

Mateusz Poradecki*

GREP w pracy redaktora – wprowadzenie do zagadnienia¹

Jedną z wielu zalet wykorzystywania komputerów do prac związanych z redakcją, składem i łamaniem tekstów jest niewątpliwie możliwość automatyzacji pracy. Dużą część powtarzających się, nużących czynności możemy wykonać znacznie szybciej dzięki profesjonalnemu oprogramowaniu, ograniczając tym samym rolę człowieka do nadzoru. Większość znanych dziś pożytecznych narzędzi powstała już w latach osiemdziesiątych i dziewięćdziesiątych XX wieku, kiedy komputery zaczęły wypierać maszyny do pisania – najpierw z redakcji dużych wydawnictw, a później także z biur pisarzy i autorów. Możliwość poprawiania tekstu bez żmudnego przepisywania całych stron w przypadku zaistnienia błędu czy też nawet pisania od razu „na czysto” to jeden z najbardziej docenianych, choć dziś uważanych już za oczywisty atutów komputerów typu PC. Chociaż, jak twierdzi znany historyk Norman Davies, tylko pisanie odręczne daje poczucie bliskiego kontaktu z własną pracą i zmusza do formułowania myśli w sposób odpowiedzialny już przed utwaleniem ich na papierze, zdecydowana większość osób parających się zawodowo pisaniem raczej nie zdecydowałaby się na porzucenie wygody używania edytora tekstu. Jednak pisanie na komputerze, jeśli ma ono spełniać choćby minimalne wymagania jakościowe stawiane przez wydawnictwa i ułatwiać dalszą obróbkę tekstu, wymaga choć krótkiego przeszkolenia. Nie będziemy tu rozważać kwestii merytorycznych, lecz wyłącznie techniczne.

Świadomość użytkowników komputerów jest na raczej niskim poziomie i w ciągu ostatnich dziesięciu lat niewiele wzrosła². Nawet jeśli już zauważamy, że komputery wkraczają do coraz to nowych dziedzin życia i coraz więcej osób posługuje się nimi, wykonując coraz to więcej czynności, to nie oznacza wcale, że nasza znajomość tak zwanych pecetów jest lepsza. To raczej oprogramowanie staje się coraz łatwiejsze i bardziej ergonomiczne w użyciu. Wrodzona tendencja

* Dr, Uniwersytet Łódzki, Wydział Filologiczny, Instytut Filologii Polskiej, Katedra Edytorstwa, Uniwersytet Łódzki, ul. Pomorska 171/173, 91-404 Łódź; e-mail: mporadecki@gmail.com

¹ Zagadnienia przedstawione w artykule są wynikiem własnych prac badawczych. W języku polskim ukazała się tylko jedna pozycja, będąca poradnikiem używania narzędzia wyszukiwania GREP: M. Włodarz, *GREP w InDesign CS3–CS5*, MAWart, Warszawa 2011.

² Powyższą tezę formułuję na podstawie własnych wieloletnich doświadczeń wydawcy oraz wykładowcy na co dzień pracującego ze studentami w pracowni komputerowej.

do chodzenia na skróty i niechęć do czytania wszelkiego rodzaju instrukcji czy samouczków powoduje, że większość z nas wykorzystuje tylko fragment potencjału kryjącego się w cyfrowym świecie. Oczywiście na obecnym poziomie rozwoju informatyki nie jesteśmy w stanie skonstruować sztucznej inteligencji mogącej zastąpić redaktora w wydawnictwie, ale wiele omyłek piszących na komputerach swoje teksty autorów można wyeliminować przy minimalnym nakładzie pracy człowieka, dzięki wykonaniu tych czynności na nieomylnym i dużo szybszym komputerze.

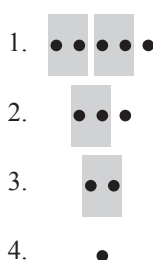
Każde wydawnictwo ma swoje standardy i wymagania dotyczące zasad przyjmowania tekstów od autorów, ale większość z nich to ogólnie przyjęte reguły – niektóre z nich skodyfikowane przez słowniki ortograficzne czy Polską Normę, a jeszcze inne wynikające z tradycji oraz innych powodów, na przykład dotyczących czytelności tekstu czy jego estetyki. Wielu z tych wymagań autorzy nie są w stanie spełnić i praca ta naturalną kolejną rzeczą zostaje przerzucona na redaktora bądź składacza.

Najpopularniejszym narzędziem, które na dodatek wbudowane jest w każdy edytor tekstu i program zecerski, jest funkcja **Znajdź/Zamień**³. Pozwala ona programowi wyszukać określony tekst i zamienić go na inny. W zależności od używanej aplikacji możemy robić to krok po kroku, przyglądając się każdemu przypadkowi, lub wydać komendę **Zamień wszystko**, jeśli jesteśmy przekonani, że nie spowoduje to powstania innych błędów. Tym sposobem można łatwo poprawić bardzo często spotykane błędy, takie jak na przykład: spacje przed kropką, przecinkiem, dwukropkiem, zamknięciem nawiasów i cudzysłowów, elipsą, wykrzyknikiem, znakiem zapytania itp. Tą metodą poprawiamy także inne powtarzające się błędy, jak choćby konsekwentną złą pisownię nazw własnych i tym podobne. Analogicznie usuwamy wielokrotne spacje, choć w tym przypadku czynność należy powtórzyć kilka razy. Komputer bowiem nie rozumie naszych intencji i wyszukuje dokładnie to, co zostanie mu zadane. Dlatego należy brać pod uwagę ograniczenia oprogramowania i zczasu się do nich dostosować. W tym przypadku procedura wygląda następująco – w pole **Znajdź** wstawiamy dwie spacje, w pole **Zamień na** – jedną. Komputer wyszukuje każdy przypadek, w którym występują obok siebie dwie spacje i zamienia je na jedną. Jeśli jednak w bezpośrednim sąsiedztwie znajdować się będzie więcej spacji, to omawiana funkcja nie spełni swojego zadania (patrz ryc. 1).

