

Krzysztof Olszowski*, Dariusz Rosłon**

AUTOMATYCZNE WSPOMAGANIE ANONIMIZACJI ORZECZEŃ SĄDOWYCH

Publikacja wszystkich orzeczeń sądowych jest warunkiem realizacji praw i wolności konstytucyjnych, takich jak prawo do sądu, [...] może przyczynić się do ujednoczenia orzecznictwa, [...] treść i problematyka poruszana w orzeczeniach może stanowić ważny sygnał dla legislatury co do problemów związanych ze stosowaniem prawa¹.

1. Wstęp

Zwiększająca się w ostatnich latach świadomość prawa w społeczeństwie powoduje nasilone zainteresowanie informacją prawną. Przepisy o dostępie do informacji publicznej, otwierające szeroko dostęp do informacji jeszcze niedawno zastrzeżonej dla wąskich grup profesjonalistów, stają się obiektem zainteresowania szerokiej publiczności.

Znacznym zainteresowaniem cieszy się konkretne orzecznictwo sądów. Czytelników nie interesują już gazetowe omówienia, czy komentarze w literaturze fachowej. Chcą mieć precyzyjny tekst wyroku zaraz po jego ogłosze-

* Krzysztof Olszowski – pracuje w firmie SoftProdukt, specjalizuje się w systemach dla administracji publicznej.

** Dariusz Rosłon – pracuje w firmie SoftProdukt, specjalizuje się w systemach dla administracji publicznej.

¹ M. Jagielski, M. Niedużak, *Publiczna dostępność orzeczeń sądowych*, Warszawa 2010, s. 5.

niu. Omówienie gazetowe ciekawego wyroku to tylko motywacja do poszukiwania pełnego tekstu sentencji i uzasadnienia. Powoduje to presję na „wydobywanie” wyroków wszelkimi sposobami.

Jak pokazują liczne dyskusje na temat publicznego dostępu, poszukujący nie zadowolają się dostępem do wybranych, najważniejszych orzeczeń. Argumentują, że kryterium ważności nie jest uniwersalne dla różnych grup odbiorców – sędziów, pełnomocników procesowych, obywateli, przedsiębiorców. Dla każdego ważny może być inny aspekt orzeczenia.

2. Publiczny dostęp jako zagrożenie bezpieczeństwa danych niejawnych

Ponieważ środki techniczne znakomicie poprawiają szansę szybkiego udostępniania informacji orzeczniczej, pojawia się pytanie kilka lat temu niesłychane: dlaczego nie udostępnić wszystkich jawnych orzeczeń? Nowe „zwirtualizowane” pokolenie wchodzące w aktywne życie nie przyjmuje argumentu o niezbędnych ograniczeniach z przyczyn prawnych lub technicznych. Żądanie wskazania ważnego interesu prawnego zaczyna być traktowane jako niepoważny wykręt urzędniczy. Dobitną ilustracją zmieniających się postaw opinii publicznej a także przedstawicieli władzy są wydarzenia wokół podpisania umowy międzynarodowej ACTA (ang. *Anti-Counterfeiting Trade Agreement*). Furię „ulicy” i środowisk opiniotwórczych pospołu wywołała nie tylko treść umowy, ale także dyskrecjonalny tryb jej przygotowania. W efekcie rząd ujawnił wewnętrzne dokumenty negocjacyjne, bez pytania tejsze „ulicy” o ważny interes prawny.

Bywa gorzej – zainteresowani informacją prawną nikogo nie pytają o zgodę na użycie, ani nie żądają udostępnienia. Po prostu nielegalnie wynoszą z sądów hurtowe ilości wyroków na nośnikach elektronicznych (pendrivach). Wystarczy zmyślny student – praktykant i nieuważny pracownik sekretariatu procesowego. Techniczne cuda z dawnych filmów szpiegowskich to dziś pospolite wyposażenie aplikanta w sądowej czytelnicy akt. Co nie jest dostępne na plikach komputerowych, można błyskawicznie sfotografować. Oczywiście – z całym bogactwem danych osobowych. Potwierdzają to liczne doświadczenia nauczycieli akademickich, którzy natykają się na ożywiony handel wymienny pomiędzy studentami prawa. Pochodzenie takich „tekstów źródłowych” jest często spowite gęstą mgłą.

