

Jan Rusinek

PROGRAMOWANIE HYBRYDOWE. PRZYKŁAD
ZASTOSOWANIA W STATYSTYCE, EKONOMETRII I
DYDAKTYCE

[**słowa kluczowe:** algorytmy, informatyzacja procesu dydaktycznego, programowanie hybrydowe, pascal, TeX, statystyka, ekonometria, arkusz kalkulacyjny]

Streszczenie

W pracy omawiane jest zastosowanie programowania hybrydowego wykorzystującego modyfikacje pewnych wcześniej napisanych programów w różnych językach do uzyskiwania i przetwarzania wielu danych statystycznych i ekonometrycznych pozyskanych z ankiet studenckich oceniających wykładowców.

1. Wstęp

Przez programowanie hybrydowe rozumiemy przygotowanie programu albo algorytmu z użyciem dwóch lub więcej języków programowania.

Najczęstsza sytuacja, kiedy używamy programowania hybrydowego to taka, kiedy jeden fragment algorytmu jest dostępny tylko w jednym języku, a drugi tylko w innym. Z taką sytuacją będziemy mieli do czynienia np. kiedy odpowiedni program graficzny (np. TeX) potrafi złożyć poprawnie wzór matematyczny, ale nie potrafi dokonać odpowiednich obliczeń, natomiast typowy język programowania bez problemu wykonuje skomplikowane obliczenia, a zmuszenie go do efektów graficznych jest trudne lub wymaga dużego wysiłku.

Bywa też tak, że w każdym z języków można wykonać dany projekt w całości, ale rozdzielenie go na różne oszczędza pracę. Tak może się zdarzyć kiedy na przykład mamy już gotowe algorytmy lub ich fragmenty napisane w różnych językach i tłumaczenie ich na jeden wspólny język zabrałoby sporo czasu (nie mówiąc już o możliwości zrobienia w trakcie tego tłumaczenia błędów).

W pracy zaprezentowano właśnie taką sytuację. Autor miał gotowe programy - część z nich w TeX-u, a część utworzonych wiele lat temu w pascalu. Drobna modyfikacja tych programów i odpowiednie ich połączenie pozwoliło **szybko** uzyskać zamierzony cel.

2. Sformułowanie celu

Kilka lat temu został napisany przez autora w TeX-u program do tworzenia i obróbki ankiet studenckich oceniających wykładowców.¹ Jest on opisany w [5].² Program ten dzięki komputerowej obróbce wyników jako „produkt uboczny” daje pliki z pełną informacją jak kolejni studenci oceniali kolejnych wykładowców, czyli odpowiadali na poszczególne pytania. Pytania te były następujące:

1. **Klarowność prowadzenia zajęć;**
2. **Życzliwość wobec studentów;**
3. **Przejrzystość zasad zaliczania;**
4. **Atrakcyjność zajęć;**
5. **Przydatność materiałów dydaktycznych.**

Na potrzeby ankiety program wyliczał tylko średnią ocenę danego wykładowcy w danym pytaniu oraz globalną średnią danego pytania i danej grupy, aby oceniany wykładowca mógł się zorientować jak wypada w porównaniu z innymi wykładowcami.

Pojawiła się idea, aby wykorzystać zebrane pełne informacje i obliczyć inne parametry statystyczne np. mediany, odchylenia standardowe, korelacje pomiędzy poszczególnymi pytaniami, korelacje pomiędzy wykładowcami itp., a nawet pewne związki ekonometryczne.

Posiadanie tych danych w plikach sugerowało, że można stworzyć narzędzie do automatycznego uzyskiwania tych parametrów.

Użycie do tego wyłącznie zmodyfikowanego wymienionego programu napisanego w TeX-u nie było zbyt perspektywiczne – Tex niezbyt dobrze radzi sobie z obliczeniami numerycznymi.³

Ponadto autor miał już gotowe programiki napisane w pascalu wiele lat temu razem ze studentami w ramach jakichś zajęć dydaktycznych. Programiki te wyliczały większość podstawowych parametrów statystycznych wykorzystując dane podane w prosty sposób, a mianowicie jako kolejne linie pliku

¹Ankiety takie są obligatoryjne od wielu lat na wszystkich uczelniach.

