Ku polskim klasom dokumentów dla LaTeXa

Marcin Woliński
wolinski@mimuw.edu.pl

Streszczenie

Niniejszy artykuł dotyczy moich własnych klas dokumentów dla LaTeXa. Bezpośrednim impulsem do ich przygotowania było pewne przykre zjawisko. Mianowicie za każdym razem, gdy zdarza mi się zabierać do składu polskiej książki w LaTeXu, muszę zaczynać pracę od wprowadzania wciąż tych samych zmian do standardowych klas. Drugi czynnik kształtujący moje klasy, to chęć wprowadzenia rozszerzeń sprawiających, że klasy staną się łatwiej/bardziej konfigurowalne niż standardowe.

Tekst jest poświęcony wybranym problemom typograficznym, na które natknąłem się w trakcie implementacji moich klas. Klasy te nie są jeszcze gotowe. W końcowej części tekstu przedstawiono stan prac i możliwe dalsze rozszerzenia.

Wprowadzenie

Zagadnienie polonizacji LaTeXa można rozbić na dwa poziomy. Poziom podstawowy (który można by nazwać poziomem mikrotypografii) obejmuje takie problemy, jak dostęp do narodowych znaków diakrytycznych w fontach, obsługa odpowiedniości między minuskułami i majuskułami, obsługa znaków pisarskich specyficznych dla naszego języka, poprawne dzielenie wyrazów przy przenoszeniu do nowego wiersza. Rozpatrując rzecz z punktu widzenia LaTeXa, są to elementy niezależne od klasy dokumentów stosowanej do składu. Istnieją co najmniej dwa pakiety (PLaTeX i Babel) realizujące ten rodzaj polonizacji.

Drugi poziom (który można by nazwać makrotypograficznym) to takie sprawy jak układ kolumny tekstowej, wielkość świateł, sposób konstrukcji tytulariów i wiele innych szczegółów, o których opowiada niniejszy tekst. W LaTeXowej infrastrukturze za te elementy odpowiada klasa dokumentów i stąd inicjatywa przygotowania „polskich klas dokumentów”.

Moje własne klasy dokumentów stanowią zastępniki trzech klas zawartych w dystrybucji LaTeXa. Mianowicie klasa `mwart` jest odpowiednikiem klasy `article`, `mwrep` — `report`, `mwbk` natomiast — klasy `book`. Autor pragnie zaznaczyć, że nie jest tak nieskromny, by umieścić swoje inicjały w nazwach klas. Litery `mw` są skrótem od słów „mój własny”. Tak więc np. `mwart` to „mój własny artykuł”.

Ze względu na opisane powyżej rozbicie poziomów, klasy te wymagają użycia pakietu polonizującego. Biorąc pod uwagę brak zgodności między użytkownikami LaTeXa co do stosowanego pakietu, wydaje się to właściwym rozwiązaniem.

Oto krótki dokument przygotowany z użyciem klasy `mwart`:

```
%& --translate-file=il2-pl
\documentclass{mwart}
\usepackage{polski}

\begin{document}
  Pójdź kińże tę chmurność w głąb flaszy.
\end{document}
```

Jeżeli używana implementacja TeXa nie obsługuje przekodowywania polskich znaków, wymagane jest załadowanie pakietu `inputenc`.

(Uwaga techniczna: niniejszy artykuł nie został złożony przy pomocy moich klas. Przykład ich użycia można znaleźć w dokumentacji pod adresem wskazanym na końcu artykułu).

Dalsza część tekstu jest poświęcona omówieniu wybranych zagadnień typograficznych, które trzeba było uwzględnić w ramach implementacji klas.

Wcięcia akapitowe

Norma branżowa [1] zaleca stosowanie w składzie do 24 cycer wcięć o wielkości 1 firetu, powyżej 24 cycer $1\,{}^1\!/_2$ firetu. Instrukcja technologiczna [3] dopuszcza w tym ostatnim przypadku wcięcia od 1 do 2 firetów. W moich klasach trzymam się ściśle zaleceń normy.

W zaleceniach normatywnych brak odniesień do anglosaskiego zwyczaju składania bez wcięcia pierwszego akapitu po tytularium. Wydaje się zatem, że polska tradycja polega na stosowaniu jednakowych wcięć we wszystkich akapitach tekstu. Tak też dzieje się w moich klasach. Ponieważ jednak zwyczaj pomijania wcięć wydaje się zyskiwać pewną popularność (por. np. niniejsze materiały konferencyjne), można go uaktywnić stosowną opcją.

