

<https://jacek.kwasniewski.org.pl>

Jacek Kwaśniewski



Lęk przed śmiercią jako siła kulturotwórcza

Jak przeszukiwać literaturę naukową z pomocą AI

Część 4

Budowa rdzeni semantycznych dla modelu embeddingowego

Spis treści

Literatura o embeddingach i oryginalność metody rdzeni semantycznych	2
Problem	3
Czym jest rdzeń semantyczny – definicja, cel, zakres zastosowania.....	4
Kotwica tematyczna (agens procesu).....	5
Agens – definicja operacyjna	5
Typy agensa: wykonawczy i przyczynowy	5
Dlaczego kotwica stoi na początku rdzenia	5
Kotwica to nie temat zdania	6
Czy kotwica może nie stać na początku?	6
Struktura logiczna rdzenia.....	6
Element procesualny	7
Model kanoniczny (czasownikowy).....	7
Model nominalny (nominalizacja procesu).....	8
Przedmiot procesu musi być jawnie wskazany	8
Kotwica musi pozostać agensem.....	9
Rozróżnienie: przedmiot procesu vs pole działania	9
Kiedy nominalizacja ma sens	10
Długość rdzenia.....	10
Rozbijanie rdzeni zbyt szerokich	10
Jeden proces na jeden rdzeń	11
Jedno pole semantyczne	11
Zakaz pytań, negacji i modalności	11
Zakaz strony biernej	12
Konsystencja leksykalna	12
Redukcja przymiotników i przysłówków.....	12
Unikanie metafor i języka figuratywnego	13
Oddzielanie perspektyw naukowych.....	13
Kompatybilność embeddingów i zasada jednego okna czasowego	14
Checklista walidacyjna rdzeni semantycznych	15
Zasady budowy rdzeni semantycznych – skrót.....	16

Literatura o embeddingach i oryginalność metody rdzeni semantycznych

Metoda budowy rdzeni semantycznych przedstawiona w tym dokumencie nie posiada bezpośredniego odpowiednika w istniejącej literaturze o embeddingach tekstu. Nie wynika to z braku badań nad reprezentacjami semantycznymi, lecz z faktu, że współczesne prace koncentrują się na poziomie modeli, wektorów i przekształceń przestrzeni, a nie na formalnej konstrukcji jednostek znaczeniowych, które użytkownik wprowadza do systemu.

Najbliższe obszary literatury, z których ta metoda czerpie inspiracje, to:

1. Embedded representations (prace o embeddingach tekstowych)

Badania nad reprezentacjami wektorowymi (BERT, GPT, fastText, GloVe i nowsze generacje modeli OpenAI) opisują:

- jak tekst staje się wektorem,
- jak modele rozpoznają relacje semantyczne,
- jak działa podobieństwo kosinusowe.

Jednak nie opisują, jak precyzyjnie konstruować zdania wejściowe tak, by wektory były maksymalnie jednoznaczne. To pozostaje domeną praktyki.

2. Interpretowalne embeddingi i semantyczne przestrzenie

Literatura o „interpretable embeddings” i „semantic feature spaces” (głównie arXiv) bada, jak reprezentacje wektorowe ujawniają ukryte cechy, w tym:

- kierunki semantyczne,
- operatory kulturowe,
- relacje symboliczne.

To najbliższy teoretycznie obszar, bo dotyczy zagadnienia przejrzystości znaczenia, jednak nie daje żadnych operacyjnych zaleceń dotyczących budowy zdań, tylko opisuje zachowanie przestrzeni.

3. Praktyka systemów RAG i wyszukiwania semantycznego

W literaturze dewelopersko-inżynierskiej pojawiają się wskazówki o:

- optymalizacji zapytań,
- redukcji szumu leksykalnego,
- używaniu stałego słownika,
- unikaniu negacji i pytań w zapytaniach wektorowych.

To są sygnały intuicyjnie zgodne z moją metodą, ale nadal - nie istnieje żadna formalna „szkoła rdzeni”.

4. Oryginalność podejścia rdzeni semantycznych

To, co przedstawiam w tym dokumencie, jest autorską, spójną metodą reprezentacji kulturowej w przestrzeni embeddingów. Jej kluczowe elementy nie występują w literaturze naukowej ani technicznej jako zestaw reguł:

- definicja rdzenia jako minimalnej jednostki procesu kulturowego,
- twardy zakres długości 7–14 słów,
- eksplicytne rozróżnienie kotwicy, procesu, przedmiotu i pola,
- zakaz negacji, pytań, modalności i metafor,
- zakaz strony biernej jako warunek jednoznaczności kierunku działania,
- formalny model dwóch typów procesualności: kanonicznej (czasownikowej) i nominalnej,
- logika operatora sprawczości, w tym model autoreferencyjny,
- spójny słownik rdzeni jako narzędzie stabilizacji przestrzeni wektorowej.

To podejście łączy trzy dziedziny, które normalnie nie występują razem:

1. Antropologię i teorię kultury – bo strukturalizuję procesy kulturowe;
2. Inżynierię reprezentacji wektorowej – bo projektuję zdania pod embedding;
3. Praktykę wyszukiwania semantycznego (RAG, FAISS) – bo rdzenie są operacyjne, nie teoretyczne.

W rezultacie otrzymuję pierwszy w praktyce konsekwentny „kodeks rdzeni semantycznych”, którego nie znajdziemy w żadnej publikacji naukowej.