²Opisy innych zastosowań podobnych algorytmów w dydaktyce można znaleźć w [3] i [4].

³Są wprawdzie pewne makra wspomagające obliczenia, ale są wysoce niedoskonałe.

tekstowego lub dwóch plików tekstowych w wypadku parametrów opisujących współzależność.

Okazało się, że wystarczyło odpowiednio zmodyfikować program TeX-a tak, aby tworzył te pliki z danymi i wpleść właściwe programy „pascalowe” w strukturę głównego programu, aby otrzymując pełną automatyzację otrzymywania potrzebnych wyników statystycznych.

3. Plik generowany przez skaner i jego wstępna obróbka

Poniżej przedstawiamy fragment pliku generowanego przez skaner zawierającego pełne informacje oceny wykładowców w danej grupie

```
Nr identyfikacyjny;1013
Zestaw;11
Pytanie 1;A
Pytanie 2;A
Pytanie 3;A
Pytanie 4;A
Pytanie 5;A
Pytanie 6;A
Pytanie 7;A
Pytanie 8;A
Pytanie 9;A
Pytanie 10;A
Pytanie 11;A
Pytanie 12; B
Pytanie 13; B
Pytanie 14; B
Pytanie 15; C
Pytanie 16;
Pytanie 17;
Pytanie 18;
Pytanie 19;
Pytanie 20;
Pytanie 21; C
Pytanie 22; C
Pytanie 23; C
Pytanie 24; C
Pytanie 25; C
...
```

Numer identyfikacyjny jest nieistotny i nie jest wcale „identyfikacyjny” oprogramowanie skanera wymaga podanie jakiegokolwiek numeru. Zestaw 11 oznacza, że użyty został pierwszy formularz odpowiedzi. Na formularzu jest 70 wierszy, co oznacza, że jeden formularz wystarczy dla oceny $70/5 = 14$ wykładowców. Jeśli w grupie jest ich więcej niż 14, trzeba użyć drugiego formularza, wtedy przypisany mu jest zestaw 22.

Na przykład linijka `Pytanie 2:A` oznacza, że pierwszy student pierwszemu wykładowcy za trzecie pytanie przyznał ocenę 6, linijka `Pytanie 12; B` oznacza, że trzeci student trzeciemu wykładowcy za 2 pytanie przyznał ocenę 5. Student nie oceniał wykładowcy numer 4 (pytania 16–20), a wykładowcy numer 5 (pytania 21–25) przyznał wszystkie oceny 4.

Taka postać pliku jest niezbyt dogodna dla TeX-a (ale również dla innych języków) do pobierania części danych w celu ich odpowiedniego przetwarzania.⁴ Dlatego w pierwszym kroku TeX został „zmuszony” do przetworzenia tego pliku do pliku bardzo dogodnego dla przetwarzania. Ma on nazwę `wynank.ttt` i jego analogiczny fragment będzie następujący:

```
\wyn{1}{1}{1}{1}{1}{6}
\wyn{1}{1}{2}{1}{2}{6}
\wyn{1}{1}{3}{1}{3}{6}
\wyn{1}{1}{4}{1}{4}{6}
\wyn{1}{1}{5}{1}{5}{6}
\wyn{1}{1}{6}{2}{1}{6}
\wyn{1}{1}{7}{2}{2}{6}
\wyn{1}{1}{8}{2}{3}{6}
\wyn{1}{1}{9}{2}{4}{6}
\wyn{1}{1}{10}{2}{5}{6}
\wyn{1}{1}{11}{3}{1}{6}
\wyn{1}{1}{12}{3}{2}{5}
\wyn{1}{1}{13}{3}{3}{5}
\wyn{1}{1}{14}{3}{4}{5}
\wyn{1}{1}{15}{3}{5}{4}
\wyn{1}{1}{16}{4}{1}{0}
\wyn{1}{1}{17}{4}{2}{0}
\wyn{1}{1}{18}{4}{3}{0}
\wyn{1}{1}{19}{4}{4}{0}
\wyn{1}{1}{20}{4}{5}{0}
\wyn{1}{1}{21}{5}{1}{4}
\wyn{1}{1}{22}{5}{2}{4}
\wyn{1}{1}{23}{5}{3}{4}
\wyn{1}{1}{24}{5}{4}{4}
\wyn{1}{1}{25}{5}{5}{4}
```

