

ANDRZEJ ŁACHWA

PODZIAŁ AKTÓW NORMATYWNYCH NA DOKUMENTY*

I. WSTĘP

Przy projektowaniu pełnotekstowych¹ systemów pośredniego² wyszukiwania informacji, których bazy danych zawierać mają pierwotne teksty prawne, tj. teksty aktów normatywnych, powstaje problem podziału owych tekstów na tzw. dokumenty³. Od sposobu przeprowadzenia takiego podziału zależy w dużej mierze sprawność działania odpowiedniego systemu.

Systemy wyszukiwawcze mają za zadanie zaspokajanie indywidualnych potrzeb informacyjnych swoich użytkowników, wyrażonych w formułowanych przez nich kwerendach⁴. Zaspokajanie owych potrzeb następuje przez odnalezienie w bazie danych i wydanie użytkownikowi odpowiedniej liczby dokumentów zawierających informacje mające pewną doniosłość dla rozstrzygnięcia pytania interesującego danego użytkownika (w przypadku systemów przechowujących teksty prawne, chodzi tu o pytania prawne).

Efektywność pracy systemów, o których mowa, uzależniona jest w du-

* Niniejsze opracowanie powstało w związku z projektem badawczym ANAPHORA realizowanym od 1979 r. w Zakładzie Humanistycznych Zastosowań Informatyki Uniwersytetu Jagiellońskiego. Autor dziękuje członkom zespołu opracowującego ten projekt: prof. dr. F. Studnickiemu, mgr B. Polanowskiej, mgr E. Stabrawie i mgr. J. M. Fallowi za ich uwagi w dyskusjach nad algorytmem.

¹ Przez systemy pełnotekstowe rozumiemy systemy, w których kwerenda użytkownika (vide przypis 4) odwołuje się do właściwości formalnych tekstów oryginalnych, a nie do właściwości formalnych opisów owych tekstów wykonanych przez organizatorów danego systemu.

² Przez systemy pośredniego wyszukiwania informacji rozumiemy tu systemy, które nie rozstrzygają pytań interesujących ich użytkowników, a jedynie dostarczają im takich fragmentów odpowiednich tekstów, z których odpowiedź na owe pytania może być wyprowadzona w drodze odpowiednich wnioskowań.

³ Przez dokumenty rozumiemy tu fragmenty tekstów uzyskane w drodze podziału dokonywanego przy wyłącznym odwołaniu się do kryteriów formalnych.

⁴ Kwerenda jest to komunikat skierowany do systemu wyszukiwawczego przez jego użytkownika, określający wymagania, które spełniać winny poszukiwane przez niego dokumenty.

zym stopniu od jakości tzw. urządzeń semantyzujących⁵ zastępujących systemowi — w pewnej przynajmniej mierze — niedostającą mu kompetencję semantyczną, a także, jak już wspomnieliśmy, od sposobu podziału bazy danych na dokumenty.

Poniżej przedstawiony zostanie prosty algorytm podziału pierwotnych tekstów prawnych na dokumenty. Algorytm ten przyporządkowuje każdemu z tych tekstów swoistą strukturę dogodną dla celów wyszukiwawczych.

II. NIEZALEŻNOŚĆ TREŚCIOWA DOKUMENTU

1. Jak już wspomnieliśmy baza danych odtwarzająca oryginalne teksty aktów normatywnych podzielona jest na tzw. dokumenty. Powstaje pytanie, jaką rozciągłość powinien mieć dokument na to, by system wyszukiwawczy mógł prawidłowo spełniać powierzone mu zadania.

2. Wyobraźmy sobie funkcjonowanie systemu, w którym jako poszczególne dokumenty traktowane są całe teksty aktów normatywnych. Przy takim rozwiązaniu najmniejszą porcję informacji wydawanej przez system stanowiłyby cały akt normatywny. System wyszukiwawczy reagowałby wtedy na kwerendy wprowadzane do niego przez użytkowników wydawaniem bądź jednego, bądź wielu aktów normatywnych. Stąd zdarzałoby się często, że użytkownik otrzymywałby — jako wynik poszukiwania — dużą liczbę takich aktów, np. materiał odpowiadający kilkunastu, czy nawet kilkudziesięciu stronom tekstu drukowanego. Odszukanie w tym materiale fragmentów niezbędnych dla odtworzenia interesujących użytkownika, w danym przypadku, norm prawnych wymagałoby wtedy przestudiowania go w całości, tj. przy włączeniu takich jego części, które nie zawierają informacji interesujących użytkownika. Obarczanie użytkownika takim zadaniem mijałoby się oczywiście z celem, któremu służyć mają systemy wyszukiwawcze. Jak widać takie rozwiązanie sprawy podziału bazy danych na dokumenty nie nadaje się do zaakceptowania.