To nowa metodologia, kompatybilna z mechaniką embeddingów, ale wykraczająca poza dotychczasowe standardy.

Problem

Analizowany temat to kulturotwórcza rola lęku przed śmiercią (LPS). Literatura tego przedmiotu jest obszerna, szacuję ją na kilkanaście do kilkudziesięciu tysięcy stron.

Jakiś czas temu zadałem sobie pytanie: co konkretnie znaczy, że lęk przed śmiercią jest silnie kulturotwórczy. Odpowiedzi mieszczą się pomiędzy dwoma biegunami. Pierwszy operuje na skali globalnej. Stwierdza, że cała kultura jest odpowiedzią na LPS i podaje wyrywkowe przykłady (np. Becker, TMT). Drugi koncentruje się na konkretach. Pokazuje, że lęk przed śmiercią (LPS) ma wpływ na tysiące artefaktów kulturowych, od praktyk żałobnych przez zmiany języka po normy moralne i kierunki rozwoju nauki.

Pytanie brzmi, co cała literatura przedmiotu o tym wszystkim mówi. Stwórzmy problemowy przegląd naszej wiedzy. Zbudujmy listę szczegółowych pytań/zagadnień na temat kulturotwórczej roli lęku przed śmiercią (LPS) i skierujmy ją do literatury.

Niech do każdego pytania/zagadnienia, na przykład ‘Lęk przed śmiercią współkształtuje zmiany estetyki nagrobków i treści epitafiów w XX i XXI wieku’ czy ‘Lęk przed śmiercią wpływa na sposób komunikowania o śmierci w mediach i edukacji’ zostaną przypisane fragmenty literatury, które dokładnie o tym mówią, dokładnie to zagadnienie analizują.

Spotkanie pytań z literaturą nastąpi w tak zwanym modelu embeddingowym. Zamieniamy on literaturę i pytania na postać cyfrową. Model koduje standardowe (co do rozmiarów) fragmenty tekstów i nasze pytania na pojedyncze wektory w 1536-wymiarowej przestrzeni tensorowej. Jest to jednocześnie przestrzeń semantyczna, w której odległości między

wektorami odzwierciedlają relacje znaczeniowe - podobieństwo pojęć, kontekstów i większych struktur tekstowych.

Fragmenty tekstów i pytania zamienione na wektory nazywają się embeddingami. Wszystkie embeddingi przejdą do kolejnego programu, modułu wyszukiwawczego FAISS. Program FAISS pełni funkcję silnika wyszukiwania wektorowego: wczytuje z modelu embeddingowego embeddingi tekstów oraz embeddingi pytań i przeszukuje przestrzeń 1536-wymiarowych wektorów w celu znalezienia najbardziej semantycznie zbliżonych fragmentów tekstów i pytań. Co z tym zrobimy dalej, zależy od nas.

Pytania, również te w języku naturalnym, kierowane do literatury umieszczonej w modelu embeddingowym muszą mieć konstrukcję, która daje się zamienić na „dobry” embedding. Co to znaczy dobry, o tym traktuje ten tekst.

Czym jest rdzeń semantyczny – definicja, cel, zakres zastosowania

Rdzeń semantyczny to zwięzłe, jednoznaczne twierdzenie o procesie, czyli o mechanizmie: kto/co na co działa, w jaki sposób i w jakim polu kultury. Nie jest to ani cytat z tekstu, ani streszczenie artykułu, ani pytanie badawcze. Rdzeń jest zaprojektowany po to, aby można go było przekształcić w swego rodzaju zapytanie pod adresem literatury na określony temat, która została przetworzona cyfrowo i wprowadzona do specjalnego programu, zwanego modelem embeddingowym (ME).

Model embeddingowy (ME) rozumie treść literatury w tym sensie, że można mu zadać pytanie problemowe z nią związane i uzyskać fragment(y) tekstów, które dokładnie o tym problemie traktują. Mówiąc inaczej, rdzeń jest ciągiem wyrazów, które można przekształcić w stabilny embedding – wektor semantyczny, który dobrze reprezentuje konkretną dynamikę kulturową.

Zasady opisane w tym dokumencie są ogólne – można je stosować w bardzo różnych projektach, niezależnie od tego, czy kotwicą jest „rewolucja cyfrowa”, „modernizacja religijności”, „rodzina nuklearna”, czy cokolwiek innego.

Przykłady, które pojawiają się dalej, pochodzą z konkretnego projektu badawczego:

„Kulturotwórcza rola lęku przed śmiercią (KR-LPS)”

i dotyczą tego, jak lęk przed śmiercią (LPS) wpływa na rytuały, symbole, normy, narracje, praktyki żałobne czy struktury społeczne. LPS jest tu więc ilustracją zasad, a nie warunkiem ich stosowania.

Rdzeń ma pełnić równocześnie kilka funkcji:

- być transparentny semantycznie – w jednej lekturze (1–2 sekundy) wiadomo, o jaki proces chodzi,
- być stabilny embeddingowo – ten sam rdzeń powinien zawsze przyciągać teksty o podobnym typie zjawiska,
- być operacyjny – dać się łatwo zweryfikować: czy spełnia kryteria długości, procesualności, jednego pola, kotwicy itp.

