

# System diagnostyki Słowosieci

Agnieszka Dziob, Barbara Nowakowska

link: <https://tracker.clarin-pl.eu/diagnostics/>

System diagnostyki Słowosieci został oparty na zapytaniach do bazy danych, napisanych w języku SQL<sup>1</sup>. Wszystkie zapytania można podzielić na trzy grupy:

- diagnozę podstawową,
- diagnozę synonimii,
- diagnozę szczegółową.

## Diagnoza podstawowa

Aby synset i znajdujące się w nim jednostki nie były osobnymi bytami, lecz stanowiły element sieci zależności, musi zostać spełnione pewne minimum opisu. To, czy owo minimum zostało spełnione oraz zrealizowane poprawnie, sprawdzają zapytania oznaczone jako **diagnoza podstawowa**. Wymagania, które muszą zostać spełnione, są swoiste dla każdej części mowy, a w przypadku przymiotników i po części czasowników, także dla danej domeny (kategorii pomocniczej). Wyjątkiem jest relacja bliskoznaczności, której zasady użycia są ściśle związane z system kwalifikatorów.

**Zasady działania każdego z zapytań diagnozy podstawowej, czyli co jest przez nie sprawdzane:**

- **dla rzeczownika**

obecność relacji pionowej, w przypadku rzeczownika: **hipo-hiperonimii** lub relacji **typ-egzemplarz**; wyjątek od tej reguły stanowi obecność **wzajemnej mero-/holonimii** lub relacji jednostek: **żeńskość, deminutywność, ekspresywność|augmentatywność** - jednostki i synsety z tymi relacjami nie wymagają relacji pionowej

---

<sup>1</sup> Zapytania zostały udostępnione na Clarin-PL Nextcloud: <https://nextcloud.clarin-pl.eu/index.php/apps/files/?dir=/System%20diagnostyki%20S%C5%82owosieci%20-%20zapytania&fileid=215180>

- **dla czasownika**

obecność relacji pionowej, w przypadku czasownika: **hipo-hiperonimii**

- **dla przymiotnika relacyjnego**

relacja pionowa nie jest tu obowiązkowa, dlatego zapytanie sprawdza, czy występująca relacja hipo-hiperonimii łączy ze sobą tylko przymiotniki relacyjne; sprawdzane jest także, czy w relacji jednostek jest obecna **synonimia międzyparadygmatyczna dla relacyjnych** lub **derywacyjność**; przyczyny i konsekwencje tej decyzji zostały opisane w innej części dokumentu

- **dla przymiotnika materiałowego**

obecność relacji pionowej, w przypadku przymiotnika materiałowego: **hipo-hiperonimia** (tylko przymiotnik materiałowy z innym przymiotnikiem materiałowym), lub **wartość cechy**

- **dla przymiotnika jakościowego**

obecność relacji pionowej, w przypadku przymiotnika jakościowego: **hipo-hiperonimia** (tylko przymiotnik jakościowy z innym przymiotnikiem jakościowym) lub **wartość cechy**; wyjątek od tej reguły stanowi obecność relacji **stopnia wyższego** lub **stopnia najwyższego** - jednostka z tą relacją nie wymaga relacji pionowej

- **dla przysłówka**

obecność relacji pionowej, w przypadku przysłówka: **hipo-hiperonimia** (tylko przysłówek z innym przysłówkiem) lub **wartość cechy**; wyjątek od tej reguły stanowi obecność relacji **stopnia wyższego** lub **stopnia najwyższego** - jednostka z tą relacją nie wymaga relacji pionowej

Dodatkowo dla każdej części mowy sprawdzana jest poprawność zastosowania relacji bliskoznaczności.

## **Diagnoza synonimii**

Diagnoza synonimii w swojej istocie polega na sprawdzaniu, czy w jednym synsecie znajdują się jednostki o zgodnych kwalifikatorach oraz, w przypadku czasowników stanowych, przymiotników jakościowych, przymiotników relacyjnych i przymiotników materiałowych, czy są w synsetach tylko same ze sobą.