3. Wyobraźmy sobie z kolei, dla kontrastu, system, w którym jako poszczególne dokumenty traktowane są wszystkie występujące w odpowiednich tekstach zdania gramatyczne. Rozwiązanie takie prowadziłoby do uzyskania wyników charakteryzujących się dużą dokładnością formalną.

⁵ Por. F. Studnicki, *Semantyka banków informacji prawniczej*, Państwo i Prawo 1979, nr 4.

Przez dokładność formalną wyniku rozumiemy tu jego właściwość polegającą na tym, że w zbiorze dokumentów wyszukanych w danym przypadku przez system liczba dokumentów relewantnych⁶ dla rozstrzygnięcia pytania interesującego użytkownika byłaby wysoka w stosunku do liczby dokumentów nierelevantnych. Duża dokładność formalna nie przesądza jednak sama w sobie o jakości uzyskanego wyniku. Jak wyjaśnimy w dalszym ciągu, dokładność taką mogą mieć również wyniki, przy których część możliwej do uzyskania informacji nie dociera do użytkownika systemu.

W prawidłowo zbudowanym tekście poszczególne zdania powiązane są ze sobą licznymi związkami semantycznymi nieobojętnymi ze stanowiska wyszukiwania zawartych w nim informacji. Nadto nie ulega wątpliwości, że sens niemal każdego z takich zdań uzależniony jest od kontekstu, w którym ono występuje, tj. między innymi od sensu zdań, które mu w danym tekście towarzyszą. Wydanie użytkownikowi dokumentu stanowiącego zdanie wyrwane z kontekstu (prowadziłoby najczęściej do zubożenia zawartej w nim informacji. W przypadkach krańcowych odtworzenie informacji o którą chodzi mogłoby się nawet okazać niemożliwe.

Przy omawianym tu rozwiązaniu ubytkom informacji można by zapobiec tylko wtedy, gdyby udało się wyposażyć system wyszukiwawczy w zdolność rozpoznawania powiązań semantycznych zachodzących między zdaniami tekstu. Umożliwiłoby to w szczególności wydawanie przez system takich fragmentów tekstu, które stanowiąc sekwencje występujących w tekście zdań mogłyby uchodzić za najmniejsze spośród jego fragmentów treściowo niezależnych (por. II.5). Uzyskanie takiego stanu rzeczy byłoby jednak równie trudne, jak wyposażenie systemu w zdolność pełnego rozumienia tekstów języka naturalnego, od czego jesteśmy jeszcze bardzo daleko.

Jak widać również rozwiązanie polegające na powierzeniu roli dokumentów występującym w tekście zdaniom nie nadaje się do zaakceptowania.

4. Przeprowadzone tu rozważania wydają się prowadzić do wniosku, że status dokumentu przysługiwać winien fragmentom tekstu możliwie najmniejszym spośród tych, które wyróżnione zostały przez normodawcę i potraktowane przez niego jako określone jednostki podziału, z uwagi na ich dostateczną niezależność treściową. Postulat, by były to jednostki możliwe najmniejsze uzasadniony jest tym, że jak wspomnieliśmy wyżej nadawanie dokumentom zbyt wielkich rozmiarów

⁶ O kryterium relewancji por. F. Studnicki, *Wprowadzenie do informatyki prawniczej. Zautomatyzowane wyszukiwanie informacji prawnej*, Warszawa 1978, s. 44-49.

prowadzi do występowania silnych szumów wyszukiwawczych. Podział tekstu pochodzący od normodawcy nazywać będziemy w dalszym ciągu podziałem naturalnym.