W trakcie wyszukiwania (1) komputer odnajduje podwójne spacje (pierwszy szary prostokąt), zamienia je na pojedyncze i przechodzi do dalszej części tekstu. Ponownie odnajduje podwójne spacje (drugi szary prostokąt), zamienia je na pojedyncze i przechodzi dalej – spacja pojedyncza zostaje pominięta. W przypadku

³ Nie biorę tu pod uwagę wariantów nazewnictwa tej funkcji, która w różnych aplikacjach i ich wersjach może brzmieć trochę inaczej.

pięciu spacji po jednokrotnym przeszukaniu zostaną one zamienione na trzy (2). Po drugim – na dwie (3). Czyli dopiero wielokrotne powtórzenie tej funkcji przynosi oczekiwane rezultaty – powinna być ona powtarzana dopóki nie otrzymamy komunikatu, że liczba znalezionych przypadków wynosi zero. Największą wadą tego narzędzia i zarazem tej metody jest ograniczona możliwość wyszukiwania tylko spacji jednego rodzaju, podczas gdy profesjonalny skład wykorzystuje ich kilka, między innymi: chudą, włoskową, cyfrową, twardą (nierozdzielającą) o stałej szerokości, twardą o zmiennej szerokości, firetową, półfiretową, interpunkcyjną. Komputer traktuje każdą z nich jako osobny znak i przy takim użyciu tego narzędzia pomija kombinacje różnych spacji.



Ryc. 1. Czarne kropki oznaczają spacje, szare prostokąty to obszary znalezione przez funkcję **Znajdź/Zamień**.

Funkcja **Znajdź/Zamień** ma także zastosowanie w przygotowaniu tekstu do składu. Wiele udogodnień, które na późniejszych etapach pracy zaowocują dużą oszczędnością czasu, może zostać wprowadzonych do tekstu już na etapie redakcji. Jednym z nich jest na przykład wymiana trzech kropek na elipsę czy dywizów na pauzy lub półpauzy. Narzędzie to może pomóc w ujednoceniu zapisów oznaczeń walutowych (jeśli redaktor zdecyduje, że w całej publikacji będzie stosować zapis **PLN** zamiast **zł**) czy skrótów (**dz. cyt.** na **op. cit.**). Bardzo często używa się funkcji **Znajdź/Zamień** do wstawiania tzw. **twardych spacji** po pojedynczych znakach. W języku polskim pozostawianie **u**, **w**, **i**, **o**, **z**, **a**⁴ na końcu wiersza jest błędem. Natomiast przenoszenie ich do następnej linijki za pomocą **wymuszonego lamania wiersza**⁵ przyczynia się do powstawania w tekście zbyt dużych światel międzywyrazowych, tak zwanych dziur, które są jednym z największych grzechów typografii (ryc. 2).

Wstawianie **twardej spacji** po wyżej wymienionych znakach jest nieco uciążliwe – należy bowiem całą operację przeprowadzić osobno dla wielkich i małych

⁴ Kolejność „uwioza” jest niealfabetyczna ale nie przypadkowa – jest to najłatwiejszy do zapamiętania układ tych liter.

⁵ W Adobe InDesign: ^n, w Microsoft Word funkcjonuje pod nazwą **ręczny koniec wiersza: ^l**.

liter (z włączoną funkcją **Uwzględniaj wielkość liter**) czyli dwanaście razy. Obok wspomnianych pojedynczych znaków jest jeszcze wiele innych przypadków, gdzie powinny zostać użyte twarde spacje, na przykład: po inicjałach, tytułach naukowych, wewnątrz dat, między cyframi (2 524 569,05 zł) itd. Zastosowanie się do wszystkich reguł może być jednak bardzo uciążliwe, zwłaszcza jeśli uświadomimy sobie, że w przypadku cyfr jest to sto kombinacji, i bardzo często okazuje się, że próba automatycznej zamiany wszystkich, jakby mogło się wydawać, możliwych przypadków może trwać znacznie dłużej niż ich ręczna korekta.

Oddawana do rąk czytelników praca składa się z trzech części, jak się wydaje dobrze oddających specyfikę pracy uczelni teologicznej. Pierwszy rozdział jest zapisem wystąpienia na Międzynarodowym Sympozjum zatytułowanym: *Dialog jako konieczny model życia w społeczeństwie pluralistycznym i w Kościele*. Zostało ono zorganizowane w listopadzie 2009 roku. Problematyka dialogu jawi się w naszych czasach jako szczególnie ważna i aktualna. W społeczeństwie pluralistycznym poszukującym ciągle swoich nowych kształtów nie ma skuteczniejszej metody budowania jego jedności i spójności mimo indywidualnych różnic. W ten sposób te, co różni, staje się naszą uciegocenia, a nie musi dzielić. Postawa dialogu została także głęboko wpisana we współczesną historię naszego Kościoła już przez Sobór Watykański II, a jeszcze mocniej podkreślona w czasie długiego pontyfikatu papieża Polaka, dzisiaj już błogosławionego Jana Pawła II.