„Operacyjny” znaczy taki, który można systematycznie audytować. Rdzeń da się przepuścić przez checklistę (patrz koniec dokumentu) i jednoznacznie odpowiedzieć, czy jest poprawny,

czy wymaga poprawki. W tym sensie rdzeń jest jednostką pracy analitycznej – najmniejszym klockiem, na którym stoi system wyszukiwania i analizy semantycznej

Kotwica tematyczna (agens procesu)

Kotwica jest pierwszym elementem rdzenia i pełni funkcję agensu, czyli czynnika sprawczego procesu, w przypadku naszego projektu, procesu kulturowego. W rdzeniach semantycznych kotwica nie jest „tematem zdania”, lecz źródłem dynamiki: to ona generuje potrzebę przekształcenia, oswojenia, reinterpretacji czy reorganizacji.

W projekcie KR-LPS (kulturotwórcza rola lęku przed śmiercią) kotwicą jest „Lęk przed śmiercią”, rozumiany nie jako indywidualne przeżycie psychiczne, lecz jako czynnik kulturotwórczy – coś, co wymusza na kulturze określone odpowiedzi (rytualne, symboliczne, normatywne, narracyjne).

Agens – definicja operacyjna

Agens to czynnik sprawczy procesu. Oznacza to, że:

- inicjuje proces (bez niego proces by nie zaszedł albo miałby inną postać),
- wyznacza kierunek działania (od kotwicy do przedmiotu procesu),
- organizuje całą sytuację kulturową opisywaną rdzeniem.

W rdzeniach semantycznych chodzi przede wszystkim o agens przyczynowy, a nie o agensa wykonawczego. To nie jest ktoś, kto „coś robi”, tylko czynnik, który powoduje, że coś musi być zrobione.

W tym sensie lęk przed śmiercią (LPS) jest agensem: nie „oswaja” sam w sobie, ale wywołuje potrzebę oswojenia; nie „przekształca rytuałów rękami aktorów”, ale powoduje, że rytuały są przekształcane.

Typy agensa: wykonawczy i przyczynowy

- Agens wykonawczy – klasyczny podmiot działania („artysta maluje obraz”, „kapłan odprawia rytuał”). W rdzeniach semantycznych pojawia się rzadko, bo rdzenie nie opisują pojedynczych aktów, tylko mechanizmy kulturowe.
- Agens przyczynowy – czynnik, którego obecność generuje proces (np. lęk przed śmiercią, tabu, normy moralne, powstanie nowych technologii, rozwój nowych instytucji). To jest główny typ agensa, z którym pracujemy.

W praktyce rdzenie prawie zawsze bazują na agensie przyczynowym.

Dlaczego kotwica stoi na początku rdzenia

Umieszczanie kotwicy na początku rdzenia:

- ujednolica strukturę całego korpusu,
- od razu pokazuje, co jest czynnikiem sprawczym,
- pomaga modelowi embeddingowemu „zrozumieć”, że wszystkie te procesy są projekcjami jednego czynnika,

- ułatwia klastrowanie (rdzenie z tą samą kotwicą grupują się w spójne zbiory),
- pozwala szybko sprawdzić, czy dany rdzeń w ogóle należy do danego projektu.

Dlatego standardem jest rozpoczynanie rdzenia od kotwicy – zwłaszcza w projektach takich jak kulturotwórcza rola lęku przed śmiercią (KR-LPS), gdzie jeden czynnik ma być konsekwentnie śledzony w wielu polach kultury.

Kotwica to nie temat zdania

Kotwica nie jest tym samym, co „temat zdania” czy „obszar zainteresowania”.

Zdanie typu: „Lęk przed śmiercią a symbolika żałobna” nie jest poprawnym rdzeniem, bo nie zawiera procesu ani kierunku - to tylko zestawienie dwóch pól.

Kotwica ma pełnić funkcję agensu przyczynowego, a nie tylko nazwy obszaru badań. Rdzeń musi odpowiadać na pytanie: co LPS robi z danym polem?, a nie: z czym lęk przed śmiercią współwystępuje?

Czy kotwica może nie stać na początku?

Teoretycznie można zbudować poprawny rdzeń, w którym kotwica nie stoi na początku zdania – np. gdy składnia wymaga innego szyku, albo gdy przedmiot procesu musi być wysunięty na czoło ze względu na przejrzystość.

W praktyce jednak:

- zdecydowana większość rdzeni może (i powinna) być sformułowana z kotwicą na początku,
- ta postać jest najbardziej jednoznaczna semantycznie,
- zapewnia największą stabilność embeddingową i spójność korpusu.

Dlatego w projektach takich jak KR-LPS (kulturotwórcza rola lęku przed śmiercią) przyjmujemy zasadę domyślną:

Kotwica stoi na początku rdzenia, a wyjątki traktujemy jako przypadki rzadkie, posiadające szczególnie uzasadnione.

Struktura logiczna rdzenia

Na poziomie znaczenia struktura każdego rdzenia jest taka sama, niezależnie od tego, czy używamy formy czasownikowej (model kanoniczny), czy nominalnej (model nominalny). Rdzeń zawsze da się rozpisać według schematu:

kotwica (agens) → element procesualny → przedmiot procesu → pole / kanał / kontekst

Przykład: „Lęk przed śmiercią kształtuje doświadczenie żałoby poprzez praktyki żałobne w kulturze miejskiej XX wieku.”