Polecenie `\wyn` jest poleceniem sześćoargumentowym. Pierwszy argument jest numerem formularza (1 lub 2), drugi numerem studenta (jest on przydzielany w zależności od kolejności czytania formularzy przez skaner), trzeci numerem linii w formularzu, czwarty numerem wykładowcy, piąty numerem pytania, szósty ocena wystawioną przez studenta. Liczba zero oznacza, że na to pytanie brak odpowiedzi.

⁴Użytkownicy skanera nie mają wpływu na formę pliku, skaner zakupuje się razem z oprogramowaniem niestety bez pliku źródłowego.

Programowanie hybrydowe. Przykład zastosowania...

A oto fragment algorytmu TeX-a, który wykonuje to przetworzenie.

```
%MACRO DO PRZETWARZANIA PLIKU GENEROWANEGO PRZEZ SKANER
%NA PLIK DOGODNY DLA TeX-a
\def\wczytujem{
\ifeof\wczytu\let\next=\relax
\else\read\wczytu to\twojalinia
\ifx\zestj\twojalinia
\advance\stj1\edef\stud{\the\stj}\fo=1\ile=0\else\fi
\ifx\zestd\twojalinia
\advance\std1\edef\stud{\the\std}\fo=2\ile=0\else\fi
\ifx\zestt\twojalinia
\advance\stt1\edef\stud{\the\stt}\fo=3\ile=0\else\fi
\setbox6=\hbox{\twojalinia}\dlu=\wd6\relax
\ifdim\dlu>1pt
\expandafter\ab\twojalinia\relax
}%
\ifx\cc\LLp\advance\liczbastud1\else
\ifx\cc\Lp{}
\else
\expandafter\ident\twojalinia;
\pile=0
\ifx\aa\AA
\pile=6
\else
\ifx\aa\BB
\pile=5
\else
\ifx\aa\CC
\pile=4
\else
\ifx\aa\DD
\pile=3
\else
\ifx\aa\EE
\pile=2
\else
\ifx\aa\FF
\pile=1
\else
\pile=0
\fi\fi\fi\fi\fi\fi
\advance\ile1
\abc=\ile\advance\abc4\abd=\abc\divide\abc5
\wykj=\abc\multiply\abc5\advance\abd1\advance\abd-\abc
\immediate\write7{\string\wyn{\the\fo}{\stud}
{\the\ile}{\the\wykj}{\the\abd}{\the\pile}}
\dopun
\fi\fi
```

```
\else\fi
\let\next=\wczytujem\fi
\next}
```

Taki plik jest bardzo łatwy do wykorzystania przez TeX-a w celu wyodrębnienia czy połączenia odpowiednich danych, co prezentujemy poniżej.

4. Programy wykonujące odpowiednie obliczenia statystyczne wplecione w TeX-a

Są to dwa pliki. Pierwszy o nazwie `pojed.exe` po uruchomieniu go z parametrem będącym nazwą pliku tekstowego (na użytek tego omawianego zagadnienia pliki mają mieć rozszerzenie `uuu`) czyta kolejne linie tego pliku jako dane liczbowe i oblicza podstawowe parametry związane z tymi danymi: średnią, medianę, kwartyle, klasyczny współczynnik asymetrii, kurtozę, odchylenie standardowe, odchylenie przeciętne (patrz np. [7]).

Z kolei programik `korej.exe` wykonuje obliczenia statystyczne związane z korelacją danych odwołując się do dwóch plików tekstowych.