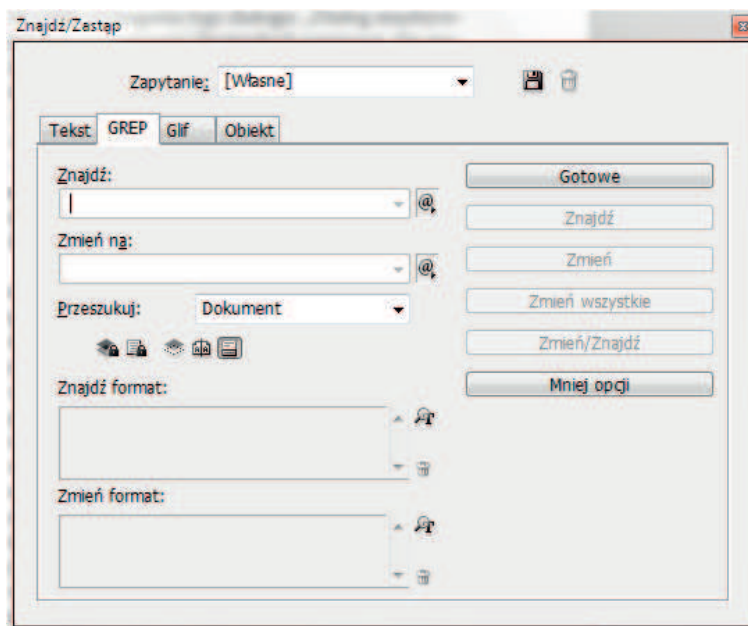
Ryc. 2. Przykład zbyt dużych odstępów w tekście

Naprzeciw oczekiwaniom użytkowników wychodzą producenci oprogramowania, doskonaląc swoje produkty. Jednak postęp jest zdecydowanie zbyt wolny i nie nadąża za wciąż rosnącymi potrzebami nowoczesnej typografii.

W najnowszych wersjach programów do składu i edycji tekstu pojawiają się nowe możliwości, pozwalające na szybsze i łatwiejsze wyszukiwanie oraz zamianę tekstu na poprawny. I tak na przykład w programie Adobe InDesign narzędzie **Znajdź/Zamień** zostało rozbudowane o wyszukiwanie wyrażen **GREP**.

GREP (ang. *Global Regular Expression Print* – drukowanie globalnych wyrażen regularnych) to program składowy systemu operacyjnego UNIX autorstwa Kena Thompsona. Służy on do wyszukiwania powtarzających się wyrażen regularnych. Używając go, możemy odnaleźć w tekście na przykład wszystkie wyrazy zaczynające się wielką literą „A”, dowolną cyfrę lub kombinację znaków i cyfr. Stosując operatory, lokalizatory i wieloznaczniki możemy także wymienić jakiś rodzaj tekstu (nie konkretny tekst, jak ma to miejsce w przypadku klasycznego narzędzia **Znajdź/Zamień**) na inny. Zadajemy na przykład wyszukanie wszystkich wyrażen w cudzysłowie, a następnie zamieniamy je na te same wyrażenia pisane kursywą i jednocześnie usuwamy cudzysłowy. Niestety akurat ta funkcja, choć podawana w instrukcji jako przykład stosowania omawianego narzędzia, nie działa – InDesign ma problemy ze znajdowaniem cudzysłowów w języku polskim. Można też tym sposobem wykonać inne żmudne, a zarazem całkiem skom-

plikowane operacje, zwłaszcza jeśli połączymy samą operację znajdowania i zamieniania z wykorzystaniem stylów akapitowych i/lub znakowych, na przykład wymianę ręcznie wstawionych myślników w wyliczeniach na automatyczne.



Ryc. 3. Narzędzie **Znajdź/Zastąp** w Adobe InDesign CS 5

Jak działa GREP? Mechanizm wyszukiwania i zastępowania znalezione-go tekstu jest w zasadzie taki sam jak tradycyjny **Znajdź/Zastąp** – różnica tkwi w składnikach, których szukamy. Używając tradycyjnych mechanizmów, możemy odnaleźć konkretny tekst, na przykład kombinację znaków **·u°** (spacja **u** spacja twarda)⁶. Czyli, jak już wspominałem, znalezienie wszystkich pojedynczych znaków literowych występujących w języku polskim wymaga powtórzenia tej czynności aż dwanaście razy:

·u· zamień na ·u°	·U· zamień na ·U°
·w· zamień na ·w°	·W· zamień na ·W°
·i· zamień na ·i°	·I· zamień na ·I°
·o· zamień na ·o°	·O· zamień na ·O°
·z· zamień na ·z°	·Z· zamień na ·Z°
·a· zamień na ·a°	·A· zamień na ·A°

⁶ W niniejszym artykule będę używał oznaczeń: · – spacja, ° – spacja twarda, nie są one zgodne z zapisem GREP, lecz ułatwi to lekturę artykułu.

Używając narzędzia GREGP, jesteśmy w stanie uzyskać ten sam efekt, stosując jedno wyszukiwanie: $(\cdot(\backslash\mathbf{u}))\cdot$ zamień na $\$1^\circ$ – czyli znajdź **grupe** (oznaczoną przez nawias) składającą się ze **spacji** oraz **klasy wyrażen** zdefiniowanych przez użytkownika (oznaczonej przez nawias kwadratowy), do klasy należą: każda mała litera (\mathbf{l}) i każda wielka litera (\mathbf{u}), następnie poza grupą znajduje się **spacja**. Znalezione wyrażenie zamieniamy na kombinację **znaleziona grupa** (cyfra po symbolu dolara oznacza numer znalezionej grupy) i **spacja twarda**. I właśnie w tej operacji kryje się zarówno największa tajemnica, jak i zaleta tego narzędzia – jest ono w stanie zwrócić wynik wyszukiwania każdorazowo inny, tj. odpowiedni do tego, co zostało znalezione: jeśli odnaleziona zostanie kombinacja $\cdot\mathbf{a}\cdot$ to wynikiem zamiany będzie $\cdot\mathbf{a}^\circ$, a jeśli komputer odnajdzie $\cdot\mathbf{Z}\cdot$ to zamieni je na $\cdot\mathbf{Z}^\circ$.

