabi
User → "रजिस्ट्रेशन बताओ"
→ Translate → "tell me registration"
→ Match dataset
→ Answer translated back → Hindi

```
from deep_translator import GoogleTranslator

from langdetect import detect

import difflib


# ===== Dataset (English only) =====

faq_data = {

    "what is your name": "I am your assistant chatbot.",

    "hello": "Hello! How can I help you today?",

    "bye": "Goodbye! Have a great day.",

    "tell me about new registration": "To use the application, first click the sign-up button. Then, fill in your personal details and submit the form.",

    "how can i register as a farmer": "Visit the registration page on our website, click 'Register', and fill in your personal and farm details.",

    "is aadhaar card mandatory": "Yes, Aadhaar card is required for most government farming schemes.",

}


def get_response(user_input):

    try:

        # Detect user language

        user_lang = detect(user_input)

    except:
```

```
    user_lang = "en"


# Step 1: Convert user question → English
if user_lang != "en":
    query_en = GoogleTranslator(source="auto", target="en").translate(user_input)
else:
    query_en = user_input


# Step 2: Find closest match in dataset
match = difflib.get_close_matches(query_en.lower(), faq_data.keys(), n=1, cutoff=0.6)
if match:
    answer = faq_data[match[0]]
else:
    answer = "Sorry, I don't have information about that."


# Step 3: Convert answer back to user's language
if user_lang != "en":
    answer = GoogleTranslator(source="en", target=user_lang).translate(answer)


return answer
```

```
# ==== Example Runs ====

print(get_response("plz registartion"))       # English spelling mistake

print(get_response("ਕਿਰਪਾ ਕਰਕੇ ਰਜਿਸਟ੍ਰੇਸ਼ਨ"))  # Punjabi

print(get_response("रजिस्ट्रेशन बताओ"))       # Hindi
```