

Reguły anonimizacji

Wersja z dnia 4 lipca 2012 r.

1. Wprowadzenie

Publikację orzeczeń w Portalu orzeczeń zawsze musi poprzedzać proces anonimizacji ich treści. Anonimizacja jest to proces przetwarzania treści w taki sposób, aby uniemożliwić identyfikację występujących w dokumencie osób fizycznych (a dokładniej, sprawić, by ta identyfikacja nie była możliwa bez "nadmiernych kosztów, czasu lub działań"). Z drugiej strony, należy starać się, by zanonimizowany dokument w jak największym stopniu zachował swoją czytelność i nie zgubił żadnych treści natury ogólnej. Można zatem powiedzieć, że proces anonimizacji jest swoistym kompromisem między udostępnianiem informacji w celu ich upublicznienia, a potrzebą przestrzegania obowiązujących norm prawnych gwarantujących ochronę prawną określonych dóbr i wartości.

Należy zwrócić uwagę, że pewne informacje bardzo często umożliwiają łatwą identyfikację podmiotów prawa, zwłaszcza osób fizycznych, w związku z czym powinny być usuwane z treści orzeczeń (dokumentów). Są to między innymi: imiona, nazwiska, adresy osób fizycznych, dane identyfikujące inne podmioty prawa np. osoby prawne. Z drugiej strony trzeba mieć na względzie, iż niektóre informacje odnoszące się do określonych instytucji publicznych powinny być publicznie dostępne i nie należy ich usuwać z treści orzeczeń. W dalszej części dokumentu przedstawione zostaną reguły określające, jakie konkretnie dane podlegają, a jakie nie podlegają anonimizacji.

Anonimizacja polega na wykonaniu jednej z czterech czynności:

- Zamianie całej frazy na pojedynczy inicjał. Stosujemy ją przy anonimizacji miejscowości i nazw geograficznych oraz występujących samotnie w tekście imion lub nazwisk.
- Zamiana całej frazy na pojedynczy inicjał, po którym następuje wielokropek „(...)..”
Stosujemy ją w przypadku nazw urzędów, pojazdów oraz handlowych nazw różnych substancji czy też innych produktów, w sytuacji gdy mają więcej niż 1 wyraz.
- Zamiana całej frazy na parę inicjałów. Stosujemy ją jedynie w przypadku osób.
- Zamiana pojedynczych wyrazów lub dłuższych ich ciągów na wielokropek „(...)..”
Stosujemy ją w pozostałych przypadkach.

2. Anonimizowane frazy

W celu ułatwienia anonimizacji sporządzone zostały bardziej szczegółowe zasady, wg których należy ją przeprowadzać. W kolejnych podpunktach wymienione zostaną różne kategorie fraz (sformułowań) występujących w tekście, wraz z informacją, jak należy z tymi frazami postępować.

2.1. Osoby

Osoby występujące w treści orzeczenia zamieniane są na inicjały. Stosowane są reguły:

- imię i nazwisko zamieniane są na inicjały np. *Jan Kowalski* anonimizowany jest do *J. K.*
- jeżeli we frazie występuje więcej imion bądź nazwisk, brane są pod uwagę tylko pierwsze imię i ostatnie nazwisko np. *Jan Marian Kowalski-Nowak* anonimizowany jest do *J. N.*
- jeżeli w tekście występuje więcej osób o tych samych inicjałach, w zanonimizowanej wersji pojawiają się liczby pomagające określić, której z nich dotyczy dany fragment. Przykładowo tekst *Jan Kowalski* oraz *Jakub Konieczny* założyli spółkę. Przewodniczącym został *Jan Kowalski*. anonimizowany jest do *J. K. (1)* oraz *J. K. (2)* założyli spółkę. Przewodniczącym został *J. K. (1)*.
- jeżeli w tekście występuje pojedyncze imię, bądź nazwisko, w zanonimizowanej wersji powinny pojawić się pojedyncze inicjały. Przykładowo tekst *Jan* oraz *Tomasz* poszli w odwiedziny do państwa *Wróblewskich*. anonizowany jest do *J.* Oraz *T.* poszli w odwiedziny do państwa *W.*. Takich pojedynczych inicjałów nie numerujemy.