Kolejne przykłady zobrazują potencjał tego narzędzia – jeśli w przypadku **uwioza** tradycyjnym sposobem wyszukiwania czynność trzeba powtórzyć dwaście razy to w wyrażeniu złożonym z długich ciągów cyfrowych jest to znacznie większa liczba operacji. Zgodnie z zasadami zapisu długich liczb należy wstawiać spację co trzy cyfry, na przykład **132·435·657,00·zł**, czyli należałoby odnaleźć wszystkie kombinacje **1·0, 1·1, 1·2, 1·3, ..., 2·0, 2·1, 2·2, ...** aż do **9·0, 9·1, 9·2** itd. Razem daje to sto kombinacji, które możemy zastąpić jednym wyrażeniem GREGP: $(\mathbf{d})\cdot(\backslash\mathbf{d})$ zamień na $\$1^\circ\2 , czyli kombinację składającą się z dwóch grup (w każdej **jakakolwiek cyfra**) rozdzielonych spacją, zwróć w postaci tych samych dwóch **grup** rozdzielonych **spacją twardą**. Jeśli w poddawanym obróbce tekście pojawiają się często cyfry w kontekście na przykład finansowym, możemy rozbudować wyszukiwaną frazę o jednostkę walutową: $(\backslash\mathbf{d})\cdot(\backslash\mathbf{dzl})$ zamień na $\$1^\circ\2 . Druga grupa została poszerzona tu o zdefiniowaną przez mnie klasę oznaczającą dowolną liczbę lub konkretne wyrażenie **zł**. Oznacza to, że komputer będzie wyszukiwał wyrażen złożonych zarówno z cyfr, jak i spacji oraz wyrażenia **zł**. W razie potrzeby można wyszukiwaną klasę rozszerzyć o kolejne wyrażenia zawierające na przykład symbole innych walut.

Kolejnym przypadkiem, w którym powinniśmy zastosować spację nierozdzielającą, jest styk inicjału imienia i nazwiska. Alfabet łaciński ma 26 znaków, przyjmijmy, że tekst, nad którym pracujemy, zawiera nazwiska z różnych części świata. Nie możemy więc pominąć liter takich jak **Y** czy **V** i ponadto musimy dodać polskie znaki dla imion jak **Żaneta, Łucja** czy **Światosław**. W tym przypadku wyszukiwanie tradycyjne musiałyby się odbyć aż w ok. 900 krokach. Nie miałyby to sensu nawet w encyklopedii, nie mówiąc o krótszych tekstach. I tu także przychodzi z pomocą GREGP, a konkretnie wyrażenie $(\backslash\mathbf{u})\cdot(\backslash\mathbf{u})$, czyli (**każda wielka litera kropka**) **spacja** (**każda wielka litera**) zamień na $\$1^\circ\2 , czyli **znaleziona grupa pierwsza, spacja twarda, znaleziona grupa druga**. **J·Kowalski** zostaje zastąpione przez **J^oKowalski**. W tym miejscu warto zwrócić uwagę na niedoskonałości tej metody – skutkuje ona tylko w przypadku pojedynczych inicjałów. Aby poradzić sobie z osobami używającymi dwóch lub trzech imion, na przykład

J. R. R. Tolkien czy G. R. R. Martin, należałoby napisać bardziej skomplikowane wyrażenie lub po prostu przeprowadzić całą procedurę trzy razy⁷. W tym miejscu należy zaznaczyć, że w obecnej chwili żaden program komputerowy nie jest w stanie poprawnie zinterpretować tekstu, więc przy przeprowadzaniu powyższej procedury mogą zaistnieć błędy. Przeanalizujmy poniższy tekst, wymyślony na potrzebę tego przykładu:

Po kolacji razem z Adamem poszliśmy na partyjkę brydża do pana J. Mieszkał on w pobliskiej willi.

Fragment „J. Mieszkał” zostanie mylnie zinterpretowany jako inicjał z nazwiskiem. Błąd, który w tym miejscu powstanie, będzie dla czytelnika niewidoczny i najprawdopodobniej nie wpłynie na układ tekstu na stronie. Niemniej istnieje realne zagrożenie powstania w tekście dziur, zwłaszcza jeśli autor często w swoim utworze stosowałby tego typu rozwiązania, a książkę wydano by w małym formacie z wąską kolumną⁸.

Wśród najczęstszych błędów⁹, które powstają podczas tworzenia tekstów przez osobę nieznającą wszystkich zasad typografii i tajemnic składu, pojawiają się:

- ·· wielokrotne spacje,
- ·, spacja przed przecinkiem,
- ·. spacja przed kropką,
- ·; spacja przed średnikiem,
- (· spacja po otwarciu nawiasu lub przed zamknięciem ·),
- dywizy między cyframi zamiast półpauz: **s. 65–89** zamiast **s. 65–89**,
- brak spacji w lokalizacjach **s.43**,
- brak spacji w datach: **1989r.**,
- brak spacji po inicjałach lub między nimi: **J.R.R.Tolkien**¹⁰.

Każdy z nich można wyeliminować przy pomocy bardzo prostych wyrażeń GREP lub nawet używając tradycyjnego narzędzia **Znajdź/Zastąp**. Powyższe

⁷ Do chwili oddania do druku niniejszego artykułu nie udało mi się stworzyć wyrażenia, które rozwiązałyby problem przy jednorazowym wyszukiwaniu – nie wykluczam, że przy obecnych ograniczeniach GREP jest to niemożliwe.

