

Reguły anonimizacji

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 konkretne dane podlegają, a jakie nie podlegają anonimizacji.

Anonimizacja polega na wykonaniu jednej z czterech czynności:

1. Zamianie całej frazy na pojedynczy inicjał. Stosujemy ją przy anonimizacji miejscowości i nazw geograficznych oraz występujących samotnie w tekście imion lub nazwisk.
2. 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.
3. Zamianie całej frazy na parę inicjałów. Stosujemy ją jedynie w przypadku imion i nazwisk osób fizycznych.
4. Zamianie pojedynczych wyrazów lub dłuższych ich ciągów na wielokropek „()”. Stosujemy ją w pozostałych przypadkach.

Treść orzeczenia	Wersja zanonimizowana
Pan Nowak mieszkał w Kędzierzynie - Koźlu nad rzeką Odrą.	Pan N. mieszkał w K. nad rzeką O..
Wysiadła z samochodu marki Fiat 126p, roztaczając zapach perfum Chanel No. 5.	Wysiadła z samochodu marki F. (...), roztaczając zapach perfum C. (...).
Anna Kowalska nigdy nie lubiła Piotra Nowaka.	A. K. nigdy nie lubiła P. N..
Zakłady Produkcji Kolejek "Ciuchcia" mają siedzibę w powiecie oleśnickim.	Zakłady (...) mają siedzibę w powiecie (...).

Tabela 1: Przykłady typów anonimizacji.

2. Anonimizowane treści orzeczeń

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 treści orzeczenia, wraz z informacją, jak należy z tymi frazami (pojęciami) postępować.

2.1. Dane osobowe osób fizycznych (zwłaszcza imion i nazwisk)

Imiona i nazwiska osób fizycznych występujące w treści orzeczenia zamieniane są na inicjały. Stosowane są następujące 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. Prezesem został Jan Kowalski. anonimizowany jest do J. K. (1) oraz J. K. (2) założyli spółkę. Prezesem 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. anonimizowany jest do J. oraz T. poszli w odwiedziny do państwa W. . Takich pojedynczych inicjałów nie numerujemy.

WYJĄTEK 1: Anonimizacji nie podlegają podmioty:

1. PEŁNIĄCE FUNKCJE PUBLICZNE

2. FUNKCJONARIUSZE PUBLICZNI

Tytułem przykładu wskazać można na:

- Sędziów składu orzekającego.
- Protokolantów.
- Prokuratorów.
- Komorników sądowych.

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

Treść orzeczenia	Wersja zanonimizowana
SSA Jan Kowalski	SSA Jan Kowalski
Protokolant: Jan Kowalski	Protokolant: Jan Kowalski
Przeciwko pozwanemu Janowi Kowalskiemu z pozwem o zapłatę.	Przeciwko pozwanemu J. K. z pozwem o zapłatę.
Powód Anna Barbara Kowalska-Nowak podjęła decyzję.	Powód A. N. podjęła decyzję.
Jan Kowalski oraz Jakub Konieczny założyli spółkę. Prezesem został Jan Kowalski.	J. K. (1) oraz J. K. (2) założyli spółkę. Prezesem został J. K. (1).
Jan oraz Jakub byli synami Zygmunta Podolskiego i nosili również nazwisko Podolski.	J. oraz J. byli synami Z. P. i nosili również nazwisko P.
(por. W. Dubis [w:] E. Gniewek, Kodeks cywilny. Komentarz, Warszawa 2006, s. 748)	(por. W. Dubis [w:] E. Gniewek, Kodeks cywilny. Komentarz, Warszawa 2006, s. 748)
świadek Clooney	świadek C.

Tabela 2: Przykładowe osoby i ich anonimizacja.

2.2. Nazwy 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.

WYŁĄCZENIE 1: Anonimizacji nie podlegają miasta opisujące siedzibę sądu, na przykład *Sąd Apelacyjny we Wrocławiu* nie podlega anonimizacji.

WYŁĄCZENIE 2: Anonimizacji nie podlegają nazwy miejscowości określające miejsce wydania nieanonimizowanej książki.

Treść orzeczenia	Wersja zanonimizowana
Sąd Apelacyjny we Wrocławiu	Sąd Apelacyjny we Wrocławiu
od wyroku Sądu Okręgowego w Opolu	od wyroku Sądu Okręgowego w Opolu
Powód zawarł we Wrocławiu umowę o kredyt długoterminowy.	Powód zawarł we W. umowę o kredyt długoterminowy.
Wydarzenie miało miejsce w Jeleniej Górze.	Wydarzenie miało miejsce w J.
Świadek widział oskarżonego w Kamiennej Górze w sobotę rano.	Świadek widział oskarżonego w K. w sobotę rano.
(por. W. Dubis [w:] E. Gniewek, Kodeks cywilny. Komentarz, Warszawa 2006, s. 748)	(por. W. Dubis [w:] E. Gniewek, Kodeks cywilny. Komentarz, Warszawa 2006, s. 748)

Tabela 3: Przykładowe miejscowości i ich anonimizacja.