5. Niezależność treściowa określonego fragmentu tekstu jest jego właściwością stopniowalną. W szczególności przyjmujemy, że dany fragment tekstu jest w tym większym stopniu niezależny treściowo, w im mniejszym stopniu jego prawidłowe rozumienie uzależnione jest od wykorzystywania informacji zawartej w jego bliższym i dalszym kontekście. Uzyskanie pełnej niezależności treściowej fragmentu możliwe jest tylko w przypadkach wyjątkowych. Na kształtowanie się treści fragmentu wpływać może niekiedy zawartość innych fragmentów wielce od niego odległych. Z silnym uzależnieniem treściowym od zawartości innych części tekstu mamy do czynienia np. we fragmentach zawierających tzw. zwroty anaforyczne⁷, tj. zwroty odsyłające interpretatora danego fragmentu do informacji zawartej w innych fragmentach. Podobnie przedstawia się rzecz w przypadku, gdy dany fragment tekstu zawiera terminy zdefiniowane w innym fragmencie. Gdy chodzi o teksty aktów normatywnych to należy również pamiętać o tym, że całkowite oderwanie zawartości treściowej danego fragmentu tekstu od zawartości jego fragmentów pozostałych naruszyłoby jedną z głównych dyrektyw wykładni prawniczej, a mianowicie dyrektywę wykładni systematycznej.

III. STRUKTURA AKTÓW NORMATYWNYCH

1. Teksty aktów normatywnych są tworamii językowymi o szczególnej strukturze. Ze stanowiska teorii tekstu można w niej wyróżnić następujące składniki:

- 1) część, którą określać będziemy jako wstępną,
- 2) część zasadniczą oraz najczęściej również
- 3) część końcową.

Nadto do tekstów takich dołączone są niekiedy tzw. załączniki.

Część wstępna zawiera zwykle numer przyporządkowany danemu aktowi w dzienniku publikacyjnym (np. numer pozycji dziennika ustaw),

⁷ O pracach nad automatycznym rozwiązywaniem międzydokumentowych zwrotów anaforycznych występujących w tekstach prawnych, por. F. Studnicki, B. Polanowska, E. Stabrawa, J. M. Fall, A. Łachwa, *A semantic approach to automated resolving of interdocumental cross-references in legal texts*, w: *Computing in the Humanities*, R. W. Bailey (ed.), North-Holland Publishing Company, OC ICCH/5, 1982; F. Studnicki, B. Polanowska, E. Stabrawa, J. M. Fall, A. Łachwa, *The research project ANAPHORA (in its present state of advancement)*, w: *COLING 32 – Abstracts*, E. Hajicová (ed.), Charles University, Praga 1982.

określenie rodzaju aktu (np. ustawa, dekret, rozporządzenie, zarządzenie), datę wydania aktu i jego tytuł określający zwykle przedmiot regulacji. Część, o której mowa, zawierać może nadto wstęp, w którym podana jest podstawa prawna jego wydania i niekiedy tzw. preambułę, tj. wprowadzenie wyjaśniające okoliczności związane z wydaniem aktu, podające cel jego wydania itp.

Część zasadnicza podzielona jest niemal zawsze przez normodawcę na fragmenty stanowiące charakterystyczne dla aktów normatywnych jednostki podziału tekstu. Jednostki te nazywać będziemy jednostkami legislacyjnymi różnych poziomów (por. IV.5). Z częścią zasadniczą nie podzieloną na fragmenty spotykamy się tylko w aktach normatywnych bardzo krótkich.

Naturalny podział aktu normatywnego na jednostki legislacyjne (tj. podział naturalny) może być zrealizowany różnymi środkami. W szczególności na różnych poziomach tego podziału wystąpić mogą jednostki legislacyjne określane jako „części”, „działy”, „rozdziały”, „artykuły”, „paragrafy”, „ustępy”, „punkty” oznaczane liczbami arabskimi, rzymskimi, dużymi i małymi literami alfabetu, jednostki legislacyjne nie mające osobnych oznaczeń zaś wydzielone środkami typograficznymi (np. przez tzw. wcięcia) oraz duże fragmenty nie numerowane lecz oznaczone tytułem. Regułą jest, że fragmenty wyróżnione na wyższym poziomie podziału wykazują wyższy stopień niezależności treściowej niż fragmenty wydzielone na poziomach niższych.

Część końcowa zawierać może takie składniki, jak podpisy osób sprawujących w procesie legislacyjnym określone funkcje, oznaczenia charakteru owych funkcji, daty itp.

Rolę załączników pełnić mogą m. in. wykazy, plansze graficzne, teksty regulaminów, teksty statutów itp.

2. Akt nowelizacyjny ma strukturę nadaną mu przez jego podział naturalny, niezależny od podziału (charakteryzującego odpowiedni akt nowelizowany nawet w przypadku, gdy nowelizacja zmienia strukturę owego aktu. O podziale aktu nowelizującego na dokumenty decyduje wyłącznie struktura jego podziału naturalnego.