⁸ Więcej o możliwych błędach i zagrożeniach spowodowanych stosowaniem GREP w dalszej części artykułu.

⁹ Jest ich oczywiście znacznie więcej, lecz to obszerny temat na osobne badania.

¹⁰ Kwestia rozdzielania inicjałów spacjami jest dyskusyjna – zarówno te ze spacjami, jak i bez spacji mają swoich zagorzałych zwolenników. Autor niniejszego artykułu przychyła się do rozdzielania inicjałów spacjami chudymi, które z jednej strony rozdzielają kolejne człony, niemniej są węższe niż zwykle, rozciągliwe spacje i wskazują, że inicjały wraz z nazwiskiem stanowią całość. W tym przypadku udaje się zachować wartość estetyczną oraz merytoryczną.

przykłady prezentują dość proste błędy, przyjrzyjmy się zatem przypadkom bardziej skomplikowanym:

- numeracja przy imionach monarchów **Ludwik XVI**,
- daty: **XIX w., 1920 r., w XX wieku, w roku 1987**,
- daty: **15 sierpnia 1677, 13 X 1555**,
- tytuły przed nazwiskami: **prof. J. Kowalski, płk B. Nowak, mec. Z. Janiak**,
- cyfry i jednostki wielkości fizycznych: **10 m, 25 m², 19° C, 25 h**,
- działania matematyczne: **2 + x = 4**,
- zapisy bibliograficzne: **s. 11, t. I, roz. 13, § 4**,
- adresy: **ul. Piotrkowska 30, al. Kościuszki 65**.

We wszystkich tych wyrażeniach należy stosować odpowiednio do konkretnego przypadku spacje (twarde, chude, cyfrowe). Czy da się napisać wyrażenia uniwersalne, pozwalające na poprawienie wszystkich przypadków? Na tym etapie badań nie sposób odpowiedzieć na powyższe pytanie. Z pewnością można wyeliminować część z nich. Dość prosty jest przypadek dat złożonych z cyfr rzymskich i arabskich, na przykład **27 XI 2014**.

Wyrażenie do wyszukiwania tego typu dat może wyglądać następująco¹¹:

$$(\backslash d^+) \cdot ([XVI]^+) \cdot (\backslash d^+ \backslash d^+ \backslash d^+)$$

czyli **grupa** składająca się z **jakichkolwiek cyfr** arabskich (+ oznacza tu dowolną wielokrotność poprzedzających go wieloznaczników) **spacja**, **grupa** składająca się z **klasy** zawierającej **XVI** czyli wszystkie rzymskie oznaczenia stosowane do zapisu miesięcy oraz + determinujący szukanie wyrażen złożonych z wielokrotnie występujących znaków (na przykład w wyrażeniach typu XIX czy VIII), **spacja**, **grupa** złożona z czterech cyfr arabskich.

Wyrażenie zwrotne:

$$\$1^\circ \$2^\circ \$3$$

czyli **znaleziona grupa pierwsza**, **spacja twarda**, **znaleziona grupa druga**, **spacja** i **znaleziona grupa trzecia**.

¹¹ Do budowy wyrażen wyszukiwania można użyć wielu różnych składników GREGP. Alternatywne wyrażenie wyszukiwania w tym przypadku może też wyglądać następująco: $(?<=\backslash d^+) \cdot ([XVI]^+) \cdot (?=\backslash d^+ \backslash d^+ \backslash d^+)$, czyli znajdź **spacja**, **grupa** składająca się z **klasy** zawierającej **XVI** występujące **raz lub więcej**, **spacja**, poprzedzone (funkcja dopasowania **Dodatnie wybieganie wstecz**) jakikolwiek cyframi występującymi **raz lub więcej**, i po którym to wyrażeniu występują cztery cyfry (funkcja dopasowania **Dodatnie wybieganie w przód**), zamień na: $^\circ \$1^\circ$ czyli **spacja twarda**, **znaleziona grupa pierwsza**, **spacja twarda**.

Podobnie jest z datami zawierającymi słowne określenie miesiąca. W tym przypadku w wyrażeniu wyszukiwanym należy podać **alternatywę** zawierającą nazwy wszystkich miesięcy w dopełniaczu liczby pojedynczej: **stycznia|lutego|marca** itd. Można także zaryzykować budowę wyrażenia bardziej ogólnego: **\w+**, **wieloznacznika** określającego **każdy znak mogący wystąpić w słowie**. W tym drugim przypadku należałoby się upewnić, czy wszystkie wyrażenia typu **cyfra lub więcej cyfr, jeden lub więcej znaków występujących w słowie, cztery cyfry** odnoszą się tylko do dat¹². Stosowanie zbyt ogólnych wyrażen zawsze niesie ze sobą niebezpieczeństwo wprowadzenia do tekstu błędów, których wcześniej nie było. Z drugiej strony wyrażenia zbyt szczegółowe zmuszają nas do stworzenia dużej ich liczby, aby objąć wszystkie przypadki wyszukiwań.

Prostotą i łatwością określenia większości przypadków charakteryzują się wielkości fizyczne, elementy zapisów bibliograficznych i części wzorów matematycznych¹³. Dość łatwo stworzyć katalog zawierający oznaczenia stron, tomów, rozdziałów itp. oraz metrów, amperów, godzin, litrów itd. Dla jednostek miary i powierzchni wyrażenie wyszukiwane mogłoby wyglądać następująco:

(?<=\d) (m|m2|m3|cm|cm2|cm3|dm|dm2|dm3|km|km2)\>¹⁴.

Dla wagi:

(?<=\d) (mg|g|dag|kg|t)\>¹⁵.

Dla zapisów bibliograficznych:

(?<=s\.|t\.|roz\.|§) (\d).