Ogłoszony w grudniu 2010 r. raport „Publiczna dostępność orzeczeń sądowych” przygotowany przez Helsińską Fundację Praw Człowieka, Forum Obywatelskiego Rozwoju i Akademię Leona Koźmińskiego w ramach projektu

„Lepszy Wymiar Sprawiedliwości” zawiera analizę stanu dostępności orzecznictwa i liczne rekomendacje.

Autorzy raportu uważają, że polski system sądownictwa powinien zmierzać w kierunku pełnej dostępności treści całego orzecznictwa dla obywateli za pośrednictwem internetu. Warunkiem udostępniania orzeczeń obywatelom jest jednak ich uprzednia anonimizacja, czyli utajnienie danych osobowych, uniemożliwiająca identyfikację konkretnych osób. Aby usprawnić proces udostępniania autorzy raportu postulują odpowiednie przeszkolenie personelu sekretarskiego w tym zakresie.

Zatem skoro nie da się powstrzymać wypływu informacji, lepiej ujawnić ją w sposób kontrolowany i bezpieczny. Ten pogląd od niedawna zaczął sobie torować drogę w resorcie sprawiedliwości.

Informacja bezpieczna, a jednocześnie pożyteczna merytorycznie, to informacja poddana odpowiedniej anonimizacji, czyli usunięciu danych pozwalających na identyfikację osób – stron postępowania sądowego. Dotyczy to zarówno nazwisk osób, firm, adresów jak i kwot pieniężnych.

Ponieważ anonimizować trzeba wszystkie orzeczenia w krótkich terminach, nakłada to zwiększone obowiązki na personel pomocniczy w sądach. W świetle zapowiadanego przez resort sprawiedliwości zmniejszenia obsady etatowej na stanowiskach pomocniczych, należy oczekiwać, że trzeba będzie wykonać więcej pracy mniejszym personelem. Spowoduje to nieunikniony wzrost presji na tempo prac sekretarskich, takich jak anonimizacja. W skrócie ma być szybko, a zarazem precyzyjnie. Te wymogi na ogół trudno pogodzić.

Problem sygnalizują badania omówione w raporcie „Publiczna dostępność orzeczeń sądowych”. Czytamy w nim m.in., że *„zdarzają się ciągle przypadki udostępniania orzeczeń na podstawie D.i.p. bez dokonania anonimizacji. Duża część orzeczeń jest również anonimizowana nieskutecznie. Na przykład zdarza się, że pomimo powzięcia pewnych działań mających na celu usunięcie danych osobowych, ze względu na zastosowaną technikę możliwe jest ich odczytanie. Niektóre z uzyskanych orzeczeń były anonimizowane w sposób umożliwiający odwrócenie procesu i usunięcia przeszkód w poznaniu dokładnych danych osobowych”*².

3. Dwa podejścia do zagadnienia anonimizacji

Warto przyjrzeć się poglądom na zagadnienie anonimizacji prezentowanym przez miarodajne ośrodki informatyzacji sądownictwa.

² Ibidem, s. 25.

Dwie tezy wyrażają przeciwstawne podejście do tej samej kwestii. Pierwszą prezentują innowatorzy ze środowiska CBKE Uniwersytetu Wrocławskiego, zasłużonego dla badań nad informatyzacją sądownictwa powszechnego, którzy stwierdzają, że *„anonimizacja będzie w pełni automatyczna. Musimy rozwiązać tylko pewne niuanse, żeby system mógł znaleźć wszystkie dane podlegające jej obowiązkowi”*³.

Zaskakuje inżynierski nieledwie optymizm, z jakim traktuje się „niuansę”. Być może jest to optymizm początkowego okresu badawczego, kiedy tezę traktuje się podświadomie jako „prawie dowiedzioną”.