3. Kwerenda skierowana do systemu wyszukiwawczego dotyczy zwykle stanu prawnego w określonym dniu lub określonym czasie. System ma wtedy odnaleźć i wydać użytkownikowi dokumenty będące fragmentami aktów normatywnych, które obowiązywały w owym czasie. Często jednak zdarza się, że nie istnieje tekst jednolity aktu normatywnego obowiązującego w danym czasie, istnieje natomiast ciąg aktów, którego elementem pierwszym jest akt normatywny poddany następnie zmianom kolejno przez następujące po sobie akty nowelizujące. Nowelizacje takie mogą m. in. zmieniać strukturę naturalnego podziału tego pierwszego, we

wspomnianym ciągu, aktu. W takim przypadku system bazy danych⁸ będzie budował aktualny tekst jednolity i właśnie ten tekst będzie następnie dzielony na dokumenty.

IV. DEFINICJE

Obecnie wprowadzamy kulika definicji, które pozwolą zapisać nasz algorytm w sposób ścisły i przejrzysty.

1. *F r a g m e n t e m* nazywamy każdy ciąg wyrazów i znaków⁹ wyróżniony przez normodawcę jako jednostka podziału części zasadniczej aktu normatywnego wraz z przyporządkowanym tej jednostce oznaczeniem ((np. „§ 4”, „Przepisy wstępne”, „Art. 245”), jeżeli takie posiada. Dla uproszczenia dalszych rozważań także całą część zasadniczą nazywać będziemy fragmentem.

2. *F r a g m e n t* nazywamy *e l e m e n t a r n y m* jeżeli jest pojedynczym zdaniem gramatycznym albo sekwencją zdań nie podzieloną przez normodawcę na fragmenty będące zdaniem lub sekwencją zdań (por. przykład 1).

3. (Mówimy, że fragment A jest podzielony bezpośrednio na fragmenty A_1, A_2, \dots, A_n i że fragmenty te są jego bezpośrednimi składnikami, jeżeli nie istnieje fragment B , który zawierałby w sobie jakiś fragment A_i i był jednocześnie zawarty w A .

4. Fragmenty A_1, A_2, \dots, A_k nazywamy bezpośrednio sąsiadującym i z fragmentem A , jeżeli fragmenty A_1, A_2, \dots, A_k , a także fragment A są bezpośrednimi składnikami pewnego fragmentu B .

5. Mówimy, że część zasadnicza leży na pierwszym poziomie podziału. Mówimy, że fragmenty A_1, A_2, \dots, A_m leżą na k -tym poziomie podziału, jeżeli istnieją fragmenty B_1, B_2, \dots, B_m leżące na $k-1$ -szym poziomie podziału, i których bezpośrednimi składnikami są fragmenty A_1, A_2, \dots, A_m . Poziom k nazywamy niższym od poziomu $k-1$.

6. Podział danego fragmentu A na fragmenty A_1, A_2, \dots, A_n nazywamy *k o m p l e t n y m*, jeżeli jest to podział "bezpośredni i jeżeli zarazem suma fragmentów A_1, A_2, \dots, A_n daje cały fragment A z dokładnością

⁸ Systemem bazy danych nazywamy bazę danych wraz z pakietem procedur służących do wykonywania pewnych standardowych operacji na owej bazie (np. wprowadzania danych). Por. C. J. Date, *Wprowadzenie do baz danych*, Warszawa 1081,

⁹ Tabele, plansze itp. twory graficzne traktować będziemy jako znaki, podobnie do znaków alfabetu („a”, „b”, ... „§”, „:”).

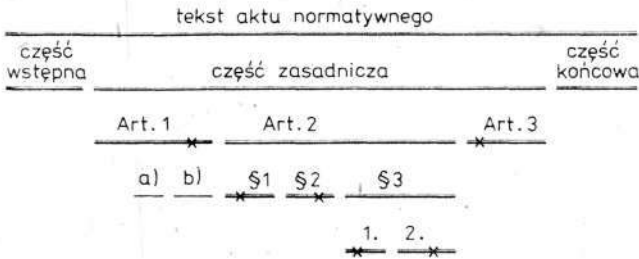
do oznaczenia tego fragmentu (por. przykład 2). Podział 'bezpośredni' nazywamy niekompletnym, jeżeli nie jest podziałem kompletnym.