Należy zwrócić uwagę na wyrażenia zakończone lokalizatorem **\>** oznaczającym **koniec słowa**. Jest to zabezpieczenie przed słowami zaczynającymi się na literę taką jak w skrócie, na przykład:

W roku 1989 tramwaje stały się bardziej popularne niż autobusy.

¹² W tym przypadku także należałoby przeprowadzić badania nad tego typu wyrażeniami – czy istnieją inne o podobnej budowie, lecz innej funkcji i znaczeniu.

¹³ Wzory skomplikowane przygotowywane są w specjalnych edytorach i importowane do składu jako pliki graficzne, nie ma więc potrzeby ich analizować.

¹⁴ Dla zachowania porządku proponuję stworzyć osobne wyrażenia dla różnego rodzaju kategorii, np. dla jednostek miar, objętości, czasu itd.

¹⁵ W przykładach używam tylko tych najpopularniejszych jednostek, które mogą pojawić się w większości tekstów. Wysiłek włożony w przygotowanie pełnych zestawień używanych w specjalistycznych tekstach jednostek najprawdopodobniej okazałby się zbędny, nie napotykały bowiem często na przykład na joktogramy czy zettagramy.

Fragment „9 t” bez lokalizatora **koniec słowa** zostałby zinterpretowany jako dziewięć ton.

Podobne wyrażenia można stworzyć dla części tytułów i nazwisk, na przykład:

- dr Jan Kowalski
- mgr Zbigniew Nowak
- prof. dr hab. Jacek Kowalski

Jednak skomplikowany system tytułatur naukowych nie pozwala na stworzenie uniwersalnego wyrażenia, gdyż możliwych kombinacji jest zbyt wiele, a niekiedy wręcz nie jest możliwe określenie wszystkich przypadków:

- prof. dr hab. ks. Waldemar Nowak,
- prof. dr hab. płk Wacław Kowalski,
- dr n. med. mjr Lucjan Kowalski,
- dr hab. n. med. gen. Wojciech Nowak prof. UJ,
- dr hab. ks. Antoni Nowak prof. UKSW,
- prof. zw. dr hab. n. med. Czesław Kowalski.

Podobnie jest w przypadku adresów, w których nie tylko człon określający rodzaj lokalizacji (ulica, plac, aleja, osiedle itp.) nie powinien być rozdzielany od nazwy, ale także numer posesji i lokalu czy mieszkania. Nie ma problemu z prostymi adresami:

- ul. Piotrkowska 104,
- ul. Łąkowa 20 m. 1,
- al. Kościuszki 65.

Komplikacje jednak mogą wystąpić z powodu różnego typu zapisu adresów w każdej miejscowości i elementów mogących się w zapisie pojawić:

- **ul. Św. Teresy od Dzieciątka Jezus 15b** (Łódź),
- **ul. Generała Leopolda Okulickiego „Niedźwiadka” 10** (Łódź),
- **ul. Generała Józefa Bema 40** (Łódź),
- **ul. Gen. J. Bema 20** (Ozorków),
- **ul. Józefa Bema** (Warszawa),
- **ul. Bema** (Chorzów)¹⁶,
- **al. Ofiar Terroryzmu 11 września** (Łódź)¹⁷.

Jak widać, w powyższych przykładach nie ma możliwości precyzyjnego oznaczenia końca adresu. Nie można oprzeć się na liczbie słów, nie można uznać, że wszystkie człony nazwy zaczynają się wielką literą. Końca adresu nie oznacza

¹⁶ Nie udało się ustalić, czy w tym przypadku chodzi także o gen. Józefa Bema,

¹⁷ Do czasu ukończenia prac nad artykułem w danej lokalizacji nie było żadnej posesji oznaczonej numerem, nie znaczy to jednak, że sytuacja ta nie uległa zmianie.

kropka czy nawet cyfra. Ani treść, ani forma nie dają możliwości napisania wyrażenia obejmującego wszystkie możliwości, nawet jeśli użylibyśmy kilku wyrażzeń wyszukiwania. Czy warto zatem stosować takie wyszukiwanie? Jeśli nawet nie uda nam się wyszukać wszystkich adresów czy osób z tytułami w tekście, to i tak warto zastosować wspomniane procedury, aby wyręczyć redaktora lub korektora choć w części tego żmudnego zajęcia.

Kolejnym zadaniem, do którego można wykorzystać GREP, jest układanie tekstu w odpowiedniej kolejności. Zademonstruję to na przykładzie bibliografii. Niejednokrotnie w tekstach naukowych stosowany jest system lokacji cytatów i źródeł przez podanie ich w przypisach dolnych, na przykład:

- V. Boland, *Conditionson talks threaten to offend Turks*, „Financial Times”, 7 października 2004, s. 11.
- W. Cimoszewicz, *Polska wizja Europy*, „Rzeczpospolita”, 19 lutego 2002, s. 7–8.
- J. Bielecki, *Traktat z Nicei zagrożony*, „Rzeczpospolita”, 2–3 lutego 2002, s. 3.
- M. Czech, *Duch Jagiellonów nie błądzi po Polsce*, „Gazeta Wyborcza”, 8 września 2009, s. 5–6.
- C. Chatelot, *La Pologne part l’assaut des futures institutions europeennes*, „Le Monde”, 3 października 2003, s. 45.
- M. A. Cichocki, *Ile Boga w Unii?*, „Więź”, nr 8–9, 2002, s. 13.