Narzędziem łączącym te dwa języki jest polecenie `\write18{plik wykonywalny}`. TeX po napotkaniu tego polecenia zatrzymuje się, wychodzi „na zewnątrz” i uruchamia `plik wykonywalny`. Po zakończeniu działania tego pliku przechodzi dalej. Plik wykonywalny w trakcie pracy tworzy nowe pliki, które mogą być potem przez TeX-a wczytane i wykorzystane. Intensywnie używane jest też makro `ifthen.sty`([1]), pozwalające łatwiej tworzyć w TeX-u pętle i iteracje. Podstawowe polecenia TeX-a można znaleźć w [2].

W związku z tym w zależności od tego jakie dane chcemy rozpatrywać musimy utworzyć odpowiedni plik (lub pliki) tekstowy a następnie poddać go obróbce przez program `pojed.exe` lub `korej.exe`. Dla przykładu przypuśćmy, że chcemy zanalizować globalne dane dotyczące drugiego pytania. Wtedy z olbrzymiego pliku `wynank.ttt` trzeba wybrać te linie, w których 5 parametr jest równy dwa, zapisać do nowego pliku odpowiadający mu 6 parametr i nowo otrzymany plik (będzie on miał nazwę `1f2p.uuu` - czyli zawierał dane dotyczące 2 pytania) poddać obróbce przez program `pojed.exe` i otrzymać plik z potrzebnymi wynikami, wstawić ten plik do pdf-a z odpowiednią informacją i przejść do wyliczania następnych parametrów.

Odpowiadający tej operacji fragment algorytmu TeX-a wykonujący powyższe zadanie kolejno dla wszystkich pięciu pytań jest następujący:

```
\def\fp#1#2{\immediate\openout7=#1f#2p.uuu
\abc=#1\abd=#2%\abe=#3
\def\wyn##1##2##3##4##5##6{\ifnum\abc=##1
```

Programowanie hybrydowe. Przykład zastosowania...

```
\ifnum\abd=##5
\immediate\write7{##6}
\else\fi
\else\fi}
{}
\input{wynank.ttt}
\immediate\closeout7}
\ilew=0\ilepp=0\ilef=1
\whiledo{\ilepp<5}{\advance\ilepp1\fp1{\the\ilepp}}
```

Fragment pliku 1f2p.uuu odpowiadający pierwszym 5 wykładowcom i pierwszemu studentowi wygląda tak:

```
6
6
5
0
4
```

Plik otrzymany w efekcie obróbki całego pliku 1f2p.uuu przez program pojed.exe wygląda następująco⁵

```
\edef\liczbadan{300}
\edef\ilejed{0}
\edef\iledwa{5}
\edef\iletrzy{9}
\edef\ileczty{32}
\edef\ilepie{66}
\edef\ilesze{188}
\edef\sred{5.41}
\edef\med{5.70}
\edef\qjed{4.94}
\edef\qtrzy{6.10}
\edef\ocw{0.58}
\edef\ostan{0.91}
\edef\kws{-1.66}
\edef\kurt{2.40}
```

Fragmenty algorytmu wykorzystujące powyższy plik (kolejno dla wszystkich 5 pytań) i drukujące wyniki do pdf-a (lub dvi) są następujące:

```
\def\piszpojed{
liczba danych  $=$ \liczbadan $;$  jedynek -  $\$ ilejed $;$ 
```

⁵Mediana jest wyliczana tak, jakby ocena n ($n = 1, 2, 3, 4, 5, 6$) była daną z szeregu rozdzielczego z przedziału $(n - 1/2; n + 1/2)$.

```

dwójek -  $\backslash\ile{dwa}$ ; trójek -  $\backslash\ile{trzy}$ ;
czwórek -  $\backslash\ile{cztyry}$ ; piątek -  $\backslash\ile{piec}$ ; szóstek -  $\backslash\ile{szesz}$ ;
średnia  $\backslash\med$ ;
mediana  $\backslash\med$ ;  $\backslash\Q_1=\backslashqjed$ ;  $\backslash\Q_3=\backslashqtrzy$ ;
odchylenie stand.  $\backslash\ostan$ ;
odchylenie ćwiartkowe  $\backslash\ocw$ ;
klasyczny wsp. sym.  $\backslash\kws$ ;
kurtoza  $\backslash\kurt$ ;}
...
\whiledo{\ile<5}{\advance\ile1
\immediate\write18{call pojed 1f\the\ile p}
{\par\bf Parametry \the\ile\ pytania:\\}
\input{pojed.tex}\pizspojed}
...
\input{pojed.tex}
\pizspojed