- Kotwica (agens) – główny czynnik sprawczy rdzenia (w projekcie KR-LPS: „Lęk przed śmiercią”).

- Element procesualny – proces, który się dokonuje (np. kształtuje, transformuje, oswaja, reinterpretuje, amortyzuje, kształtowanie, transformacja, oswajanie, reinterpretacja, amortyzacja).
- Przedmiot procesu – to, co bezpośrednio podlega działaniu (lęk, doświadczenie żałoby, normy, relacje, symbolika, rytuały).
- Pole / kanał / kontekst działania – to, w czym i poprzez co proces się realizuje (rytuały, ikonografia, instytucje, praktyki żałobne, typ kultury, epoka).

W całym tekście używamy konsekwentnie:

- „przedmiot procesu” – dla tego, co jest przekształcane,
- „pole / kanał działania” – dla tego, poprzez co proces przebiega.

Element procesualny

Każdy rdzeń musi zawierać element procesualny, ponieważ rdzeń opisuje mechanizm kulturowy, a nie temat, kategorię czy powiązanie. Procesualność to centrum znaczenia rdzenia: wskazuje, co się dzieje i jak przebiega zmiana.

Element procesualny może przyjąć dwie równorzędne formy:

- model kanoniczny (czasownikowy) – proces wyrażony jest dynamicznym czasownikiem,
- model nominalny (nominalizacja procesu) – proces wyrażony jest rzeczownikiem odczasownikowym i wymaga operatora sprawczości.

Oba modele są zgodne ze strukturą z punktu 3:

kotwica (agens) → operator sprawczości / czasownik procesualny → przedmiot procesu → pole/kontekst działania

Model kanoniczny (czasownikowy)

To forma podstawowa, najprostsza i najbardziej przejrzysta.

Procesualność jest tu wyrażona czasownikiem w stronie czynnej, np.:

kształtuje, przekształca, wytłumia, inspiruje, organizuje, kanalizuje, intensyfikuje, rozluźnia, modeluje, redefiniuje.

Struktura rdzenia wygląda wtedy tak:

kotwica → robi coś (czasownik procesualny) → przedmiot procesu → pole / kanał / kontekst

Dlaczego czasownik czynny?

Bo jednoznacznie określa kierunek działania, którego embedding nie musi się „domyślać”.

Czego nie używać?

Pseudo-czasowników statycznych:

- dotyczy,
- odnosi się do,
- pozostaje w relacji z,

- związany jest z.

Te formy nie opisują procesu, lecz relację tematyczną — i są bezużyteczne dla embeddingu.

Przykłady dobrych rdzeni czasownikowych

- „Lęk przed śmiercią w XX wieku zmienił rytuały pogrzebowe w kulturze miejskiej.”
- „Lęk przed śmiercią współkształtuje rozróżnienia dobra i zła.”
- „Lęk przed śmiercią kształtuje debatę o znaczeniu śmierci w kulturze.”

Wszystkie mają:

- jasny proces,
- określony przedmiot działania,
- jedno pole kulturowe,
- stronę czynną.

Przykłady złych rdzeni czasownikowych

- „Lęk przed śmiercią zmienia kulturę.” – pole zbyt ogólne.
- „Lęk przed śmiercią a dobro i zło.” – brak procesu.
- „Lęk przed śmiercią a debata o śmierci.” – samo zestawienie pojęć.

Model nominalny (nominalizacja procesu)

W tym modelu proces wyrażony jest nominalizacją (rzeczownikiem odczasownikowym):

oswajanie, amortyzacja, kształtowanie, reinterpretacja, kanalizacja, symbolizacja, instytucjonalizacja.

Jednak nominalizacja nigdy nie może stać samodzielnie, ponieważ nie zawiera informacji o tym, kto/co uruchamia proces. Rzeczownik odczasownikowy („oswajanie”, „amortyzacja”, „kształtowanie”, „reinterpretacja”) sam z siebie nie mówi jeszcze, co jest osvajane, amortyzowane, kształtowane czy reinterpretowane.

Dlatego model nominalny zawsze wymaga operatora sprawczości, czyli element łączący kotwicę z procesem.

Operator sprawczości występuje w dwóch rodzajach:

(1) Operator sprawczości czasownikowy (czasownik inicjujący)

inicjuje, prowadzi do, wywołuje, uruchamia

(2) Operator sprawczości kauzalny (faza sprawcza)

przez lęk, pod wpływem lęku, wskutek lęku, motywowane lękiem

To jest jeden i ten sam element logiczny — różne formy, jedna funkcja.

Przedmiot procesu musi być jawnie wskazany

Nominalizacja typu:

- oswajanie
- amortyzacja

- kształtowanie
- reinterpretacja

sama w sobie nie mówi, co jest osvajane, amortyzowane czy reinterpretowane.

Dlatego rdzeń musi zawierać wyraźny przedmiot procesu:

- „oswajanie lęku...”
- „amortyzacja lęku...”
- „redefinicja symboliki żałoby...”

Formy typu:

„oswajanie poprzez ikonografię śmierci”

są błędne — brakuje przedmiotu procesu.

Kotwica musi pozostać agensem

Najczęstszy błąd modelu nominalnego:

„Lęk przed śmiercią – kulturowe kształtowanie doświadczenia żałoby...”

