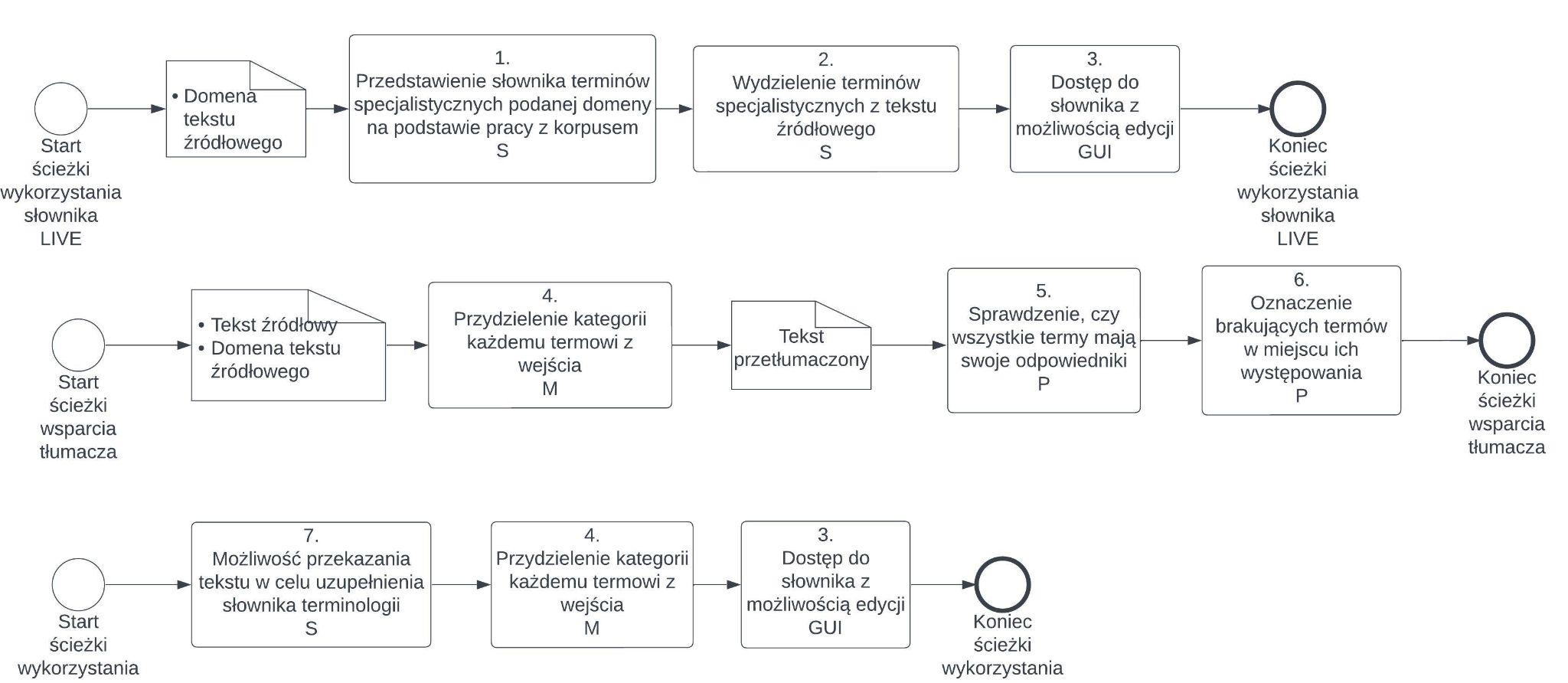
**TermHarbor - Architektura i harmonogram prac**

1. **Architektura systemu**



1. Przedstawienie słownika terminów specjalistycznych podanej domeny na podstawie pracy z korpusem
   1. Osoba odpowiedzialna: Sebastian
   2. Input: Domena
   3. Output: słownik zawierający pary ‘słowo oryginalne’:’słownik’, gdzie wartościami w słowniku są: słowo, lista tłumaczeń w języku docelowym
   4. Algorytm/mechanizm działania:   
      Ekstrakcja terminów: Tokenizacja, oznaczanie części mowy, identyfikacja sekwencji słów lub kolokacji, filtracja terminów ogólnych, budowanie reguł z wykorzystaniem biblioteki spacy.
2. Wydzielenie terminów specjalistycznych z tekstu źródłowego
   1. Osoba odpowiedzialna: Sebastian
   2. Input: Tekst źródłowy
   3. Output: Słownik zawierający terminy specjalistyczne wyekstrahowane z tekstu źródłowego
3. Dostęp do słownika z możliwością edycji (opcjonalne)
   1. Osoba odpowiedzialna: Maurycy
   2. Input:
   3. Output:
   4. Mechanizm działania:

Użytkownikowi wyświetla się lista termów danej domeny. Każdy term posiada indywidualne menu boczne, po którego naciśnięciu można dotrzeć do opcji usunięcia termu lub jego edycji. Edycja polega na wyświetleniu wszystkich informacji o termie z możliwością ich zmiany.

1. Przydzielenie kategorii każdemu termowi z wejścia.
   1. Osoba odpowiedzialna: Maurycy
   2. Input: Słownik zawierający terminy specjalistyczne wyekstrahowane z tekstu źródłowego
   3. Output: Słownik z przypisanymi tagami kategorii do danego słowa
2. Sprawdzenie, czy wszystkie termy mają swoje odpowiedniki
   1. Osoba odpowiedzialna: Paweł
   2. Input: tekst źródłowy, tekst przetłumaczony (lub jego fragment), słownik terminologii języka źródłowego i docelowego, lista termów tekstu źródłowego, lista termów tekstu docelowego
   3. Output: Oznaczenie poprawności, jeżeli wszystkie termy zostały zawarte w tekście lub lista termów nieoznaczonych wraz z numerem zdania/sekcji/linii, w której się znajdują
3. Oznaczenie brakujących termów w miejscu ich występowania
   1. Osoba odpowiedzialna: Paweł
   2. Input: lista termów nieoznaczonych
   3. Output: Indeks miejsc, w których mogłyby występować brakujące słowa
   4. Algorytm: Analiza słów okalających brakujący wyraz w tekście źródłowym i zastosowanie tej metody w tekście docelowym
4. Możliwość przekazania tekstu w celu uzupełnienia słownika terminologii
   1. Osoba odpowiedzialna: Sebastian
   2. Input: tekst
   3. Output: terminologia zapisana w słowniku

**2. Harmonogram prac**

|  |  |  |  |
| --- | --- | --- | --- |
| Termin | Punkt kontrolny | Produkt | Priorytet |
| 31.10.2023 | Wstępna specyfikacja wymagań | Dokument wymagań projektowych | 1 |
| 12.11.2023 | Przygotowanie prezentacji koncepcji projektu | Prezentacja multimedialna i schemat wystąpienia | 2 |
| 14.11.2023 | Prezentacja koncepcji projektu | Wystąpienie | 1 |
| 31.11.2023 | Architektura i harmonogram prac | Rozdział dokumentu założeń projektowych | 1 |
| 30.12.2023 | KPI (kryteria sukcesu) i analiza ryzyka | Rozdział dokumentu założeń projektowych | 1 |
| 19.01.2024 | Przygotowanie posteru | Poster i schemat wystąpienia | 2 |
| 23.01.2024 | Sesja posterowa | Wystąpienie | 1 |
| 31.01.2024 | Badania oraz prace nad poszczególnymi modułami projektu | Koncepcja rozwiązań i algorytmów | 1 |
| 24.02.2024 | Lokalne repozytorium projektu zgodne z wytycznymi | Repozytorium lokalne | 1 |
| 30.03.2024 | Pierwszy commit na repozytorium produkcyjnym | Repozytorium GitHub | 1 |
| 05.05.2024 | Pierwszy kamień milowy - stworzenie kompletnych, osobnych modułów projektu | Repozytorium GitHub |  |
| 19.05.2024 | Drugi kamień milowy - połączenie istniejących modułów projektów | Repozytorium GitHub |  |
| 03.06.2024 | Testy integracyjne | Repozytorium GitHub |  |
| 10.06.2024 | Ostatni commit na repozytorium produkcyjnym | Repozytorium GitHub | 1 |
| 10.06.2024 | Raport z wykonania projektu | Raport PDF | 1 |