Druga, ostrożniejsza teza sformułowana przez Wydział Informatyki Sądowej Naczelnego Sądu Administracyjnego brzmi: *„Proces pozbawiania orzeczenia danych mogących doprowadzić do łatwej identyfikacji przez osoby postronne stron postępowania sądownoadministracyjnego [...] nie musi mieć charakteru jednolitego. Każdorazowo poprzedzać go powinna analiza, czy w konkretnej sprawie występują dane podlegające prawnej ochronie”*⁴.

Jest to stanowisko praktyków o największym w kraju doświadczeniu w masowym udostępnianiu orzeczeń (ok. 100.000 rocznie). Wieloletnia już praktyka daje podstawę do realistycznego podejścia do potrzeb właściwej anonimizacji. Chodzi przecież o takie opracowanie sentencji, a szczególnie uzasadnienia wyroku, żeby tekst niósł zawartość informacyjną.

Pogląd ten popierają badania autorów wspomnianego raportu, którzy stwierdzają, że *„Warunkiem udostępniania orzeczeń obywatelom jest ich uprzednia anonimizacja. Wskazane jest odpowiednie przeszkolenie personelu sekretarskiego w tym zakresie. [...] Osoba pracująca nad przygotowaniem orzeczenia z reguły orientuje się w danych, które należy usunąć, aby uniemożliwić identyfikację konkretnych osób. Co więcej, obsługa sekretarska współpracująca z danym sędzią czy danym składem sędziowskim ma z nim ułatwiony kontakt, co upraszcza udział samego autora (autorów) danego orzeczenia w procesie anonimizacji oraz pozwala na wyjaśnienie ewentualnych wątpliwości, jakie mogą się w danej sytuacji nasunąć. To właśnie sędzia lub sędziowie, którzy podjęli dane rozstrzygnięcie, są najlepiej przygotowani do odpowiedzi na pytanie: które konkretnie informacje mogą pozwolić na identyfikację osoby, a w efekcie winny zostać usunięte z treści orzeczenia podlegającej publikacji lub udostępnieniu”*⁵.

³ M. Jałoszewski, *Orzeczenia w Internecie. Korzyści dla sądów i obywateli*, „»Na wokandzie« Kwartalnik Informatyczny Ministerstwa Sprawiedliwości” 2011, nr 5 (8), s. 38.

⁴ Naczelny Sąd Administracyjny, *Informacja o działalności sądów administracyjnych w 2008 roku*, red. Kancelaria Prezesa NSA, Warszawa 2009, s. 382.

⁵ M. Jagielski, M. Niedużak, op.cit., s. 40.

Z punktu widzenia nauk informatycznych należy podkreślić, że aktualny stan wiedzy w zakresie analizowania języka naturalnego nie daje podstaw do nadmiernego optymizmu co do skuteczności automatycznych systemów. Postępy nauki w ciągu ostatnich 30 lat są co prawda znaczne, ale ciągle niezadowalające. Najsprawniejsze systemy analizy leksykalnej nie są w stanie całkowicie zastąpić żywego tłumacza nawet dla tekstów technicznych. Tym bardziej nie należy oczekiwać skutecznego odgadnięcia intencji sędziego piszącego uzasadnienie, będące zawiłym tekstem prawniczym, zanurzonym w realiach rozstrzyganej sprawy. Fachowa praca redakcyjna przy anonimizacji jest domeną człowieka, choć można ją wspomagać.

Grupa specjalistów współpracujących z NSA przeprowadziła eksperymenty mające na celu naświetlenie problemu od strony praktycznej. Próbną eksploatacją programu wspomaganie anonimizacji GALL w wybranych sądach administracyjnych potwierdza, że obecnie nie jest możliwa w pełni automatyczna anonimizacja orzeczeń.

Nawet dysponując wzorcami tekstów do anonimizacji (np. nazwisko skarżącego) nie zawsze jest możliwe rozpoznanie wszystkich wystąpień wzorca w tekście. Problem stwarza fleksja języka polskiego, „literówki”, a także wystąpienia słowa w innym znaczeniu (np. nazwiska i nazwy miast będące rzeczownikami pospolitymi, jak „belka”, „brama”, „łódź”, „wola”).

Pełna automatyzacja stwarza też ryzyko przekraczania granic anonimizacji, czyli nieuzasadnionego pozbawienia publikowanego dokumentu istotnych informacji.