- po myślniku powstaje osobna fraza
- kotwica staje się nagłówkiem tematycznym
- zanika sprawczość

Dlatego potrzebny jest operator sprawczości:

- „Lęk przed śmiercią – kulturowe kształtowanie przez lęk doświadczenia żałoby...”

lub

- „Lęk przed śmiercią inicjuje kulturowe kształtowanie doświadczenia żałoby...”

I dopiero wtedy kotwica pozostaje agensem

Rozróżnienie: przedmiot procesu vs pole działania

W modelu nominalnym częsty błąd to pomylenie:

- przedmiotu procesu (to, co jest przekształcane)

z

- polem / kanałem (to, poprzez co proces się dokonuje)

Przykład poprawny:

„Lęk przed śmiercią – osvajanie lęku ikonografią śmierci w sztuce i literaturze.”

- przedmiot procesu = lęk
- pole = ikonografia śmierci
- kontekst = sztuka i literatura

Kiedy nominalizacja ma sens

Nominalizacja jest użyteczna, gdy:

- forma czasownikowa byłaby nienaturalna lub zbyt ciężka,
- trzeba uniknąć strony biernej,
- potrzebna jest większa precyzja pojęciowa (np. amortyzacja, instytucjonalizacja),
- zależy nam na kontrolowanej strukturze rdzenia.

Nominalizacja nie jest preferowana „dla stylu”.

Modelem podstawowym pozostaje rdzeń czasownikowy — nominalny jest alternatywą, stosowaną świadomie i tylko wtedy, gdy rzeczywiście pomaga.

Długość rdzenia

Obowiązuje twarda reguła: 7–14 słów, licząc wszystkie wyrazy.

- poniżej 7 słów – brakuje zwykle przedmiotu procesu lub kontekstu; rdzeń staje się hasłem („Lęk przed śmiercią a moralność”); embedding robi się zbyt szeroki i niejednoznaczny,
- powyżej 14 słów – zaczyna się mini-esej: dopowiedzenia, kwalifikatory, nawisy.

Nawis to dodatkowy fragment zdania, który nie jest konieczny do opisu procesu (np. rozwlekłe komentarze o naturze współczesności). W rdzeniu trzeba takie fragmenty wycinać – one dokleją modelowi kilka niepotrzebnych wymiarów.

Dobry rdzeń:

- ma jeden proces,
- jasno wskazuje, „na czym” ten proces zachodzi,
- zawiera minimalny, ale wystarczający kontekst (czas, typ kultury, perspektywę),
- mieści się w 7–14 słowach bez kombinowania.

Zbyt krótki rdzeń generuje embedding zbyt ogólny; zbyt długi – embedding zaszumiony, bo model zaczyna „ważyć” dodatkowe dopowiedzenia zamiast rdzenia znaczenia.

Rozbijanie rdzeni zbyt szerokich

Gdy rdzeń obejmuje więcej niż jeden efekt działania, trzeba go rozbić na dwa lub więcej rdzeni. Sygnalizują to:

- dwa niezależne czasowniki procesualne,
- dwa wyraźnie różne przedmioty procesu,
- dwa odległe konteksty.

Zły:

„Lęk przed śmiercią buduje narracje heroiczne i osłabia solidarność wspólnotową.”

Dwa dobre rdzenie:

„Lęk przed śmiercią buduje narracje heroiczne wokół poświęcenia.”

„Lęk przed śmiercią osłabia solidarność wspólnotową w warunkach kryzysu.”

Rozbijanie nie jest formalizmem – chodzi o to, żeby każdy embedding miał jeden dominujący kierunek.

Jeden proces na jeden rdzeń

To konsekwencja poprzedniego punktu, ale warto ją pamiętać jako osobną zasadę:

1 rdzeń = 1 proces.

Rdzeni „łączonych” nie tworzymy. Nie ma rdzeni, które jednocześnie:

- kształtują rytuały,
- rozbijają więzi,
- i generują narracje heroiczne.

Jeśli pojawia się pokusa, by wszystko to zmieścić w jednym rdzeniu, odpowiedź jest prosta: trzeba zrobić trzy rdzenie.

Jedno pole semantyczne

Pole semantyczne to spójny obszar kultury, w którym zachodzi proces: rytuały, symbolika, normy moralne, emocje, narracje, praktyki żałobne, relacje pokoleniowe, struktury rodzinne.

Rdzeń powinien odnosić się do jednego takiego pola. Jeśli łączy pola odległe ontologicznie, embedding zaczyna „mieszać światy”.

Zły typ łączenia:

- rytuały + biotechnologia,
- symbolika + gospodarka,
- emocje + technologie cyfrowe (bez jasnego łącznika).

Jeżeli tekst rzeczywiście opisuje przecięcie pól (np. symbolika żałobna w mediach cyfrowych), można to uchwycić w jednym rdzeniu – ale musi być jasne, że to jedno zjawisko kulturowe, a nie przypadkowe spięcie dwóch obszarów.

Zakaz pytań, negacji i modalności

Rdzeń ma postać twierdzenia, nie:

- pytania („czy”, „jak”, „dlaczego”),
- negacji („nie kształtuje”, „nie wpływa”),
- modalności („musi”, „powinien”, „prawdopodobnie”, „być może”).

Te formy są potrzebne w tekście naukowym (problematyzacja, wątpliwości, hipotezy), ale na poziomie rdzenia są szkodliwe. Embedding nie „rozumie” logiki – modalności i negacje traktuje jako kolejne słowa. W efekcie zamiast czystego wektora procesu powstaje mieszanina procesu + nastawienia epistemicznego.