WYJĄTEK 1: Anonimizacji **nie** podlegają następujące podmioty:

- sędziowie składu orzekającego,
- protokolanci,
- prokuratorzy,
- komornicy sądowi.

WYJĄTEK 2: Anonimizacji **nie** są poddawani autorzy cytowanych książek, komentarzy, glos oraz artykułów naukowych.

2.2. Miejscowości

Miejscowości występujące w treści orzeczeń zamieniane są na *pojedynczy inicjał*. W przypadku, gdy nazwa miejscowości składa się z kilku wyrazów, tak jak np. *Kąty Wrocławskie*, w zanonimizowanej wersji pojawia się inicjał tylko pierwszego z nich.

WYJĄTEK 1: Anonimizacji nie podlegają miasta opisujące siedzibę sądu, na przykład fraza „Sąd Apelacyjny we Wrocławiu” nie podlega anonimizacji.

WYJĄTEK 2: Anonimizacji nie podlegają miasta określające miejsce wydania nieanonimizowanej książki.

2.3. Nazwy geograficzne

Przymiotniki związane z nazwami geograficznymi zmieniane są na wielokropek „(...)”. W szczególności dotyczy to województw i powiatów. Rzeczownikowe nazwy geograficzne, pisane wielką literą, zamieniane są, tak jak miejscowości, na pojedynczy inicjał. Dotyczy to między innymi:

- rzek,
- jezior,
- szczytów górskich,
- dzielnic miast.

Natomiast w nazwach dróg ich numery (i inne oznaczenia) zamieniamy na wielokropek „(...)”.

2.4. Firmy, spółki, spółdzielnie, lokale usługowe

W nazwach:

- firm,
- spółek,
- spółdzielni

występujących w treści orzeczenia, wszystkie nazwy własne oraz słowa mogące posłużyć do zidentyfikowania danej organizacji zamieniane są na wielokropek „(...)”. Pozostają jedynie określenia organizacyjno-prawne, takie jak:

- typy spółek np. spółka z ograniczoną odpowiedzialnością, spółka jawna, spółka cywilna,
- typy przedsiębiorstwa, np. przedsiębiorstwo produkcyjno-handlowo-usługowe,
- skróty: s. a., sp. z o. o.

2.5. Organizacje, fundacje, instytucje

Organizacje, fundacje i instytucje anonimizowane są podobnie jak firmy, spółki, spółdzielnie (patrz punkt poprzedni).

WYJĄTKI::

- Nie anonimizujemy nazw organów, urzędów oraz instytucji publicznych (choć anonimizujemy informację o miejscu, w którym znajduje się ten urząd lub organ – właściwość miejscową podmiotu).
- Nie anonimizujemy nazw organizacji międzynarodowych, których członkami są państwa (na przykład Światowa Organizacja Zdrowia, Międzynarodowy Fundusz Walutowy, UNESCO).
- Nie anonimizujemy nazw organizacji ustalających różnego rodzaju normy, takich jak: American National Standards Institute, Polski Komitet Normalizacyjny.
- Nie anonimizujemy Polskiej Akademii Nauk i analogicznych instytucji z innych krajów.
- Nie anonimizujemy rodzajowych nazw szkół wyższych, takich jak: Politechnika, Uniwersytet, Akademia Medyczna, Akademia Wychowania Fizycznego, Akademia Muzyczna, Akademia Sztuk Pięknych.
- Nie anonimizujemy organów samorządu zawodowego lub gospodarczego (takich jak na przykład Naczelna Izba Lekarska, Krajowa Izba Biegłych Rewidentów, Naczelna Rada Adwokacka, izba rzemieślnicza, izba gospodarcza, izba rolnicza, Związek Rzemiosła Polskiego).
- Nie anonimizujemy nazw sądów, takich jak Sąd Rejonowy, Naczelny Sąd Administracyjny.
- Nie anonimizujemy nazwy – Skarb Państwa.