Najłatwiejszym sposobem sporządzenia bibliografii jest skopiowanie przypisów i sporządzenie z nich listy¹⁸. Taka lista musi zostać jednak przetworzona: należy usunąć lokalizację (numery stron, tomów itp.), ułożyć pozycje alfabetycznie i zamienić kolejność inicjałów imienia i nazwiska. Przy listach liczących kilkaset pozycji bardzo spowalnia to pracę. Jak zatem wykorzystać GREP? Pierwsza procedura to usunięcie niepotrzebnych elementów. Należy znaleźć wszystkie kombinacje **s. spacja**¹⁹ **jakakolwiek cyfra lub wiele cyfr**, następnie **t. spacja klasa** zawierająca litery używane do rzymskiej numeracji itp. i zamienić na **nic**. Następnie należy wykorzystać kolejną zaletę tego narzędzia, jaką jest możliwość dowolnego układania znalezionych grup. W omawianym przypadku wyrażenie wyszukiwania może wyglądać następująco:

`\<(\u\.) (\u[-\w]+)`

¹⁸ Bardzo często prace nadsyłane przez autorów nie posiadają bibliografii i trudno skłonić ich do jej sporządzenia, a terminy wydania wymagają szybkiego składu.

¹⁹ W tym przypadku należy raczej wstawić symbol `\s` oznaczający jakikolwiek odstęp, ponieważ jeśli wcześniej porządkowaliśmy tekst, to w tym miejscu może znajdować się spacja twar-da lub cyfrowa.

czyli znajdź **na początku akapitu grupę** składającą się z **wielkiej litery, spacji, grupę** składającą się z **wielkiej litery, jednego lub więcej dowolnych znaków występujących w słowie**²⁰ i zamień na:

\$2°\$1

czyli **znalezioną grupę drugą, spację twardą**²¹ i **znalezioną grupę pierwszą**. Efektem naszych poczynań będzie lista:

- Boland V., *Conditionson talks threaten to offend Turks*, „Financial Times”, 7 października 2004.
- Cimoszewicz W., *Polska wizja Europy*, „Rzeczpospolita”, 19 lutego 2002.
- Bielecki J., *Unia otwarta dla kandydatów, Polska po szczycie w Nicei*, „Rzeczpospolita”, 12 grudnia 2000.
- Czech M., *Duch Jagiellonów nie błądzi po Polsce*, „Gazeta Wyborcza”, 8 września 2009.
- Chatelot C., *La Pologne part l’assaut des futures institutions europeennes*, „Le Monde”, 3 października 2003.
- M. Cichocki A., *Ile Boga w Unii?*, „Więź”, nr 8–9, 2002.

Jedyny problem powstał w ostatniej pozycji na liście – autor używa dwóch imion i przeniesione na właściwe miejsce zostało tylko drugie. Nie da się skonstruować takiego wyrażenia GREP, które poradziłoby sobie z przypadkami o różnej liczbie imion; inicjałów może być więcej jak w przypadku wspomnianego już J. R. R. Tolkiena. Należy wtedy napisać kilka wyrażen (trzy w zupełności wystarczą) i uruchomić je w odpowiedniej kolejności. W ten sposób można uporządkować całą bibliografię. Ostatnią czynnością jest ułożenie pozycji w kolejności alfabetycznej – do tego celu można wykorzystać zamianę tekstu na tabelę i funkcję sortowania²².

Każda broń może być obosieczna. Dlatego w tym przypadku należy bardzo ostrożnie obchodzić się z tekstem i nie próbować naprawiać wszystkich błędów, jakie się napotka bez rozważenia konsekwencji płynących z przeprowadzenia danej procedury. W niektórych przypadkach istotna jest kolejność, w jakiej będziemy usuwać błędy, na przykład wstawianie spacji po dwukropku nie może być poprze-

²⁰ W tym przypadku znak \w został wstawiony do klasy razem z dywizem – ponieważ nieuwzględnienie dywizu wyeliminowałoby z wyszukiwania dwuczłonowe nazwiska np. Kowalska-Nowak. Zob. M. Włodarz, *GREP w InDesign CS3–CS5*, MAWart, Warszawa 2011, s. 58.

²¹ W tym przypadku spacja twarda nie jest konieczna, mało prawdopodobne jest, aby trzeba było łamać wiersz w tym miejscu, można jednak wyobrazić sobie przypadek, kiedy autorów artykułu jest wielu, a skład na wąskiej kolumnie.

²² Nie da się tego zrobić przy pomocy narzędzia GREP ani w programie InDesign (chyba, że wykupimy plug-in, czyli wtyczkę do sortowania; takie wtyczki tworzone są przez niezależne od Adobe firmy). Należy wyeksportować tabelę np. do MS Word i tam posortować.

dzone usuwaniem spacji przed zamknięciem nawiasu, gdyż taka kolejność spowodowałaby powstanie błędu ([w:]). Podobny problem powstaje, gdy sąsiadują ze sobą otwarcie cudzysłowu i otwarcie nawiasu na przykład „(oraz zamknięcie nawiasu i zamknięcie cudzysłowu na przykład)”. Wstawienie spacji przed otwarciem nawiasu jest w tym miejscu błędem, choć w innych przypadkach jest obowiązkowe. Kolejne przykłady: spacje po ! lub ? nie powinny pojawiać się przed zamknięciem nawiasów ! ?) i cudzysłowów !” ?”. Bardzo problematyczny jest znak elipsy ... – może on występować w otoczeniu spacji w różnych konfiguracjach:

- – *Miałem powiedzieć... – Za drzwiami dało się słyszeć hałas – ...ana nie przyjdzie.*
- *Jan próbował przekrzyknąć wicherę. – Musisz ... dźwignię i mocno ... a potem...*
- Jako znak opuszczenia fragmentu tekstu w cytacie [...].