4. Dylemat – ręcznie czy automatycznie?

Czy zatem jesteśmy skazani tylko na „ręczną” anonimizację orzeczeń z wykorzystaniem edytorów tekstu? Może lepiej poczekać z automatyzacją na dalsze postępy informatyki?

Pochylmy się nad skalą problemu. W samych tylko sądach administracyjnych wydaje się rocznie ok. 110 tysięcy orzeczeń. Według pomiarów eksperymentalnych czas potrzebny na „ręczną” anonimizację orzeczenia przy użyciu edytora MSWord wynosi przeciętnie 20 minut.

W skali roku daje to zatem około 30 tysięcy godzin przeznaczanych na anonimizację. Innymi słowy, personel sekretarski sądów administracyjnych musi przeznaczyć 5 tysięcy osobodni na to odpowiedzialne zadanie. Nie uwzględniamy tu czasu na ewentualne wyjaśnianie wątpliwości z sędziami – autorami uzasadnień.

W przypadku sądów powszechnych mamy do czynienia z prawdziwym oceanem tekstu: około 13,3 miliona orzeczeń. Daje to w efekcie potencjalne obciążenie pracami anonimizacyjnymi w wysokości 600 tysięcy osobodni.

Jak widać, jest o co walczyć. Zadanie jest potężne i co gorsza – nieuniknione.

Przyjmijmy zatem założenie umiarkowanie pesymistyczne – nie da się zanonimizować wszystkiego automatycznie, ale można próbować podpowiadać i porządkować proces redakcyjny. Decyzje redakcyjne zostawmy człowiekowi.

5. GALL – program wspomagający anonimizację orzeczeń.

Specjalny dedykowany program GALL opracowano z uwzględnieniem zasad anonimizacji określonych w Zarządzeniu Prezesa NSA z 12.7.2007 r.

GALL był doskonalony w konsultacji z pracownikami NSA wykonującymi anonimizację orzeczeń na co dzień. Przeprowadzona analiza, statystyczna i empiryczna, treści orzeczeń pozwoliła wprowadzić dedykowane mechanizmy tekstowe wykorzystujące dane z rejestrów sądowych. Udało się dzięki temu znacznie przyspieszyć proces anonimizacji.

Celem ograniczenia błędów operatorskich zastosowano funkcje prezentacji zamienianych fragmentów tekstu („co” na „co”) oraz możliwości cofania i powtarzania wybranych anonimizacji. Poprawia to jakość anonimizacji i zwiększa bezpieczeństwo publikacji.

Na podstawie serii próbnych eksploatacji programu GALL w wybranych sądach administracyjnych stwierdzono zachęcające efekty ilościowe i jakościowe wykorzystania programu.

Po pierwsze, uzyskano znaczne skrócenie czasu anonimizacji orzeczeń. Typowe orzeczenia anonimizuje się o połowę krócej.

Jeszcze lepsze wyniki uzyskano dla orzeczeń w sprawach z wieloma skarżącymi. Tutaj żmudny proces ręcznego przygotowywania listy zamian nazwisk na inicjały dokonuje się błyskawicznie.

Potencjalne oszczędności wynikające z zastosowania programu GALL wynoszą więc 2.500 osobodni rocznie w skali sądów administracyjnych i 300.000 osobodni rocznie w skali sądów powszechnych

Po drugie, istotnie zwiększyło się bezpieczeństwo i poprawność anonimizacji.

Program wychwytywa teksty ukryte w trakcie edycji w MSWord, dzięki czemu do publikacji nie mogą się przypadkowo przedostać „resztki” redakcyjne. Wyraźne uwidocznienie zamian poprzez kolorowe podświetlenie po-

maga skupić uwagę na ważnych fragmentach tekstu. Jest to szczególnie ważne, kiedy znużenie może sprzyjać błędom edytorskim. Podobnie jest z wieloma skarżącymi – unika się znacznego ryzyka pomyłek.

6. GALL – wstępna anonimizacja automatyczna

Na podstawie danych konkretnej sprawy, pobranych z eksploatowanego w sądach administracyjnych systemu Obsługi Spraw Orzekanych (OSO) program GALL dokonuje wstępnej, automatycznej anonimizacji.