2.6. Kościoły i związki wyznaniowe

Anonimizujemy zarówno nazwy kościołów i związków wyznaniowych jak i informacje, pozwalające określić przynależność religijną konkretnych gmin wyznaniowych i parafii.

2.7. Adresy

Adresy zamieniane są na wielokropek „(...)”. Pozostawiana jest tylko część informująca o tym, czego dotyczy dana fraza np.:

- „ul.”,
- „plac”,
- „skwer”.

2.8. Numery i identyfikatory

Wszystkie identyfikatory, zarówno osób, organizacji, miejsc czy przedmiotów podczas anonimizacji powinny być zmieniane na wielokropek „(...)”. Dotyczy to w szczególności następujących kategorii:

- numery PESEL, KRS, REGON, NIP, itd.,
- numery ksiąg wieczystych,
- numery umów i faktur,
- numery licencji,
- numer paszportu, numer i seria dowodu osobistego (lub dowodu rejestracyjnego), numer legitymacji, prawa jazdy, prawa wykonywania zawodu bądź innego dokumentu,
- numery rejestracyjne samochodów, numery VIN,
- numery fabryczne różnych produktów, numery silnika pojazdu, numery podwozia, numery nadwozia,
- numery działek,
- numery szkół, przedszkoli i żłobków,
- numery telefonów i faksów,
- numer wpisu do ewidencji działalności gospodarczej,
- identyfikatory używane w internecie (z wyjątkiem tych, które przypominają normalne imiona i nazwiska).

2.9. Nazwy urzędzeń, pojazdów, nazwy różnych produktów

W celu zwiększenia bezpieczeństwa procesu anonimizacji w Portalu orzeczeń anonimizacji poddawane są także nazwy urzędów, pojazdów oraz innych produktów, przy czym nazwę jednowyrazową zamieniamy na inicjał, natomiast w przypadku nazw dłuższych do inicjału doklejamy wielokropek „(...)”.

2.9 Pozostałe słowa i frazy wymagające anonimizacji

- Nazwy miejscowości występujące w nazwach jednostek organizacyjnych (wydziałów, komórek) organów i urzędów, np. Wydział Komunikacji Urzędu Rejonowego w Brzegu.
- Nazwy dookreślające właściwość miejscową organów administracji publicznej, np. Wojewoda Dolnośląski albo Prezydent Wrocławia.
- Długie ciągi cyfr, w sytuacji, gdy nie jest oczywista ich dokładniejsza klasyfikacja, na przykład 346565356312677.
- Ciągi liczb, liter, cyfr, myślników, gdy nie jest oczywista ich dokładniejsza klasyfikacja, np. 4565-DF0-34cvf44.
- Treść haseł reklamowych i innych napisów.
- Nazwy zespołów muzycznych, drużyn sportowych.

3. Nieanonimizowane treści orzeczeń

Anonimizacji nie podlegają treści orzeczeń (frazy), które nie zawierają danych oraz informacji „niebezpiecznych”, czyli umożliwiających identyfikację podmiotu prawa. W szczególności są to treści związane z prawniczym charakterem orzeczeń, tzn. należą do kategorii języka prawnego (np. treść przepisów prawa) bądź języka prawniczego (argumentacja prawnicza, komentarz normy prawnej, wykładnia prawa czy doktryna prawa).

3.1. Terminologia prawnicza

Terminologia prawnicza obejmuje zarówno język prawny, jak i prawniczy oraz wszelkie pojęcia wynikające z instrukcji kancelaryjnej wymiaru sprawiedliwości. W praktyce stosowania owej terminologii można napotkać również sformułowania (frazy) dotyczące elektronicznych systemów wyszukiwania informacji prawniczej. Poniżej przykłady takiej terminologii nie podlegającej procesowi anonimizacji.