Końcowy wiersz akapitu

Wedle zaleceń, ilość tekstu w wierszu końcowym powinna być co najmniej dwa razy większa od wcięcia akapitowego. Realizacja tego postulatu w TEXu nastręcza pewnych problemów. Na końcu akapitu TEX kładzie klej zadany przez wartość `\parfillskip`. Klej ten służy do dopełnienia ostatniego wiersza akapitu do normalnej długości. Wielkość tego kleju wynosi zwykle `0pt plus1fil`, czyli jest on wyposażony w zdolność dowolnego rozciągania się, co pozwala na skonstruowanie dowonie niepełnego wiersza.

Omawiane ograniczenie można by wprowadzić poprzez zmianę rozciągliwości tego kleju tak, aby mógł się on rozciągnąć jedynie do szerokości łamu pomniejszonej o podwojoną wielkość wcięcia akapitowego.

Jednak nadanie skończonej rozciągliwości parametrowi `\parfillskip` powoduje przykry skutek uboczny. Jak wyjaśniono w *The TEXbook* (rozdział 19, s. 188), w takiej sytuacji TEX wyłącza mechanizm dostosowywania odstępu nad wystawioną formułą matematyczną do stopnia wypełnienia poprzedzającego wiersza tekstowego.

Ponieważ nie jestem pewien, czy jest to cena, którą warto zapłacić, nie wprowadziłem stosownej modyfikacji. Być może warto byłoby w e-TEXu usunąć omawiane ograniczenie.

Szewce i bękarty

Źródła wykazują ciekawą rozbieżność w kwestii tzw. szewców i bękartów. Według normy [2]:

> **2.2.10. Wiersz końcowy niepełny na początku kolumny (bękart).** Umieszczenie końcowego wiersza akapitu na początku kolumny jest dopuszczalne.
>
> Jeśli instytucja wydawnicza żąda formowania kolumn bez tzw. „bękartów”, powinna w dyspozycji technicznej wskazać sposób ich uniknięcia z wykluczeniem dorabiania lub gubienia wierszy przez przeskładywanie tekstów bez zmiany treści oraz zwiększania odstępów między wierszami. Dopuszcza się w takich przypadkach skrócenie obu kolumn na rozwarciu o 1 wiersz z pozostawieniem pagin na właściwym miejscu.

natomiast Instrukcja [3] na s. 189 informuje:

> i) Nie należy umieszczać na początku kolumny lub łamu niepełnego końcowego wiersza akapitu tzw. „bękarta”. Umieszczenie pełnego wiersza końcowego na początku kolumny jest dopuszczalne wówczas, gdy książka zawiera dużo akapitów.
>
> j) Zaleca się unikać, zwłaszcza w książkach zawierających małą ilość akapitów na kolumnie, pozostawiania u dołu kolumny pierwszego wciętego wiersza akapitowego. Jeśli jednak zalecenie to byłoby możliwe do wykonania tylko przez rozbicie kolumny na dodatkowe interlinie i naruszenie zasady registru, należy pozostawić ten wiersz u dołu kolumny.

Indagowani fachowcy zeznają, że bękarty są „błędem składu”, natomiast szewce można od biedy dopuścić.

W klasach zostało zablokowane tworzenie bękartów, natomiast tworzenie szewców zostało obłożone sporą grzywną (*penalty*).

Łamanie wiersza na dywizie

W moich klasach parametrowi `\exhyphenpenalty` przypisuję wartość 10 000. Parametr ten jest grzywną za łamanie wiersza po dywizie i ligaturach kończących się dywizem. Zastosowana wartość zaś oznacza zakaz łamania.

Uzasadnienie jest następujące: łamanie wiersza na dywizie wymaga powtórzenia go na początku następnego wiersza. Takiego zachowania nie da się w sensowny sposób uzyskać przy pomocy znaku `-`. Dlatego blokujemy łamanie na takim dywizie, uznając za błędny zapis `biało-czerwony`. (Prawidłowy zapis tego słowa wymaga użycia makra, które spowoduje powtórzenie dywizu po przeniesieniu: `biało\dywiz czerwony`.)

W języku polskim wokół myślnika (pauzy) stosuje się odstępy, tak więc łamanie może nastąpić na odstępie i dodatkowa grzywna nie jest potrzebna. (W języku angielskim brak odstępów i grzywna musi zostać przez TEXa dostawiona, żeby w ogóle umożliwić złamanie wiersza).

