

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 konkretnie 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 urzędzeń, 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.

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 (...).

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. Osoby fizyczne

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 następujące podmioty:

- Sędziowie składu orzekającego.
- Protokolanci.
- Prokuratorzy.
- Komornicy sądowi.

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

Treść orzeczenia	Wersja zanonimizowana
SSA Jan Kowalski	SSA Jan Kowalski
Protokolant: Jan Kowalski	Protokolant: Jan Kowalski
Przeciwko pozwanemu <i>Janowi Kowalskiemu</i> z pozwem o zapłatę.	Przeciwko pozwanemu <i>J. K.</i> z pozwem o zapłatę.
Powód <i>Anna Barbara Kowalska-Nowak</i> podjęła decyzję.	Powód <i>A. N.</i> podjęła decyzję.
<i>Jan Kowalski</i> oraz <i>Jakub Konieczny</i> założyli spółkę. Prezesem został <i>Jan Kowalski</i> .	<i>J. K. (1)</i> oraz <i>J. K. (2)</i> założyli spółkę. Prezesem został <i>J. K. (1)</i> .
<i>Jan</i> oraz <i>Jakub</i> byli synami <i>Zygmunta Podolskiego</i> i nosili również nazwisko <i>Podolscy</i> .	<i>J.</i> oraz <i>J.</i> byli synami <i>Z. P.</i> i nosili również nazwisko <i>P.</i>
(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 <i>Clooney</i>	świadek <i>C.</i>

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 *Sąd Apelacyjny we Wrocławiu* nie podlega anonimizacji.

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

Treść orzeczenia	Wersja zanonimizowana
Sąd Apelacyjny we <i>Wrocławiu</i>	Sąd Apelacyjny we <i>Wrocławiu</i>
od wyroku Sądu Okręgowego w <i>Opolu</i>	od wyroku Sądu Okręgowego w <i>Opolu</i>
Powód zawarł we <i>Wrocławiu</i> umowę o kredyt długoterminowy.	Powód zawarł we <i>W.</i> umowę o kredyt długoterminowy.
Wydarzenie miało miejsce w <i>Jeleniej Górze</i> .	Wydarzenie miało miejsce w <i>J.</i>
Świadek widział oskarżonego w <i>Kamiennej Górze</i> w sobotę rano.	Świadek widział oskarżonego w <i>K.</i> 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)

2.3. Nazwy geograficzne

Przymiotniki związane z nazwami geograficznymi zmieniają się na wielokropek „(...)”. W szczególności dotyczy to województw, powiatów czy gmin. Rzeczownikowe nazwy geograficzne, pisane wielką literą, zamieniają się, 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 „(...)”.

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 (...)

2.4. Adresy

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

- „ul.”,
- „plac”,
- „skwer”.

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. (...).

2.5. 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 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 Waleowniczy "Walex"	Zakład (...)
poszedł do baru "Wiarus"	poszedł do baru (...)
Spółdzielnia Mieszkaniowa "Akademicka"	Spółdzielnia Mieszkaniowa (...)

2.6. Organizacje społeczne, fundacje, stowarzyszenia oraz instytucje publiczne lub inne podmioty wykonujące zadania publiczne

Nazwy tych podmiotów są anonimizowane w sposób analogiczny jak nazwy firm, spółek, spółdzielni (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 szkół wyższych, takich jak: Politechnika, Uniwersytet, Akademia Medyczna, Akademia Wychowania Fizycznego, Akademia Muzyczna, Akademia Sztuk Pięknych.
- 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 gospodarecza, Izba rolnicza, Związek Rzemiosła Polskiego).

- Nie anonimizujemy nazwy Polskiej Akademii Nauk i nazw analogicznych instytucji z innych krajów.

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 Kruszyńnianach	(...) Gmina Wyznaniowa w K.
Parafia rzymskokatolicka w Bochni	Parafia (...) w B.

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).

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 (...)

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. (...)

2.10. 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-DFo-34cvf44.
- Treść hasel reklamowych i innych napisów.
- Nazwy zespołów muzycznych, drużyn sportowych.

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-DFo-34cvf44	(...)
baner reklamowy z dwustronnym napisem "najlepsze pierogi w mieście"	baner reklamowy z dwustronnym napisem (...)
zagrał zespół szantowy "Przyjaciele Grubego Bosmana"	zagrał zespół szantowy (...)

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 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.

Typ frazy	Przykładowe frazy
Sygnatury spraw, numer kart akt sprawy, numery dowodów rzeczowych	<p>Sygnatura akt I ACa 874/03</p> <p>IV GNC 3991/08</p> <p>k.222</p> <p>czeków z dowodów rzeczowych pod numerem Drz 2/1518/02</p>
Adresy dzienników promulgacyjnych, nazwy periodyków naukowych, odniesienia do różnych aktów prawnych	<p>Dz.U.2012.393</p> <p>(III CZP 141/93, OSNCP 1994, nr 5, poz. 102)</p> <p>OSA 2001/10/53, OSNC 1978/11/194, OSNPG 10-11/1990, OSNPr. 10/2002, poz. 2</p> <p>art. 187 § 1 pkt 2 k.p.c.</p>
numery orzeczeń w systemach informacji prawnej LEX oraz LEGALIS	<p>LEX nr 22050</p> <p>sygn. FSK 377/04, Legalis</p>
Oznaczenie norm technicznych	<p>Według Polskiej Normy PN-92/c-96051</p> <p>odpowiadającego Polskiej Normie PN-D-94021</p> <p>wnioski o dofinansowanie projektów do konkursu otwartego nr 5/POKL/9.4/2012</p>
Nazwy programów operacyjnych, działań, poddziałań, priorytetów, konkursów	<p>Program Operacyjny Innowacyjna Gospodarka 2007-2013</p>

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 fizycznej.

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

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

- 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 sądów, takich jak 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

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 przyjscia do pracy ze wzgledu na niezbyt gornych drog oddechowych Zmuszony byl wykonac otolaryngologiczny zabieg otwarcia przedniej sciany tchawicy
Kwota pieniezna	zasadzenia kwoty 110386zł 10gr z ustawowymi odsetkami w kwocie 46.179.60 zł. czyli 30,07zł brutto za strone sloownie: dwieście pięćdziesiąt złotych i piętnaście groszy
Nazwy państw	Przebywal na terenie Niemiec do maja mieszkania wlasnoscowego we Francji o powierzchni wyeliminowania z rynku w Polsce jedyne wówczas
Liczby z jednostkami	przebieg pojazdu w wysokosci 325.328 km zakupionej kukurydzy w ilosci 5.000 ton powierzchnie uzytkowa 80 m. kw.

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:
 1. Nie zawiera ono danych objętych klauzulą tajności.
 2. 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).

3. 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 zgłaszającym 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. (...) S.A. Zakłady (...) S.A.
Zakłady Naprawcze Taboru Kolejowego w Oleśnicy S.A.	Zakłady (...) w O. S.A.

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	Anonimizacja	Kwalifikacja 2	Anonimizacja
	1	1		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.