Dokonywane jest najpierw ukrycie strony skarżącej

- zamiana nazwy osoby prawnej na inicjały lub „podmiot A”
- zamiana nazwisk na inicjały

Oryginalna nazwa	Typ anonimizacji	Proponowana nazwa
Piękny Ogród Sp. z o.o.	Osoba prawna	P.O.
Dyrektor Izby Skarbowej w Opolu	Organ	Dyrektor Izby Skarbowej
Jan Nowak	Osoba fizyczna	J. N.
Paweł Kowalski	Osoba fizyczna	P. K.

Dopuszczono tu możliwość poprawienia wzorca do zamiany:

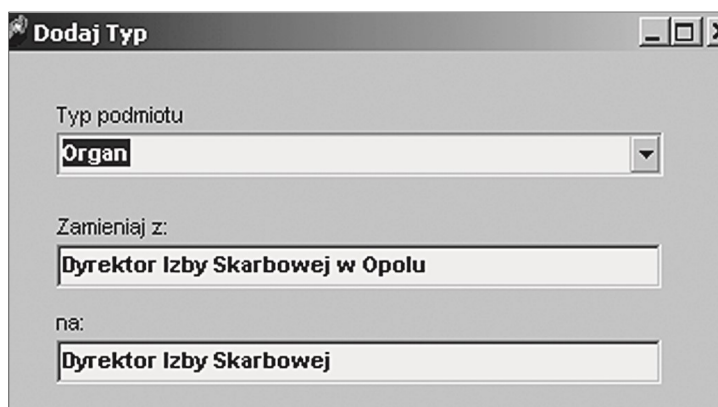


The screenshot shows a window titled "Dodaj Typ" with a key icon in the title bar. Inside the window, there are three input fields:

- The first field is labeled "Typ podmiotu" and has a dropdown menu with "Osoba fizyczna" selected.
- The second field is labeled "Zamieniaj z:" and contains the text "Paweł Kowalski".
- The third field is labeled "na:" and contains the text "P. K."

W polu „na:” można wpisać dowolny, bardziej odpowiedni wzorec.

Ukrywa się organ administracji, zgodnie z ustalonym katalogiem podmiotów zawartym w systemie OSO. Organy ukrywane są tak, aby nie zatracić sensu merytorycznego orzeczenia.



Następnie dokonuje się najbardziej żmudnych zamian, a w tym:

- numeru skarżonej decyzji (na podstawie danych z OSO)
nr [...] [PS-I/47071-22/06/TM]
- daty skarżonej decyzji (numer dnia jest zamieniany na „[...]”)
z dnia [...] [29] stycznia 2007
- danych podlegających anonimizacji „na ogólnych zasadach”, jak nazwy miast – zamiana na pierwsze litery
w N. [Niemirowie]
- nazwy ulic wraz numerami domów i lokali – zamiana na pierwsze litery
przy ul. D. [Długiej 25 m. 13]
- kwoty – zamiana na [...] [125 tys.] PLN
- osoby fizyczne – zamiana na inicjały (nie dotyczy składu orzekającego)
J.N. [Jan Nowak]
- inne numery – zamiana na [...] (nie dotyczy Dziennika Ustaw, Monitora Polskiego, Lexa)
Dowód osobisty numer [...] [AAT 967835]

Wszystkie zamiany podświetlone są na ekranie kolorami – skrót na zielono, tekst oryginalny na żółto.

7. GALL – edycja specjalna

Wstępnie zanonimizowany dokument otwiera się w specjalizowanym edytorze tekstu. Użytkownik weryfikuje, uzupełnia, poprawia i akceptuje anonimizację dokumentu.

Edytor korzysta z typowych wzorców do podpowiadania anonimizacji zaznaczonego tekstu i wyposażony jest w podstawowe funkcje:

- Powtórz anonimizację według wzorca
- Zmień daną anonimizację
- Cofnij daną anonimizację
- Cofnij wszystkie anonimizacje dla danego wzorca.