(Pół)pauzy bez odstępów stosuje się sporadycznie w zakresach liczbowych: 21–59. W tym przypadku również wydaje się rozsądne zablokowanie łamania.

Wartość `\exhyphenpenalty` wpływa również (w tym przypadku blokująco) na wszystkie użycia polecenia `\discretionary` z pustym pierwszym argumentem. Wydaje się, że nie stanowi to problemu w tym sensie, że takich użyć daje się prosto uniknąć.

Paginacja

Sposób składu żywej paginy określa Instrukcja [3] na s. 190 następująco:

> d) Paginy żywe składa się pismem mniejszym od pisma, którym złożony jest tekst książki, przy

czym można je składać antykwą lub kursywą, kapitalikami lub wersalikami. Zaleca się, aby żywa pagina wraz z odstępem od tekstu zajmowała miejsce najwyżej dwóch wierszy tekstu podstawowego.

W moich klasach domyślnie używana jest w żywej paginie kursywa stopnia o jeden mniejszego od podstawowego. Lista dostępnych stylów paginacji została uzupełniona o wersję z żywą paginą w główce strony oddzieloną od łamów cienką linią.

W [3] znajdujemy na tej samej stronie zalecenia odnośnie pomijania paginacji:

ł) Nie należy umieszczać górnej paginy w kolumnach spuszczonych. Kolejną liczbę stronicy spuszczonej można składać nonparelem w klamrach i umieszczać na środku u dołu poza formatem kolumny.

m) Nie należy umieszczać dolnej paginy w kolumnach końcowych. Zasada ta nie obowiązuje w składzie poezji, zbiorze przepisów prawnych, cennikach itp. dziełach, w których zachodzi potrzeba powoływania się na numerację danej strony, oraz jeśli pagina dolna jest składana na środku i stanowi element dekoracyjny kolumny. W tych przypadkach nie daje się paginy w kolumnach końcowych książki i kolumnach poprzedzających tytuły na oddzielnych kolumnach lub części umieszczone na oddzielnych nieparzystych stronicach.

n) Nie należy umieszczać pagin na kolumnach tytułowych, tytułach części oraz kolumnach zajętych całkowicie przez klisze. Jeśli jednak książka szkolna lub techniczna zawiera dużo ilustracji lub rysunków całostronicowych, to paginy należy na nich umieścić. Paginę należy także umieszczać w przypadku następowania kilku ilustracji całostronicowych — jedna za drugą w ciągu.

Norma [2] zawiera podobne zalecenia w punktach 2.2.6 i 2.2.7.

LaTeXowy mechanizm obsługi stylów paginacji (zwanych w LaTeXu stylami stron — `\pagestyle`) musiał zostać rozszerzony, aby obsłużyć takie subtelności.

Wprowadzono mianowicie cztery typy kolumn: kolumny zwykłe z pełną paginacją; kolumny spuszczone, na których pomija się paginę górną; kolumny niepełne, na których pomija się paginę dolną; kolumny wakatowe bez paginacji. W moich klasach wybór stylu stron poleceniem `\pagestyle` oznacza w istocie wybór całego zestawu definicji dla poszczególnych typów stron zamiast jednej definicji.

Automatyczne przypisywanie typów poszczególnym stronom (odbywające się przy pomocy polecenia `\thispagestyle`) nastręcza pewnych problemów w związku z asynchronicznym sposobem pracy TeXowego budowniczego stron. W obecnej implementacji może się zdarzyć, że dwustronicowy rozdział będzie miał źle przypisany typ pierwszej strony. Prace nad poprawieniem tego błędu są w toku.

Tytularia

Wydaje się, że w polskiej tradycji jest silnie zakorzeniony zwyczaj podawania na karcie tytułowej najpierw autora, a dopiero potem tytułu dzieła (anglosasi robią odwrotnie, jak widać dla nich ważniejsze jest dzieło niż autor, przeciwnie niż dla nas). Gdy w informacji tytułowej zawarty jest adres autora, jest chyba zgrabniej umieścić go przed tytułem. Dlatego w moich klasach kolejnością danych autora i tytułu sterują stosowne opcje klasy.

Domyślnie w informacjach tytułowych nie ma daty. Efekt taki, jak w klasach standardowych, można uzyskać stosując polecenie `\date{\today}`.

Następujące zalecenia normy [1]

2.2.1. Składanie tytułów. [...]

W tytułach nie należy przenosić wyrazów. [...]