Rdzeń ma opisywać relację tak, jakby rzeczywiście zachodziła, bo jego zadaniem jest modelowanie mechanizmu kulturowego, a nie stanu badań.

Zakaz strony biernej

Strona bierna zaciera informację o kierunku działania:

- „jest przekształcany”,
- „ulega reinterpretacji”,
- „bywa amortyzowany”.

W logice rdzeni i embeddingów jest to nieakceptowalne. Strona bierna:

- rozmywa pozycję agensa (sprawcy),
- zamienia proces w raport o fakcie („coś się wydarzyło”),
- utrudnia analizę intencjonalności,
- spłaszcza wektor: model widzi działanie, ale gorzej odczytuje „kto na kogo działa”.

Dlatego wszystkie konstrukcje bierne trzeba przepisywać:

- na stronę czynną („Lęk przed śmiercią przekształca rytuały”),
- albo – ostrożnie – na formę nominalną z zachowaniem sprawcy („Lęk przed śmiercią prowadzi do reinterpretacji rytuałów żałobnych”).

W obu przypadkach kierunek procesu pozostaje jasny.

Konsystencja leksykalna

Korpus rdzeni musi korzystać ze stałego słownika bazowego.

W projekcie KR-LPS:

- zawsze „lęk przed śmiercią” (nie zmieniać na: strach, trwoga),
- zawsze „rytuały pogrzebowe” (nie zmieniać na: obrzędy, obyczaje),
- zawsze „symbolika żałobna” (nie zmieniać na: funeralna, pogrzebowa),
- zawsze „kultura świecka” / „kultura religijna” (nie zmieniać na: laicka / sakralna).

Synonimy są kuszące stylistycznie, ale fatalne embeddingowo: każdy inny wyraz tworzy osobny wektor.

Spójność leksykalna:

- stabilizuje klastrowanie,
- ułatwia kontrolę jakości rdzeni,
- zwiększa skuteczność dopasowania pytań użytkownika do embeddingów w module wyszukiującym.

Jeśli koniecznie potrzebujesz innego słowa – tworzysz osobny rdzeń, a nie „drobny wariant”.

Redukcja przymiotników i przysłówków

Dopuszczalne są przymiotniki klasyfikujące (religijny, świecki, symboliczny, instytucjonalny, medyczny). Zakazane są przymiotniki:

- emocjonalne („głęboki”, „dramatyczny”),
- eskalujące wyjątkowość („zaskakujący”, „niezwykły”),

- oceniające („ważny”, „istotny”).

One nie dodają istotnych informacji o procesie, a wprowadzają szum. Model embeddingowy koduje je jako kolejne wymiary – nie jako „ozdobniki”.

Zamiast:

„dynamiczna i gwałtowna transformacja norm moralnych”

w rdzeniu wystarczy:

„transformacja norm moralnych w praktykach społecznych”.

Unikanie metafor i języka figuratywnego

Metafory są świetne dla ludzi, ale toksyczne dla embeddingów. Model nie ma osobnego trybu „czytam metaforycznie” – traktuje wszystkie wyrażenia jako dosłowne.

Dlatego w rdzeniach nie używamy:

- „kultura tańczy z nowoczesnością”,
- „Lęk przed śmiercią wchodzi w dialog z tradycją”,
- „żałoba niesie cień śmierci”,
- „odgrywa teatr śmierci”.

To są dobre zdania eseistyczne, ale złe rdzenie. Język rdzeni ma być techniczny, deskryptywny, bez figur retorycznych. Jeśli zdanie brzmi jak fragment eseju, a nie jak notatka analityczna – trzeba je uprościć.

Oddzielanie perspektyw naukowych

Jeżeli jeden proces jest analizowany z różnych perspektyw teoretycznych, każda z nich powinna dostać osobny rdzeń.

- Socjologiczna – normy, instytucje, role, systemy kontroli.
- Psychologiczna – emocje, mechanizmy obronne, tożsamość.
- Filozoficzna – pojęcia dobra, zła, sensu, śmierci.
- Antropologiczna – rytuały, symbole, struktury przejścia.

Nie należy wciskać dwóch perspektyw w jedno zdanie typu:

„Lęk przed śmiercią w sensie filozoficznym i socjologicznym redefiniuje dobro i normy społeczne.”

To są de facto dwa różne procesy w dwóch różnych polach analizy.

Należy zrobić dwa rdzenie, każdy z własną perspektywą i własnym polem. Wtedy embeddingi nie mieszają dyskursów, ale pozwalają świadomie włączać/wyłączać daną perspektywę w analizie.

Kompatybilność embeddingów i zasada jednego okna czasowego

Wszystkie rdzenie i wszystkie chunki tekstów muszą zostać przetworzone przez model embeddingowy w tym samym oknie czasowym. Oznacza to, że wszystkie embeddingi, które mają być porównywane w FAISS, muszą pochodzić z tej samej wersji modelu.

Dlaczego?

Modele embeddingowe (ME) są aktualizowane przez twórców — modyfikowane są ich parametry, procedury normalizacji, rozkłady wektorów, a nawet struktura przestrzeni semantycznej. Z zewnątrz użytkownik widzi ten sam endpoint API, ale rozrzut wektorowy tego samego zdania może być inny dziś i inny za miesiąc.