Warto zauważyć, że zamieniane teksty traktowane są jako integralne obiekty, co ułatwia dopracowanie anonimizacji zgodnie z intencjami autora oryginalnego tekstu.

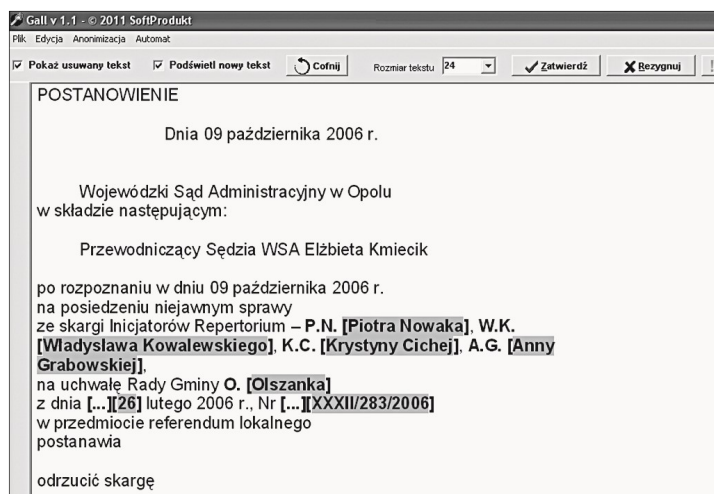
Aby uniknąć przypadkowych błędów przy anonimizacji przyjęto następujące założenia:

- Ograniczamy możliwość pisania i kasowania tekstu – w zamian dajemy funkcje zamiany fragmentów tekstu na inny.
- Zarówno tekst nowy (skrót) jak i oryginalny są wyraźnie zaznaczone w dokumencie ekranowym (na zielono i żółto).

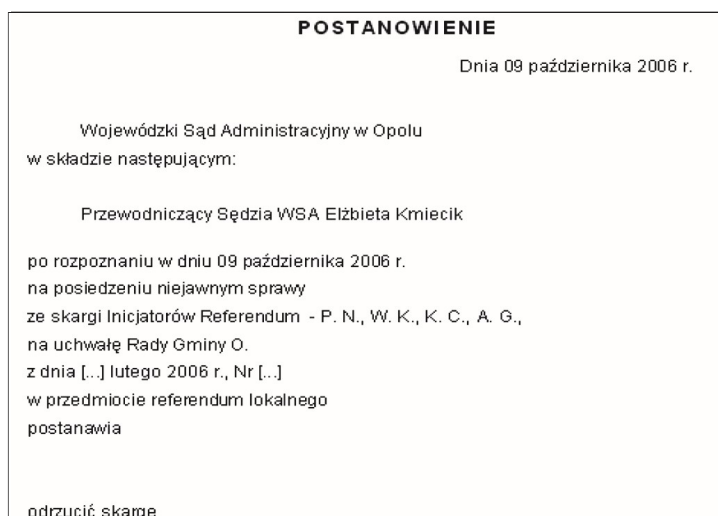
P.N. [Piotr Nowak]

- Każda anonimizacja (zamiana tekstu) może być wycofana lub poprawiona.

Korekta może dotyczyć zakresu anonimizowanego tekstu („odkąd-dokąd”) i postaci skrótu (na „co”). Cały tekst ma uwidocznione anonimizacje:



Poniżej – tekst orzeczenia po anonimizacji gotowy do wydruku lub publikacji w internecie:



Źródło: Ekran program GALL – © SoftProdukt

Automatic Support for Anonymization of Court Rulings

Public access to court rulings is crucial to the right of access to court and can help improve the quality of case law.

Public opinion demands up-to-date and comprehensive legal information. It is not enough to publish the “most important” verdicts. Each one may be most important. It seems inevitable to allow access to all rulings. The rulings must be anonymized in order to protect private data.

Manual stripping of sensitive information is a cumbersome and risky process. Automated assistance is necessary. A semi intelligent system GALL, developed for Administrative Courts, provides an effective and safe support. It analyses ruling texts searching for sensitive data e.g. names, addresses, amounts and suggests the safe abbreviations. The user can accept or edit the anonymized text. The assisted process is twice faster and significantly safer than manual anonymization.