Te kwestie musi indywidualnie rozstrzygnąć redaktor. Podobnie jak znaleźć odpowiedź na pytanie, z której strony postawić twardą spację w sytuacjach zdemontrowanych w przykładach:

- **Terminator 2 to znany film. Pokażą go w telewizji 2 lutego,**
- **w XX wieku, w podeszłym wieku.**

Kolejnym problemem są ograniczenia narzędzia, których nie da się przewidzieć, na przykład w odmianie nazwisk obcojęzycznych zakończonych na samogłoskę, na przykład **Nicholasa Stonebridge’a**. Apostrof przed **a** powoduje, że mechanizm wyszukiwania traktuje tę literę jako osobną i wstawia po niej twardą spację. Ten błąd należy skorygować, gdyż przeniesienie długiego nazwiska do następnej linijki może spowodować powstanie nieproporcjonalnie dużych odstępów w całym akapicie. Możliwe są także przypadki, kiedy występuje nagromadzenie twardych spacji w jednym miejscu, na przykład **Nicholasa Stonebridge’a i George’a Fridricha**. Twarde spacje w tym przypadku zostają wstawione po **Stonebridge’a, i** oraz **George’a** co w przypadku, kiedy nie pozwalamy na dzielenie nazwisk, sprawia, że powstaje długi ciąg wyrazów bez możliwości złamania wiersza²³.

Oprócz likwidacji błędów i porządkowania bibliografii mechanizm wyszukiwania GREP może zostać wykorzystany jako pomoc przy składzie tekstu. Jeśli połączymy jego funkcje z możliwością formatowania tekstu przy pomocy stylów akapitowych i znakowych, otrzymamy wielofunkcyjne narzędzie, którego wszystkich możliwości nie sposób omówić. Marek Włodarz podaje receptę na sformatowanie spisu treści w sposób znacznie ciekawszy, niż pozwalają na to narzędzia

²³ Należy także zwrócić uwagę na wersję oprogramowania i system operacyjny, w którym pracujemy, gdyż wyrażenie prawidłowo działające pod Windows w InDesign CS6 wcale nie musi działać tak samo w MacOS w Indesignie CS3, zob. M. Włodarz, dz. cyt., s. 28.

wbudowane w program InDesign²⁴. Ja używam kombinacji GREP i stylów między innymi do tworzenia spisów treści w tomikach wierszy zawierających utwory bez tytułu. Ręczne wyszukiwanie i zbieranie incypitów jest zajęciem żmudnym i czasochłonnym, a dzięki zastosowaniu automatyzacji można je skrócić do kilkunastu sekund. Istnieje także wiele narzędzi dostępnych w Internecie²⁵, które pozwalają na łączenie wyrażeń wyszukiwania i wyrażeń zwrotnych w łańcuchy lub wręcz umożliwiają pisanie skryptów automatyzujących wiele działań. Niejednokrotnie okazuje się jednak, że nie wszystkie pasują do naszych potrzeb lub nie są dostosowane do zasad i specyfiki języka polskiego (na przykład nieuwzględnianie polskich znaków czy stosowanie innych cudzysłowów). Dalszy rozwój GREP i jego integracja ze stylami mogą zaowocować znaczącym postępem w dziedzinie pracy z tekstem.

Czy GREP może pomóc w redagowaniu i składaniu tekstu? Niewątpliwie tak, mimo że narzędzie to ma liczne ograniczenia i wady, jego używanie znacznie przyspiesza i ułatwia pracę. Należy jednak zaznaczyć, że bardzo łatwo popełnić błąd przy tworzeniu wyrażeń wyszukiwania i zwrotnych. Dlatego każdy tekst po zastosowaniu wyrażeń GREP powinien zostać przeczytany przez redaktora. Nie sposób przewidzieć wszystkich konsekwencji automatycznej zamiany lub formatowania tekstu. Automatyczne wyszukiwanie zakłada powtarzalność – jeśli autor popełnia zawsze te same błędy, wówczas można je wyeliminować bez trudu. Największym wrogiem tej metody jest chaos. Różni autorzy popełniają różne rodzaje błędów, dlatego nie można stosować wielu czynności automatycznych bez wcześniejszego poznania tekstów. Ponadto każdy język zawiera wieloznaczności, które tylko człowiek jest w stanie zrozumieć, a nawet najlepsze oprogramowanie nie nadaży za ludzką wyobraźnię i kreatywnością. Złe zastosowanie wyrażeń GREP może prowadzić w tekście do powstania **ko(s)micznych**²⁶ błędów.

Bibliografia

Przedmiotowa

Włodarz Marek, *GREP w InDesign CS3–CS5*, MAWart, Warszawa 2011.

²⁴ Zob. tamże, s. 67–69.

²⁵ Istnieje wiele stron i list dyskusyjnych poświęconych GREP. Można ściągnąć darmowe i płatne narzędzia przeznaczone do najróżniejszych celów.

²⁶ Wstawienie w tym miejscu spacji przed otwarciem i po zamknięciu nawiasu byłoby błędem, ale może go dostrzec tylko człowiek.

Mateusz Poradecki

GREP in Editors' Work – Introduction to the subject

(Summary)

The paper considers the use of the Global Regular Expression Print tool in the Adobe InDesign application. The paper shows the results of the research on the utilisation of this tool in editors' work. It presents examples of common mistakes which can be automatically corrected using the Find/Change tool with the GREP function. The paper also indicates threats connected with making new mistakes when the tool is used without extreme care or in a wrong order. The conclusion is that the GREP tool cannot be used without human supervision as languages are too complicated and not all situations can be predicted.

Keywords: automatization of editors' work; typesetting; Global Regular Expression Print

Słowa kluczowe: automatyzacja pracy redaktora; skład i łamanie tekstu; Global Regular Expression Print