Z tego wynika fundamentalna konsekwencja metodologiczna:

Rdzenie i chunki muszą być embeddingowane jednocześnie.

Jeśli chunki zostaną przetworzone w styczniu, a rdzenie w lutym, istnieje wysokie ryzyko, że powstaną dwa zbiory wektorów o niekompatybilnych własnościach geometrycznych.

FAISS nadal policzy podobieństwo, ale:

- wektory będą miały różne „lokalne geografie”,
- kierunki semantyczne będą lekko przesunięte,
- rdzeń x nie będzie wskazywał na te same teksty, które wskazywałby miesiąc wcześniej,
- wyniki porównawcze będą niestabilne i niereplikowalne.

To właśnie nazywa się dryfem embeddingowym: powolnym, ale realnym przesuwaniem przestrzeni semantycznej między kolejnymi wersjami ME.

Konsekwencje praktyczne dla projektu KR-LPS (kulturotwórcza rola lęku przed śmiercią)

- Jednorazowe okno embeddingowania
Wszystkie chunki i wszystkie rdzenie powinny zostać embeddingowane *w jednej sesji lub tej samej dobie*, najlepiej jednym skryptem.
- Nie wolno uzupełniać korpusu rdzeni po czasie, jeśli ma on być porównywany z istniejącymi embeddingami
— chyba że przeliczysz cały korpus od zera.
- Nie wolno aktualizować rdzeni bez re-embeddingowania chunków, nawet jeśli zmianie ulega tylko kilka rdzeni. ME interpretuje teksty względnie względem całej przestrzeni — pojedyncze poprawki wymagają pełnego restartu embeddingów.
- Stabilność FAISS zależy od jednorodności wektorów
FAISS nie „wie”, że embeddingi powstały w różnych wersjach modelu.
Dla niego to tylko punkty w przestrzeni. Jeśli przestrzeń się zmieniła, FAISS nie ma żadnego mechanizmu korekty.

Zależność od czasu działania modelu

Model embeddingowy nie jest „stałym obiektem matematycznym”. Gdy OpenAI publikuje nową wersję, endpoint pozostaje ten sam, ale:

- zmieniają się statystyki normalizacji,
- zmienia się rozłożenie kierunków semantycznych,
- niektóre relacje stają się ostrzejsze, inne słabsze.

Dlatego:

Ponieważ chcemy mieć porównywalność embeddingów — wszystkie embeddingi muszą pochodzić z tego samego momentu.

To jest zasada absolutnie kluczowa i musi być przestrzegana równie rygorystycznie jak zasada 7–14 słów czy zakaz strony biernej.

Checklista walidacyjna rdzeni semantycznych

Checklista służy do jednoznacznej oceny, czy rdzeń jest poprawny embeddingowo i metodologicznie.

Każdy rdzeń musi spełnić wszystkie poniższe warunki.

A. Struktura rdzenia

1. Czy rdzeń zawiera kotwicę (agens) na początku zdania?
(wyjątki możliwe, ale muszą być świadome — standard: kotwica otwiera rdzeń)
2. Czy kotwica pełni funkcję sprawczą (agens przyczynowy), a nie bierny nagłówek?
3. Czy rdzeń zawiera element procesualny?
– czasownik czynny lub nominalizacja z operatorem sprawczości.
4. Czy jest jasno wskazany przedmiot procesu?
(to, co podlega przekształceniu)
5. Czy pole/kontekst działania jest jednoznaczne?
(rytuały, ikonografia, narracje, normy, praktyki żałobne itd.)

B. Wariant procesualny

6. Jeśli rdzeń jest czasownikowy:
– czy użyto czasownika dynamicznego?
– czy unika pseudo-czasowników typu „dotyczy”, „pozostaje w relacji z”?
7. Jeśli rdzeń jest nominalny:
– czy nominalizacja NIE stoi sama po myślniku?
– czy operator sprawczości jest obecny (czasownik inicjujący lub fraza kauzalna)?
– czy kotwica nie stała się nagłówkiem semantycznym?
– czy przedmiot procesu jest jawnie wskazany?
8. Jeśli rdzeń jest autoreferencyjny:
– czy kotwica i przedmiot procesu pełnią różne funkcje semantyczne?
– czy zależność przyczynowa pozostaje czytelna?

C. Jednoznaczność znaczeniowa

9. Czy rdzeń zawiera dokładnie jeden proces?
(zakaz łączenia kilku dynamik w jednym zdaniu)

10. Czy odnosi się do jednego pola semantycznego?
(brak mieszania np. rytuałów + gospodarki, norm + technologii)
11. Czy kierunek działania jest jednoznaczny?
(zakaz strony biernej)
12. Czy rdzeń jest zrozumiały w 1–2 sekundy?
(test „natychmiastowej percepcji”)

D. Higiena językowa

13. Czy rdzeń mieści się w twardym limicie 7–14 słów?
14. Czy unika pytań, negacji i modalności?
(„czy”, „nie”, „powinien”, „prawdopodobnie”, „być może”)
15. Czy unika metafor oraz języka figuratywnego?
(embedding traktuje metafory dosłownie)
16. Czy słownik jest spójny z pozostałymi rdzeniami?
(jedna terminologia: „lęk przed śmiercią”, „symbolika żałobna”, „rytuały pogrzebowe”)
17. Czy nie zawiera zbędnych przymiotników i przysłówków?
(dopuszczalne tylko przymiotniki klasyfikacyjne)

E. Kryterium końcowe

18. Czy rdzeń zachowuje logikę:
kotwica (agens) → proces → przedmiot procesu → pole/kontekst?
19. Czy rdzeń można bez wątpliwości umieścić w jednej grupie rdzeni (klastrze) tematycznych?