3.2. Oznaczenia czasu

Zasadą jest, że oznaczenia czasu czyli informacje o latach, miesiącach, dniach, godzinach, przedziałach czasowych pozostają nieanonimizowane.

WYJĄTEK: Informacje o dacie urodzenia konkretnej osoby.

3.3. Pozostałe nieanonimizowane treści

Treści orzeczeń (frazy) wymienione w tym podpunkcie również nie są anonimizowane, ponieważ ich pozostawienie w tekście orzeczenia nie wpływa na możliwość zidentyfikowania podmiotu wskazanego w orzeczeniu, stąd nie stanowi to zagrożenia dla jego interesu prawnego. W ten

sposób użytkownik Portalu orzeczeń może się zapoznać w szerszym stopniu z okolicznościami faktycznymi sprawy, co istotnie wpływa na możliwość zrozumienia sposobów stosowania i wykładni prawa przez wymiar sprawiedliwości.

Poniżej konkretne typy treści orzeczeń, które **nie** są anonimizowane.

Typ frazy	Przykładowe frazy
Nazwy części instytucji (wydziałów, oddziałów, katedr, zakładów itp.)	Wydział Matematyki i Informatyki Oddział Intensywnej Opieki Medycznej Zakład Medycyny Sądowej
Liczby porządkowe, liczby pewnych zdarzeń, osób, przedmiotów	pozwany nie zapłacił 3. raty, jak również raty nr 4 zadał 23 ciosy nożem miała 2 synów i 3 córki w firmie było zatrudnionych 1345 pracowników jego spółka posiadała 32 samochody dostawcze
Wykonywany (lub wyuczony) zawód, stanowisko służbowe oraz pełnione funkcje	w firmie pełniła obowiązki Głównego Księgowego najpierw pracował jako weterynarz, a następnie jako informatyk
Choroby oraz rodzaje terapii	Odmówiła przyjęcia do pracy ze względu na nieżyt górnych dróg oddechowych Zmuszony był wykonać otolaryngologiczny zabieg otwarcia przedniej ściany tchawicy
Kwota pieniężna	zasądzenia kwoty 110386zł 10gr z ustawowymi odsetkami w kwocie 46.179.60 zł. czyli 30,07zł brutto za stronę słownie: dwieście pięćdziesiąt złotych i piętnaście groszy
Nazwy państw	Przebywał na terenie Niemiec do maja mieszkania własnościowego we Francji o powierzchni wyeliminowania z rynku w Polsce jedyne go wówczas
Liczby z jednostkami	przebieg pojazdu w wysokości 325.328 km zakupionej kukurydzy w ilości 5.000 ton powierzchnię użytkową 80 m. kw.

Tabela 12: Inne frazy, które nie podlegają anonimizacji.

4. Praktyka anonimizacji

4.1. Rola człowieka w procesie anonimizacji

W procesie anonimizacji człowiek musi wykonać 3 czynności

1. Zdecydować, czy dane orzeczenie może zostać opublikowane po anonimizacji. W szczególności oznacza to sprawdzenie, czy:
 - o nie zawiera ono danych objętych klauzulą tajności,
 - o dane orzeczenie (lub jakiś jego fragment) nie narusza tajemnicy przedsiębiorcy. **Uwaga:** Zakładamy, że w każdym orzeczeniu po anonimizacji nie da się zidentyfikować przedsiębiorcy. Stąd nie jest ujawnieniem tajemnicy przedsiębiorcy zdanie, które po zanonimizowaniu brzmi: "Średnia pensja w (...) S.A. wynosiła w ostatnim kwartale 3.023 zł brutto.". Natomiast może być takim ujawnieniem informacja o tym, jak dokładnie zorganizowana jest linia produkcyjna w "(...) S.A." (bo ta informacja jest użyteczna dla konkurencji również wtedy, gdy tożsamość przedsiębiorcy pozostaje ukryta).
 - o określone orzeczenie nie należy do zbioru orzeczeń wyłączonych z obowiązku publikacji w Portalu orzeczeń (przesądza o tym Zarządzenie Prezesa Sądu, na podstawie którego w danym Sądzie funkcjonuje Portal).
2. Poprawić ewentualne błędy, koncentrując się na błędach *krytycznych* (punkt 4.3.)
3. Zatwierdzić dokument do publikacji.