### **Diagnoza szczegółowa**

Do diagnozy szczegółowej należą te zapytania, które dotyczą kwestii wychodzących poza diagnozę podstawową, lub które, ze względów technicznych, musiały zostać ujęte w osobnych zapytaniach.

#### **Opis działania każdego z zapytań diagnozy szczegółowej:**

- **czasownik stanowy a relacje: stanowość, kauzacja, procesywność, dystrybutywność, inchoatywność**

Wśród relacji synsetów dedykowanych dla czasownika są takie, którymi można łączyć wyłącznie czasowniki stanowe, oraz takie, które wykluczają taką możliwość. Zapytanie zostało skonstruowane w taki sposób, aby sprawdzało, czy zostały spełnione następujące warunki:

- relacja stanowości łączy tylko czasownik stanowy z rzeczownikiem, przymiotnikiem lub przysłówkiem
- w przypadku kauzacji, procesywności i inchoatywności, czasownik, od którego wychodzi relacja, nie jest czasownikiem stanowym
- czasownik stanowy nie wchodzi w relację dystrybutywności

Każda jednostka, która nie spełnia chociaż jednego z powyższych warunków, jest traktowana przez system jako opisana błędnie.

- **przymiotnik jakościowy: 'nie' a derywacyjność i antonimia**

Dla przymiotników jakościowych, które są słowami zaprzeczonymi, a podstawa słowotwórcza stanowi jednocześnie ich antonim, zasady postępowania nakazują dodanie dwóch relacji – antonimii i derywacyjności. Zapytanie ma za zadanie sprawdzić, czy powyższa zasada postępowania jest stosowana.

- **Przymiotnik materiałowy: relacja rola: materiał**

Pewne relacje synsetów i jednostek są bardziej precyzyjne od innych i mają węższe zastosowanie. W sytuacji, kiedy można przypisać dwie relacje, ale jedna z nich lepiej oddaje istotę związku między jednostkami lub synsetami, ma ona pierwszeństwo przed tą, która jest bardziej ogólna i uniwersalna. Jedną z takich relacji jest derywacyjność, którą stosuje się, kiedy związek między podstawą słowotwórczą a wyrazem pochodnym istnieje, ale nie można go opisać za pomocą żadnej innej relacji. Zapytanie to sprawdza, czy obecne wśród przymiotników materiałowych relacje derywacyjności są uzasadnione, a kiedy na jej miejsce powinna pojawić się relacja roli: materiału.

## **Prezentacja wyników zapytań**

Wyniki każdego z zapytań diagnostycznych są prezentowane w formie tabeli. Zawartość tabel ulega odświeżeniu po każdym kliknięciu w link przekierowujący, a więc wyświetlane dane są zawsze aktualne i odzwierciedlają rzeczywisty stan SłowoSieci.

Wyniki są szeregowane w tabelach w porządku alfabetycznym, za wyjątkiem tych dotyczących synonimii. Tam są one sortowane rosnąco według ID synsetu.

Oznaczenia kolumn oraz ich znaczenie:

- SYN\_ID – unikalny numer synsetu (ID synsetu)
- unitsstr – zawartość danego synsetu, a więc jakie jednostki się w nim znajdują oraz jakie mają domeny (kategorie pomocnicze)
- lemma – nazwa lematu wraz z numerem wariantu
- domain – domena (kategoria pomocnicza); objaśnienie oznaczeń domen znajduje się w oddzielnym dokumencie
- comment – treść pola „komentarz”
- owner – osoba, która stworzyła dane hasło lub dokonywała jego znaczącej edycji jako ostatnia
- status – status, który posiada dana jednostka:

- NPRZ - nieprzetworzony
- NOW - nowy
- BŁD - błąd
- SPR - sprawdzony
- ZN - znaczenie
- CZ.PRZ - częściowo przetworzony
- error\_comment – treść pola “opis błędu”

## Słabe strony systemu

Pomimo wszelkich starań, aby zaproponowany system diagnostyczny był jak najbardziej efektywny i funkcjonalny, wciąż pozostaje miejsce na pewne usprawnienia i udoskonalenia. Poniżej zostały opisane główne słabe strony konkretnych zapytań oraz jak można by je rozwiązać.