Zasady budowy rdzeni semantycznych – skrót

Rdzenie semantyczne – jak opisać kulturę, by mogła ją zrozumieć sztuczna inteligencja?

Gdy analizujemy kulturę – rytuały, symbole, normy, emocje zbiorowe - sztuczna inteligencja działa poprzez modele embeddingowe (ME). Nie „rozumieją” one kultury jak człowiek; przekształcają tekst w wektory, czyli punkty w semantycznej przestrzeni matematycznej. Aby te punkty były jednoznaczne, tekst wejściowy musi być wyjątkowo precyzyjny.

Dlatego powstała koncepcja rdzenia semantycznego – najmniejszej jednostki opisu procesu kulturowego, którą AI może odczytać stabilnie i konsekwentnie. Rdzeń nie opisuje wszystkiego, co wiemy. On „chwytą” sedno: *co* działa na *co*, *jak* i *w jakim kontekście*.

Rdzenie używane są w systemach wyszukiwania semantycznego (ME + FAISS, czyli), które potrafią odnajdywać w dużych zbiorach tekstów fragmenty odpowiadające określonemu mechanizmowi kulturowemu. Aby takie działanie było możliwe, każdy rdzeń musi być jednoznaczny, procesualny, spójny semantycznie i precyzyjnie zbudowany.

Każdy rdzeń zaczyna się od tego, co napędza proces – czynnika sprawczego. Ten czynnik nazywamy kotwicą. W projekcie „Kulturotwórcza rola lęku przed śmiercią” (KR-LPS) kotwicą jest wyrażenie „lęk przed śmiercią (LPS)”, rozumiany nie jako emocja jednostki, lecz siła kulturotwórcza wymuszająca rytualne, symboliczne i normatywne odpowiedzi.

Dzięki kotwicy wszystkie rdzenie stanowią jedną rodzinę i AI nie „gubi się”, o jakim zjawisku mówimy. Jeśli kotwicy zabraknie, lub jeśli zostanie postawiona gdzieś poza początkiem zdania, model może błędnie uznać rdzeń za należący do innego tematu.

Rdzeń musi być procesem. Nie może być tytułem („lęk przed śmiercią a moralność”) ani obserwacją („lęk przed śmiercią w kulturze jest ważny”). Musi opisywać dzianie się, czyli proces kulturowy: przekształcanie symboli, formowanie rytuałów, wytłumianie emocji, inspirowanie narracji.

Każdy rdzeń – niezależnie od formy językowej – ma tę samą strukturę:

K kotwica (agens) → **EP** element procesualny → **PP** przedmiot procesu → **PKK** pole / kanał / kontekst

Przykład:

Lęk przed śmiercią (**K**) → kształtuje (**EP**) → doświadczenie żałoby (**PP**) → poprzez praktyki żałobne w kulturze miejskiej XX wieku (**PKK**).

- Kotwica (agens) – główny czynnik sprawczy rdzenia (w projekcie KR-LPS: „Lęk przed śmiercią”).
- Element procesualny – proces, który się dokonuje (np. kształtuje, transformuje, oswaja, reinterpretuje, amortyzuje, kształtowanie, transformacja, oswajanie, reinterpretacja, amortyzacja).
- Przedmiot procesu – to, co bezpośrednio podlega działaniu (lęk, doświadczenie żałoby, normy, relacje, symbolika, rytuały).
- Pole / kanał / kontekst działania – to, w czym i poprzez co proces się realizuje (rytuały, ikonografia, instytucje, praktyki żałobne, typ kultury, epoka).

Pozostałe zasady budowy rdzenia

- Długość: 7–14 słów.
- Jeden proces na rdzeń: brak łączenia kilku dynamik (np. ‘buduje narracje’ i ‘osłabia solidarność’).
- Jedno pole semantyczne, żadnych połączeń typu: rytuały + biotechnologia.
- Zakaz pytań, negacji, modalności: rdzeń musi być twierdzeniem.
- Zakaz strony biernej: zawsze strona czynna lub nominalizacja z jawnym agensem.
- Stały słownik: zakaz używania synonimów (lęk, strach, obawa, panika).
- Brak metafor: embeddingi traktują metafory dosłownie.
- Oddzielanie perspektyw naukowych: osobny rdzeń dla socjologii, psychologii, antropologii itd.

Ważna zasada techniczna – jedno okno czasowe

Modele embeddingowe (ME) zmieniają się w czasie: aktualizują parametry i sposób odwzorowania znaczeń w przestrzeni wektorowej. Dlatego fragmenty tekstów (chunki) i rdzenie muszą być embeddingowane (zamieniane na wektory) w tym samym momencie. Jeśli chunki powstaną dziś, a rdzenie za miesiąc, ich wektory będą pochodziły z różnych „wersji”

przestrzeni semantycznej. FAISS nadal je porówna, ale wyniki będą mniej stabilne i trudniejsze do replikacji. Stąd reguła: wszystkie embeddingi muszą być generowane w jednym oknie czasowym.

wersja 1 07.01.2026