

White Paper: Etykietowanie tekstu wspomnieniowego na potrzeby budowy grafu wiedzy

Cel

Celem projektu jest przekształcenie tekstu wspomnieniowego o charakterze autobiograficzno-literackim w uporządkowany graf wiedzy, zawierający: - obiekty (encje) typu fizycznego i abstrakcyjnego, - relacje między encjami, - fenomeny (myśli, wrażenia, retrospekcje), - hierarchię ważności bytów (priorytet egzystencjalny), - sygnaturę brakujących bytów (BRAK).

Model etykietowania

Tekst jest dzielony na segmenty logiczne (np. miejsce, życie codzienne, praca), a następnie każdemu fragmentowi przypisuje się: - listę encji wraz z typem i podstawowymi cechami, - występujące relacje (np. `uses`, `part_of`, `located_in`), - zidentyfikowane fenomeny mentalne i ich właściciela, - sygnalizację brakujących bytów (np. `piekarnik` jako brakująca encja typu `PRZEDMIOT`).

Przykładowe typy encji

- **BUDYNEK:** kamienica
- **MIEJSCE:** dach, mieszkanie, pokoje, korytarz
- **PRZEDMIOT:** lodówka, czajnik, butla
- **ZASÓB:** prąd, gaz
- **GRUPA LUDZI / OSOBA:** mieszkańcy, sąsiedzi, policja, Krzysiek
- **INSTYTUCJA:** Tesco, Kamienica Cudów
- **FENOMEN / MYŚL:** strach, wdzięczność, autodefinicja, degradacja moralna
- **WYDARZENIE CYKLICZNE:** Sylwester
- **BRAK:** piekarnik, telewizor, internet, kablówka

Fenomeny

Myśli traktowane są jako obiekty typu `FENOMEN`, posiadające: - właściciela (jeśli możliwe), - treść i typ (np. `autodefinicja`, `lęk`, `refleksja`), - `existential_priority` — liczba określająca wagę fenomenologiczną, - możliwe relacje do encji (np. `dotyczy`, `powiązana z`).

Zasady ekstrakcji

- Nie uwzględniamy encji, które są wyłącznie kontrfaktyczne lub wyraźnie nieobecne — trafiają one do sekcji `BRAK`.
- Encje z wiedzy ogólnej mogą być implikowane (np. `łazienka` z `deski klozetowej`).
- Myśl jest bytem o wysokim priorytecie i może być powiązana z każdą encją (nawet z kamieniem).

Przykład

Fragment: „Nie mam piekarnika. Tort zjem w cukierni.” - piekarnik → typ BRAK, reason: brak sprzętu w mieszkaniu - cukiernia → typ MIEJSCE, relacja: provides → tort - tort → typ PRZEDMIOT, kontekstowy, ale realny

Myśl (implikowana): *strategia radzenia sobie z brakiem* → fenomen z afektem: adaptacja

Rezultat

W wyniku działania systemu etykietującego uzyskujemy: - strukturę encji i relacji możliwą do zapisania jako JSON/graf RDF, - punkt wejścia do RAG lub systemu generatywnego, - możliwość rekontekstualizacji wspomnień za pomocą wiedzy uporządkowanej.

Ten dokument opisuje podejście do etykietowania tekstu autobiograficznego z uwzględnieniem zarówno bytu materialnego, jak i subiektywnego (mentalnego), zachowując jego oryginalny styl i treść jako kontekstualne źródło wiedzy.