2.3. Nazwy geograficzne oraz administracyjne

Przymiotniki związane z nazwami geograficznymi zmieniane są na wielokropkę „(...)”. W szczególności dotyczy to nazw:

- województw,
- powiatów,
- gmin.

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 nazwy:

- dróg ich numery (i inne oznaczenia) zamieniamy na wielokropek „(...)”.

Treść orzeczenia	Wersja zanonimizowana
mieszkał w powiecie oleśnickim przy jeziorze Wigry Wielkie	mieszkał w powiecie (...) przy jeziorze W.
przeciwko Wojewodzie Dolnośląskiemu w sprawie regulacji rzeki Odry	przeciwko Wojewodzie (...) w sprawie regulacji rzeki O.
Jechał autostradą A4, a następnie DK 8.	Jechał autostradą (...), a następnie DK (...).
Wahał się między willą nad Zatoką Gdańską a apartamentem nad Zalewem Szczecińskim	Wahał się między willą nad Zatoką (...) a apartamentem nad Zalewem (...)

Tabela 4: Przykładowe nazwy geograficzne i ich anonimizacja.

2.4. Adresy i numery lokalizacji

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

- „ul.”,
- „plac”,
- „skwer”,
- „park”,
- „rondo”.

Treść orzeczenia	Wersja zanonimizowana
na ul. Robotniczej 72 dokonano morderstwa	na ul. (...) dokonano morderstwa
Widziałem świadka przy placu Trzech Krzyży.	Widziałem świadka przy placu (...).
Zamieszkały przy ul. Wszystkich Świętych 23/11c	Zamieszkały przy ul. (...).

Tabela 5: Przykładowe adresy i ich anonimizacja.

2.5. Podmioty prowadzące działalność gospodarczą

W nazwach:

- firm,
- spółek,
- spółdzielni,
- lokali usługowych

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 podmiotów, np.: s. a., sp. z o. o.

Treść orzeczenia	Wersja zanonimizowana
Uczniowski Klub Sportowy „Miś”	(...) Klub Sportowy (...)
Eko-Recykling Spółka jawna	(...) Spółka jawna
Zakład Walcowniczy „Walex”	Zakład (...)
poszedł do baru „Wiarus”	poszedł do baru (...)

2.6. Organizacje społeczne, fundacje, stowarzyszenia, związki zawodowe oraz inne podmioty wykonujące zadania publiczne nie będące organami administracji publicznej.

Nazwy tych podmiotów są anonimizowane w sposób analogiczny jak nazwy podmiotów prowadzących działalność gospodarczą (por. pkt 2.5).

WYJĄTKI:

- **Nie** anonimizujemy nazw organizacji ustalających różnego rodzaju normy, takich jak: American National Standards Institute, Polski Komitet Normalizacyjny.
- **Nie** anonimizujemy nazw rodzajowych **instytucji publicznych** oraz **zakładów administracyjnych** wykonujących zadania publiczne, takich jak:
 - Zakład Karny,
 - Zakład Opieki Zdrowotnej,
 - Szpital,
 - Biblioteka,
 - Przy jednoczesnej anonimizacji informacji uszczegóławiających owe podmioty, typu:
 - nr,
 - nazwa,
 - adres,
 - miejsce świadczenia usług
- **Nie** anonimizujemy nazwy **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 nazwy ogólnej **instytucji naukowych**, naukowo-badawczych, badawczo-rozwojowych o zasięgu krajowym lub międzynarodowym, np. Polska Akademia Nauk, Politechnika, Uniwersytet, Akademia Medyczna, Akademia Wychowania Fizycznego, Akademia Muzyczna, Akademia Sztuk Pięknych, itp.
 - Przy jednoczesnej anonimizacji informacji uszczegóławiających, typu:
 - instytut,
 - zakład,
 - katedra,
 - wydział,
 - rewir.

2.7. Kościoły i związki wyznaniowe

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

Treść orzeczenia	Wersja zanonimizowana
Kościół Chrześcijan Baptystów	Kościół (...)
Polski Autokefaliczny Kościół Prawosławny	(...) Kościół (...)
Muzułmańska Gmina Wyznaniowa w Kruszynianach	(...) Gmina Wyznaniowa w K.
Parafia rzymskokatolicka w Bochni	Parafia (...) w B.