7. Fragment *A* nazywamy podstawowym, jeżeli jest fragmentem elementarnym albo fragmentem podzielonym w sposób niekompletny, i jeżeli jednocześnie każdy fragment *B*, którego częścią jest fragment *A*, jest podzielony w sposób kompletny.

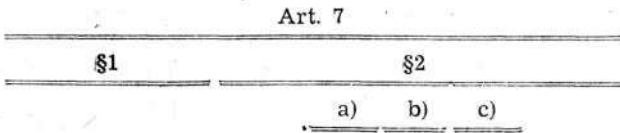
Mówimy, że fragment *B* jest podzielony w sposób podstawowy, jeżeli jego wszystkie bezpośrednie składniki są fragmentami podstawowymi (por. przykład 3).

Przykład 1

Schemat struktury podziału naturalnego pewnego tekstu z zaznaczonymi fragmentami elementarnymi (linie z krzyżykiem) i fragmentami mniejszymi od zdania (linie pojedyncze):

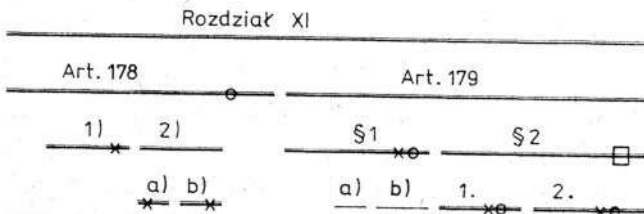


Graficzne odróżnienie podziału kompletnego (artykułu na paragrafy) i niekompletnego (paragrafu na podpunkty):



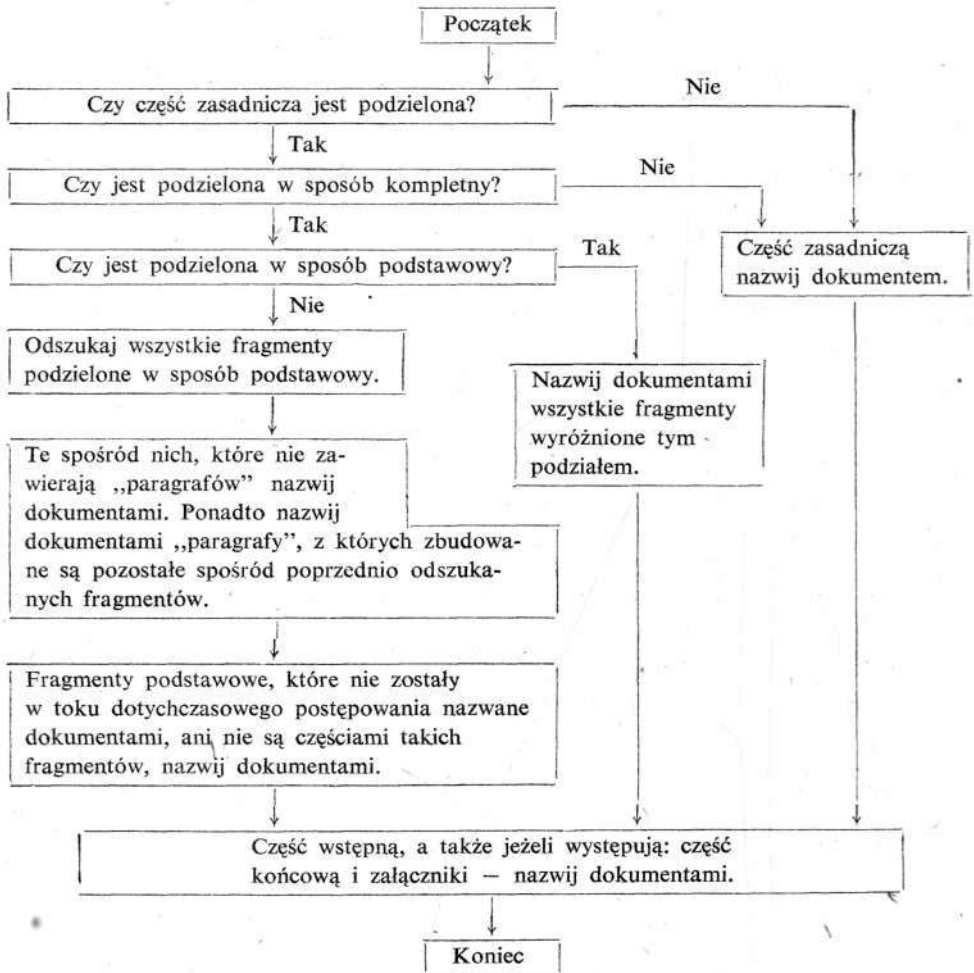
Przykład 3

Schemat struktury podziału naturalnego pewnego fragmentu tekstu z zaznaczonymi fragmentami elementarnymi (linie z krzyżykiem), fragmentami podstawowymi (linie z kółkiem) i fragmentem podzielonym w sposób podstawowy (linia z kwadratem):