Numerację dziesiętną należy składać pismem prostym bez odstępów między cyframi i kropkami, lecz z odstępem $^1/_2$ firetu od treści tytułu.

wymagają odmiennego zrealizowania tytulariów niż w jądrze LaTeXa. (Powyższe zalecenia wraz z wymaganiem dostawienia kropki po numerze podrozdziału, są chyba najbardziej uporczywie powracającymi zapytaniami na liście dyskusyjnej GUST-L).

Dodatkowe utrudnienie stanowią następujące wymagania normy [2]:

2.2.3. Kolumna spuszczona. W kolumnach spuszczonych wielkość spuszczenia powinna być jednakowa w całym wydawnictwie i wynosić od $^1/_5$ do $^1/_3$ długości pełnej kolumny.

[...] Wielkość spuszczenia należy określać od góry kolumny do pierwszego wiersza tekstu podstawowego.

Zalecenia te zostały na razie zaimplementowane częściowo: polecenie `\chapter` będące jedynym źródłem kolumn spuszczonych zostało zdefiniowane tak, aby zawsze produkować pudełko o tej samej wysokości. W środku tego pudełka umieszczany jest tytuł rozdziału tak, aby jego ostatnia linia znalazła się w ustalonej odległości od dna pudełka.

Wątpliwość natomiast budzi zalecane przez normę wliczenie do spuszczenia ewentualnego tytułu podrozdziału następującego bezpośrednio po tytule rozdziału.

Przypisy

Według norm przypisy (notki) u dołu strony należy składać pismem tego samego kroju o jeden lub dwa stopnie mniejszym od podstawowego. Wcięcie akapitowe notek musi być takie samo jak w tekście podstawowym. Tekst notki zaczyna się od powtórzenia znaku odsyłacza, po którym następuje stały odstęp pół firetu i dalej treść notki.

Jest to układ odmienny od zwyczajów LaTeXa, który wpuszcza znak odsyłacza we wcięcie pierwszego akapitu notki.

W normie [2] czytamy:

> **2.2.12. Umieszczenie przypisów**
>
> **2.2.12.1. Kolumny jednołamowe tekstu zwykłego.** Przypisy należy umieszczać u dołu kolumny.
>
> Na kolumnach szpicowych [niepełnych] przypisy należy umieszczać bezpośrednio pod tekstem.
>
> Wyjątek stanowią kolumny szpicowe krótsze o 1–3 wierszy, na których przypisy należy umieszczać u dołu kolumn. W tym przypadku odstęp między linią a przypisem powinien wynosić $1/2$ wiersza podstawowego.
>
> Dopuszcza się inne sposoby umieszczania przypisów.

Ponieważ implementacja tego punktu wydaje się wymagać ingerencji w procedurę wyjścia LaTeXa, na razie pozwalam sobie korzystać z ostatniego zdania powyższych zaleceń i traktuję notki na kolumnach niepełnych tak samo jak na pozostałych.

Wyliczenia

Zalecenia normy [1] dotyczące składu wyliczeń:

> **2.1.9. Wyliczenia.** Przy składaniu krótkich wyliczeń liczby lub litery wyliczeń justuje się tak, aby były wypuszczone w lewo, a tekst wyjustowany był w linii pionowej.
>
> Odstęp za liczbą lub literą pozostaje niezmienny i powinien wynosić do $2/3$ firetu.
>
> Jeżeli wyliczenia są długie i zajmują trzy lub więcej wierszy, należy je składać akapitowo, odstęp za liczbą lub literą w tym przypadku jest stały i powinien wynosić $1/2$ firetu.

Oczywiście rozróżnienie między wyliczeniami „krótkimi" i „długimi", jest niezbyt ostre i nie należy traktować go mechanicznie. Na przykład czy długie jest wyliczenie, którego tylko jeden punkt ma więcej niż trzy linie, a pozostałe są bardzo krótkie? Dlatego przyjmuję następujące rozwiązanie: środowiska `\begin{enumerate}` i `\begin{itemize}` są zdefiniowane tak, aby realizować możliwie wiernie zalecenia dla „krótkich" wyliczeń. Oprócz tego wprowadzam nowe środowiska `\begin{enumerate*}` i `\begin{itemize*}` dla wersji „długich". Do użytkownika należy decyzja, którego środowiska użyć dla danego wyliczenia.

Poboczne części tekstu

Do polskiej tradycji należy skład pobocznych części tekstu, to jest spisu treści, spisu tablic, spisu rysunków, skorowidza i bibliografii pismem o jeden lub dwa stopnie mniejszym od podstawowego.