Tabela 7: Przykładowe anonimizacje związane z kościołami i związkami wyznaniowymi.

2.8. Numery, symbole i identyfikatory

Wszystkie identyfikatory, zarówno osób, organizacji, miejsc czy przedmiotów podczas anonimizacji powinny być zmieniane na wielokropek „(...)”. Dotyczy to w szczególności wszelkich rejestrów oraz zbiorów danych, np.:

- 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).

Treść orzeczenia	Wersja zanonimizowana
Spółdzielnia o numerze KRS 51727	Spółdzielnia o numerze KRS (...)
księgą wieczystą kw WR1K/00062615/6	księgą wieczystą kw (...)
o nr. rej. WWL 5273	o nr. rej. (...)
Posługujący się dowodem osobistym serii ADW 23565.	Posługujący się dowodem osobistym serii (...).
Liceum Ogólnokształcące nr IX	Liceum Ogólnokształcące nr (...)
Działki o nr 17/4, 17/5 i 17/6, AM-7	Działki o nr (...)
Pod numerem tel. (0-71) 4523 543	Pod numerem tel. (...)
pisywał na forum pod identyfikatorem 12432654	pisywał na forum pod identyfikatorem (...)
popularny bloger Jakub Puchatek	popularny bloger J. P.
zalogował się jako użytkownik karolinka13	zalogował się jako użytkownik k.
dokument serii ADV 455657	dokument serii (...)
umowę o kredyt obrotowy Nr 2082/56/04	umowę o kredyt obrotowy Nr (...)

Tabela 8: Przykładowe identyfikatory i ich anonimizacja.

2.9. Nazwy urzędzeń, pojazdów, nazwy różnych produktów

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

Treść orzeczenia	Wersja zanonimizowana
samochodem osobowym marki Seat Toledo	samochodem osobowym marki S. (...)
zakup samochodu marki Peugeot	zakup samochodu marki P.
drukarka HP 4100 oraz telefon Nokia 5610	drukarka H. (...) oraz telefon N. (...)
laptop Lenovo R61	laptop L. (...)
system operacyjny Windows Vista	system operacyjny W. (...)

Tabela 9: Przykładowe nazwy produktów i ich anonimizacja.

2.10. Informacje chronione ustawowo

Anonimizacji podlegają wszelkie informacje objęte prawną ochroną wyrażoną w konkretnej ustawie.

2.11. Adresy e-mail oraz adresy URL strony internetowej

Treść orzeczenia	Wersja zanonimizowana
sonia1974@mail.com	(...)
Mail wysłano z adresu kayak_klcofas@op.pl	Mail wysłano z adresu (...)
www.portal.prawo.pl	(...)
Plik został pobrany ze strony www.plixx.pl	Plik został pobrany ze strony (...)

Tabela 10: Przykładowe adresy e-mail oraz adresy URL

2.12. Pozostałe słowa i frazy wymagające anonimizacji

2.12.1. Cyfry

2.12.1.1. Długie ciągi cyfr, w sytuacji, gdy nie jest oczywista ich dokładniejsza klasyfikacja, na przykład 346565356312677.

2.12.1.2. Ciągi liczb, liter, cyfr, myślników, gdy nie jest oczywista ich dokładniejsza klasyfikacja, np. 4565-DF0-34cvf44.

2.12.2. Treść haseł reklamowych i innych napisów

2.12.3. Nazwy

2.12.3.1. Zespołów muzycznych.

2.12.3.2. Drużyn sportowych.

2.12.3.3. Innych organizacji społecznych lub o podobnym charakterze.

2.12.4. Słownictwo:

2.12.4.1. Wulgarne.

2.12.4.2. Potocznie uważane za obelżywe.

Treść orzeczenia	Wersja zanonimizowana
Wydział Komunikacji Urzędu Rejonowego w Brzegu	Wydział Komunikacji Urzędu Rejonowego w B.
Prezydent Wrocławia	Prezydent W.
Wojewoda Dolnośląski	Wojewoda (...)
346565356312677	(...)
4565-DF0-34cvf44	(...)
baner reklamowy z dwustronnym napisem "najlepsze pierogi w mieście"	baner reklamowy z dwustronnym napisem (...)
zagrał zespół szantowy "Przyjaciele Grubego Bosmana"	zagrał zespół szantowy (...)

Tabela 11: Inne słowa i frazy oraz ich anonimizacja.

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 orzeczenia 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 prawna, prawnicza oraz odnosząca się do czynności kancelaryjnych administracji wymiaru sprawiedliwości.

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.