V. ALGORYTM

W przedstawionym poniżej schemacie blokowym algorytmu podziału tekstu aktu normatywnego na dokumenty, terminu „paragraf” należy używać tylko w odniesieniu do aktów normatywnych publikowanych w Monitorze Polskim. Dla aktów publikowanych w Dzienniku Ustaw termin ten należy zastąpić terminem „artykuł”. Z tego też powodu termin „paragraf” został tu (użyty w cudzysłowie).



Działanie przedstawionego algorytmu polega na tym, że analizuje on pierwotny tekst prawny, a w szczególności strukturę jego podziału naturalnego (jest to struktura drzewiasta wyróżniająca jednostki o różnej wielkości i w (różnym stopniu niezależne treściowo) i w wyniku tej analizy przyporządkowuje temu tekstowi nową strukturę podziału dogodną

dla celów wyszukiwawczych. Jest to struktura sekwencyjna jednostek o podobnym — znacznym — stopniu niezależności treściowej i podobnej — niedużej¹⁰ — wielkości.

Przykład 4

Dwa schematy struktur pewnego tekstu: a) schemat struktury naturalnego podziału tekstu, b) schemat struktury nadanej algorytmem;

a)

tekst aktu normatywnego						
część wstępna	część zasadnicza					
	Rozdz. I		Rozdz. II		Rozdz. III	
	§1	§2	§3	§4	§5	
					1. 2.	
					a) b)	

b)

tekst aktu normatywnego						
część wstępna	§1	§2	§3	§4	1. 2.	Rozdz. III

VI. PODSUMOWANIE

Stosując pewne uproszczenie można przyjąć, że opisany tu algorytm nadaje charakter dokumentów tym spośród jednostek stanowiących ciągi zdań lub zdania, które leżą na drugim od dołu poziomie podziału naturalnego. Wyjątkowo algorytm nadaje ten charakter jednostce ulokowanej na najniższym poziomie podziału, jeżeli jest ona paragrafem lub artykułem. Wyjątek ten usprawiedliwiony jest tym, że w zbadanych w toku przygotowywania niniejszego opracowania tekstach¹¹ występuje charakterystyczny podział aktów publikowanych w Monitorze Polskim na paragrafy złożone niekiedy z ustępów, zaś aktów publikowanych w Dzienniku Ustaw na artykuły złożone niekiedy z paragrafów oraz tym, że takie paragrafy (wyróżnione w Monitorze Polskim) i artykuły (wyróżnione w Dzienniku Ustaw) są zazwyczaj w znacznym stopniu niezależne treściowo.

¹⁰ Dotyczy to tylko dokumentów części zasadniczej, bowiem dokumenty-załączniki mogą być bardzo duże.

¹¹ Zbadany materiał tekstowy dotyczy aktów normatywnych prawa polskiego wydawanych od 1944 r.

Algorytm zbudowany został (tak; aby podziały tekstów na dokumenty otrzymywane w wyniku jego działania maksymalnie ułatwiały wyszukiwanie. Stąd też największą uwagę skoncentrowano na części zasadniczej aktu normatywnego, bowiem właśnie fragmenty tej części są najczęściej wyszukiwane i wydawane przez system wyszukiwawczy.

SUBDIVISION OF STATUTORY TEXTS INTO DOCUMENTS

Summary

Functioning of the full text legal data banks consists in providing its users with such fragments of the legal text as might be relevant to resolving definite legal questions. A proper fulfilling of this task is dependent i.e. on the way in which the text in question are subdivided into smaller units denoted by the term documents. The author discusses the problem of the optimal size of such units.

Subdividing of the texts into documents is usually carried out by using manual methods. The author presents a method of subdividing the legal texts into documents automatically. The method refers to what may be called a natural subdivision, i.e. to the subdivision applied by the law-maker. The outlines of an algorithm of the operation of automated subdividing are presented in the form the corresponding flow diagram.