Spis treści

Spis treści proponowany przez moje klasy można by nazwać ascetycznym w porównaniu z tym z klas standardowych. Wszystkie poziomy tytułów są składane pismem tego samego stopnia (tego co pozostałe części poboczne) i wszystkie są wyposażane w kropkowanie wiodące ku numerowi stronicy. Wyjątek stanowią tytuły części, centrowane i składane bez numeru stronicy.

Ulepszenie w stosunku do klas standardowych stanowi dodawanie grzywien pomiędzy punktami spisu treści. Sprawia ono, że blokowane są łamania między tytułem danego rzędu a następującym po nim tytułem niższego rzędu. Z kolei jeśli następny punkt jest tytułem wyższego rzędu, dane miejsce jest preferowane jako potencjalny koniec strony.

Rozszerzenia

Moje klasy różnią się od standardowych kilkoma własnościami, które nie są specyficzne dla naszego języka, ale stanowią poprawki (w naszym mniemaniu) usterek LaTeXa.

Jedną z takich zmian jest sposób zachowania się poleceń wprowadzających podział dokumentu na części, rozdziały, punkty, podpunkty itd. Różnice omówimy na przykładzie polecenia `\section`, pozostałe polecenia zachowują się analogicznie. W wersji najprostszej polecenie:

```
\section{Tytuł punktu}
```

powoduje rozpoczęcie punktu o zadanym tytule. Tytuł ten pojawi się też w żywej paginie (o ile wymaga tego bieżący styl paginacji) i w spisie treści.

Czasami zdarza się, że tytuł punktu jest za długi, aby zmieścić się w żywej paginie. W takim przypadku w paginie powinna znaleźć się wersja skrócona. Można ją zadać za pomocą opcjonalnego argumentu:

```
\section[Tekst paginy]{Tytuł punktu}
```

Polecenia zdefiniowane w klasach standardowych w takim przypadku również w spisie treści umieszczają tekst skrócony. Wydaje się nam to niewłaściwe. W moich klasach powyższa postać polecenia umieszcza w spisie treści ten sam tekst co w kolumnie. Wprowadzamy natomiast kolejny opcjonalny argument pozwalający zadać wariant tytułu dla spisu treści:

```
\section[Tekst paginy]
  [Tytuł w~spisie treści]
  {Tytuł punktu}
```

(Jedyne zastosowanie, które widzimy dla tego argumentu, to wymuszenie innego łamania na wiersze w spisie treści i w kolumnie).

Oprócz tego dostępne są wersje z gwiazdką wszystkich poleceń tytułowych. Za pomocą tych poleceń można uzyskać nagłówki nie numerowane. Odmiennie niż w klasach standardowych te wersje również powodują wpisy w spisie treści i w żywej paginie. Można przy nich stosować opcjonalne argumenty.

Stan prac

Jak zaznaczono na wstępie, polskie klasy dokumentów nie są jeszcze gotowe. Udało się doprowadzić do działania ich podstawowe części pozwalające złożyć prosty dokument. Nie jest to jednak wersja stabilna; w szczególności autor na razie nie gwarantuje, że następne wersje klas będą dawać identyczne wyniki składu.

Nadal wymagają przemyślenia niektóre co bardziej skomplikowane elementy, jak tablice i rysunki (*floats*). Dalszych badań i zabiegów wymaga zapewnienie zgodności z innymi pakietami. Na przykład można się spodziewać przykrych niespodzianek ze strony pakietu `hyperref`.

Ciekawy kierunek rozwoju klas może też stanowić próba wykonania składu z pełnym zachowaniem zasady registru.

Najbardziej aktualną wersję klas można znaleźć w Internecie pod adresem `http://www.mimuw.edu.pl/~wolinski/mwcls.html`

Autor zachęca czytelników do eksperymentów i będzie wdzięczny za wszelkie komentarze, szczególnie te wskazujące błędy i miejsca wymagające dalszego dopracowania.

Bibliografia

[1] BN-76/7440-02 *Zasady składania tekstów w języku polskim.*

[2] BN-76/7440-03 *Zasady formowania kolumn książek, broszur i czasopism.*

[3] *Instrukcja technologiczna. Skład ręczny i maszynowy.* Ministerstwo Kultury i Sztuki, Warszawa 1965.

[4] Barbara Osuchowska, *Poradnik redaktora i autora. Nauki ścisłe i technika.* Wydawnictwo PTWK, Warszawa 1988.

[5] Filip Trzaska, *Podstawy techniki wydawniczej.* Warszawa 1967.

◇ Marcin Woliński
`wolinski@mimuw.edu.pl`