### Diagnoza podstawowa: przymiotnik relacyjny

W dokumencie *Przymiotniki w Słowności 3.0* czytamy: "Przymiotniki relacyjne łączą się ze swoimi rzeczownikami za pomocą relacji synonimii międzyparadygmatycznej dla relacyjnych. Relacja ta oznacza transpozycję, tj. zmianę części mowy wyrazu bez zmiany znaczenia. Wszystkie przymiotniki powiązane synonimią międzyparadygmatyczną dla relacyjnych są relacyjne, **nie każdy jednak przymiotnik relacyjny musi mieć swoją relację synonimii międzyparadygmatycznej.**" Jeżeli wyjdzie się z założenia, że istnieją jednostki, dla których zapomniano dodać relację syn. mpar. dla rel., zapytanie diagnostyczne powinno być skonstruowane tak, aby pokazywało wszystkie jednostki, które nie zawierają tej relacji. Wadą takiego rozwiązania jest mieszanie się jednostek opisanych w sposób prawidłowy, czyli które nie powinny zawierać relacji syn. mpar. dla rel., z tymi, dla których brak relacji jest błędem opisu. W takiej sytuacji jedynie lingwista jest w stanie zweryfikować, które z wyników zapytania są, a które nie są uzasadnione. Rozwiązaniem mogłoby być ręczne przejście wygenerowanych wyników i oznaczenie w komentarzu tych jednostek, dla których relacja syn. mpar. dla rel. być nie powinno, a następnie zmodyfikować zapytanie w taki sposób, aby nie pokazywało wyników z tymże oznaczeniem.

### **Przymiotnik jakościowy: 'nie' a derywacyjność i antonimia**

Ze względu na pewne ograniczenia, zapytanie diagnozujące to zagadnienie wskazuje zarówno jednostki, które w myśl powyższego cytatu zostały opisane błędnie, jak i te, których zapytanie w ogóle nie dotyczy, gdyż w ich przypadku 'nie' jest ciągiem znaków przynależnych do tematu, np. 'niebieski', 'niebiański', 'niedźwiedzi', 'niemiecki', 'niechlujny'. Wydaje się, że i tutaj warto zastosować rozwiązanie zaproponowane powyżej. Dzięki temu generowane przez zapytanie wyniki będą celniejsze.

### **Kwalifikatory a diagnoza synonimii oraz relacja bliskoźnaczości**

Diagnoza zgodności kwalifikatorów dla synonimii polega obecnie na odfiltrowaniu tych synsetów, w których kwalifikatory jednostek są identyczne. W konsekwencji oznacza to, że wynikiem takiego zapytania są wszystkie synsety, które mają niejednolite kwalifikatory, bez względu na to, czy w myśl wytycznych opis jest prawidłowy, czy nie.

Podobny problem wiąże się z diagnozą relacji bliskoźnaczości. Według procedur należy wyłączać z synsetu jednostki z kwalifikatorem pot., jeśli miałyby się znajdować w jednym synsecie z jednostkami z kwalifikatorami książk., specj. lub urz. Obecnie skonstruowane zapytania nie biorą tego pod uwagę i każde łączenie bliskoźnaczością synsetu, w którym znajduje się słownictwo oznaczone jako potoczne, jest wychwytywane jako błąd.

Oba przedstawione problemy nie uniemożliwiają wykonywania czynności diagnostycznych, jednak znacząco wpływają na jakość otrzymywanych wyników. Dlatego też warto możliwie jak najszybciej je rozwiązać.