Dzięki zgłaszanym przez użytkowników wewnętrznych błędom jakość automatycznej anonimizacji stale się poprawia, zatem należy się spodziewać, że pracochłonność punktu drugiego będzie się ciągle zmniejszała (aż do momentu, w którym prawdopodobieństwo popełnienia błędu będzie na tyle niewielkie, że etap ten będzie można wyeliminować). W tym celu prosimy korzystać z funkcji – zgłoś błąd, opisując konkretne przypadki błędnej anonimizacji.

4.2. Warianty anonimizacji

Istnieją sytuacje, w których istnieje więcej niż jedna akceptowalna metoda anonimizacji. W takich sytuacjach zaleca się pozostawianie anonimizacji wybranej przez algorytm.

Przykładowa fraza	Możliwe anonimizacje
Klub Sportowy "Tęcza"	Klub (...)
	Klub Sportowy (...)
	Kancelaria (...)
Kancelaria Adwokacka Janiny Nowak	Kancelaria Adwokacka (...)
	Kancelaria Adwokacka J. N.
	(...) S.A.
Zakłady Naprawcze Taboru Kolejowego w Oleśnicy S.A.	Zakłady (...) S.A.
	Zakłady (...) w O. S.A.

Tabela 13: Przykładowe frazy i warianty ich anonimizacji.

4.3. Błędy anonimizacji

Anonimizacja opiera się na wynikach działania różnych algorytmów związanych z przetwarzaniem języka naturalnego, zaś tego typu algorytmy zawsze mają pewien procent sytuacji, w których podejmują błędne decyzje. Tak popełnione błędy mogą się różnić rangą. Najpoważniejsze błędy (błędy krytyczne) polegają na bezpośrednim ujawnieniu tożsamości osoby fizycznej. Następne w kolejności są błędy, umożliwiające identyfikację firmy, organizacji, miejscowości lub jakiejś części adresu. Nieco mniej istotne są błędy, w których niepotrzebna anonimizacja powoduje spadek czytelności tekstu, a najmniej istotne są takie, w których niepotrzebna (lub niewłaściwa) anonimizacja nie utrudnia czytania tekstu.

Pewne frazy mogą być zakwalifikowane na różne sposoby, z których każdy (zgodnie z zasadami) prowadzi do innej anonimizacji. Prezentuje to poniższa tabelka:

Przykładowa fraza	Kwalifikacja 1	Anonimizacja 1	Kwalifikacja 2	Anonimizacja 2
Mercedes Benz	Osoba	M. B.	Marka samochodu	M. (...)
Arthur Andersen	Osoba	A. A.	Firma	(...)
Kazimierz Dolny	Osoba	K. D.	Miasto	K.

W takich sytuacjach zaleca się akceptowanie wyboru algorytmu (nawet gdyby był niewłaściwy), o ile nie powoduje to istotnego zwiększenia niebezpieczeństwa identyfikacji, albo znacznego spadku czytelności tekstu.

Należy też zwrócić uwagę, że częstą przyczyną niewłaściwych anonimizacji są błędy osób, które wprowadzają tekst orzeczenia. Tytułem przykładu, algorytm zakłada, że *Jan Kowalski* i *Jan Kowlaski* są różnymi osobami, co w pewnych sytuacjach może być prawdą, ale może też się zdarzyć, że w drugim przypadku popełniony został błąd, i tak naprawdę mamy do czynienia z tą samą osobą.