Typ frazy	Przykładowe frazy
Treść aktów normatywnych oraz odniesienia do różnych aktów prawnych	- treść przepisów (norm) prawnych art. 187 § 1 pkt 2 k.p.c. - treść aktów prawa miejscowego Dz.U.2012.393 (III CZP 141/93, OSNCP 1994, nr 5, poz. 102) OSA 2001/10/53, OSNC 1978/11/194, OSNPG 10-11/1990, OSNPiPr. 10/2002, poz. 2
Adresy dzienników promulgacyjnych, nazwy periodyków naukowych	LEX nr 22050 sygn. FSK 377/04, Legalis
Numerory orzeczeń w dowolnych systemach informacji prawnej	Według Polskiej Normy PN-92/c-96051 odpowiadającego Polskiej Normie PN-D-94021
Oznaczanie norm technicznych	

3.2. Oznaczenia czasu oraz odniesienia do określonych podmiotów bądź przedmiotów w czasie.

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

WYJĄTEK: Informacje o dacie urodzenia konkretnej osoby fizycznej.

Treść orzeczenia	Wersja zanonimizowana
W latach 2005-2008 pracował jako motorniczy	W latach 2005-2008 pracował jako motorniczy
Od 15-go stycznia 1999 roku	Od 15-go stycznia 1999 roku
W niedzielę sklep był otwarty do godziny 16.30	W niedzielę sklep był otwarty do godziny 16.30
w drugi dzień świąt Bożego Narodzenia	w drugi dzień świąt Bożego Narodzenia
Jan Kowalski, ur. 2.01.1976r.	J. K., ur. (...)

3.3. Organy władzy publicznej, zwłaszcza organy administracji publicznej (w tym nazwy urzędów) oraz wymiaru sprawiedliwości.

- **Nie** anonimizujemy nazw organów lub urzędów, z zachowaniem następujących wyłączeń
 - Anonimizujemy informację o miejscu, w którym znajduje się ten urząd lub organ – właściwość miejscową podmiotu.
 - Anonimizacji podlega również nazwa wewnętrznych jednostek organizacyjnych (tj. wydziałów, komórek) organów i urzędów, np. Wydział Komunikacji Urzędu Miejskiego w Brzegu.
- **Nie** anonimizujemy nazw organizacji międzynarodowych, których członkami są państwa (na przykład Światowa Organizacja Zdrowia, Międzynarodowy Fundusz Walutowy, UNESCO).
- **Nie** anonimizujemy nazw państw
- **Nie** anonimizujemy nazw sądów (w sensie rodzajowym czy przedmiotowym - instancyjnym), czyli nazw: sądy powszechne, sądy administracyjne, Sąd Rejonowy, Naczelny Sąd Administracyjny.
- **Nie** anonimizujemy nazwy – Skarb Państwa.

3.4. Pozostałe nieanonimizowane treści orzeczeń

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
Liczby porządkowe, liczby pewnych zdarzeń, osób, przedmiotów	pozwany nie zapłacił 3. raty, jak również raty nr 4 zadął 23 ciosy nożem miała 2 synów i 3 córki w firmie było zatrudnionych 1345 pracowników jego spółka posiadała 32 samochody dostawcze
Liczby z jednostkami	przebieg pojazdu w wysokości 325.328 km zakupionej kukurydzy w ilości 5.000 ton powierzchnię użytkową 80 m. kw.
Kwota pieniężna	zasądzenia kwoty 110386zł 10gr z ustawowymi odsetkami w kwocie 46.179.60 zł. czyli 30,07 zł brutto za stronę słownie: dwieście pięćdziesiąt złotych i piętnaście groszy

Choroby oraz rodzaje terapii

Odmówiła przyjscia do pracy ze wzgledu na niezyt górnych dróg oddechowych

Zmuszony byl wykonac otolaryngologiczny zabieg otwarcia przedniej sciany tchawicy

Nazwy związkowe chemicznych

Wodorotlenek sodu, nadmanganian potasu

Wykonywany (lub wyuczony) zawód, stanowisko służbowe oraz pełnione funkcje

w firmie pełniła obowiązki Głównego Księgowego
najpierw pracował jako weterynarz, a następnie jako informatyk

Tabela 13: 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:

- Nie zawiera ono danych objętych klauzulą tajności
- 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).
- 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 Adwokacka Janiny Nowak	Kancelaria (...) Kancelaria Adwokacka (...) Kancelaria Adwokacka J. N.
Zakłady Naprawcze Taboru Kolejowego w Oleśnicy S.A.	(...) S.A. Zakłady (...) S.A. Zakłady (...) w O. S.A.

Tabela 14: 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 tabela:

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.