```

Dla przykładu, jeśli chcemy uzyskać różne dane „korelacyjne” pomiędzy pytaniami tworzymy pliki `1fnp.uuu` dla $n = 1, 2, 3, 4, 5$ i uruchamiamy dziesięciokrotnie program `korej.exe` z odpowiednimi plikami i wyniki w analogiczny sposób zapisujemy do `pdf-a`.

A oto odpowiadające tej operacji fragmenty algorytmu TeX-a.

```

\def\pizskorej{
liczba danych  $\backslash\liczbadan$ ;
I średnia  $\backslash\psred (\overline{x})$ ;
I odchylenie stand.  $\backslash\postan$ ;
II średnia  $\backslash\dsred (\overline{y})$ ;
II odchylenie stand.  $\backslash\dostan$ ;
kowariancja  $\backslash\kwar$ ; korelacja  $\backslash\korel$ ;
\bbbj=\bbj pt
\ifdim\bbbj<0pt\def\znack{}\else\def\znack{+}\fi
{\it I regresja  $y=\backslash\aa_j x\backslash\znack\backslash\bbj$ ;
\bbbj=\bbd pt
\ifdim\bbbj<0pt\def\znack{}\else\def\znack{+}\fi
II regresja:  $\backslash\backslash\aa_j y\backslash\znack\backslash\bbd$ ; współczynnik zgodności  $\backslash\backslash\wspzgg$ ;} }
\ile=0\ilep=0
\whiledo{\ilep<5}{\advance\ilep1\ile=\ilep
\whiledo{\ile<5}{\advance\ile1
\immediate\write18{call korej 1f\the\ilep p 1f\the\ile p}
{\par\bf Ogólna współzależność \the\ilep\ i \the\ile\ pytania:\\}
\input{korej.tex}\pizskorej
}}

```

Program `korej.tex` został tak napisany, że nie uwzględnia k -tych linii w **obu plikach**, jeśli przynajmniej w jednym z nich jest wartość 0, bo oznacza to, że dany student na jedno lub drugie pytanie nie odpowiedział.

Podobnie program tworzy dane dotyczące poszczególnych wykładowców, korelacji pomiędzy nimi w całości lub dla poszczególnych pytań itp. W sumie bardzo dużo pożytecznych danych. Np plik pdf z danymi statystycznymi dla grupy liczącej 33 studentów i 9 przedmiotów liczy 21 stron, w niektórych większych grupach liczba stron przekracza 100!

5. Przygotowanie do analizy ekonometrycznej

Analiza pięciu pytań, na które odpowiadają studenci sugeruje następujący wniosek: ⁶

Wydaje się, że celem, czyli zmienną zależną powinien być parametr odpowiadający pytaniu nr 1, a pozostałe są w pewnym sensie środkiem do tego celu, czyli zmiennymi zależnymi. O tym czy tak jest w istocie można się przekonać robiąc odpowiednią analizę ekonometryczną. I tu przyszedł autorowi w sukurs pewien stworzony dawno temu razem ze studentami program (a w zasadzie odpowiednie makro) stworzony dla arkusza kalkulacyjnego (open office calc⁷), te analizę wykonujący. Programik wyliczał podstawowe parametry ekonometryczne oraz wyliczał przy pomocy metody Hellwiga najodpowiedniejszy dobór zmiennych niezależnych. Metody te można znaleźć np. w [6].

Wydaje się, że najbardziej sensownym materiałem podlegającym analizie będzie wykorzystanie jako danej ekonometrycznej średniej oceny danego wykładowcy dla danego pytania.

Aby wprowadzić dane do tego programiku, trzeba je przygotować w postaci macierzowej (dane w wierszach oddzielone średnikami).

W czasie tworzenia danych statystycznych prezentowane w poprzednim rozdziale algorytmy wyliczając olbrzymia liczbę danych używały kolejno plików o tej samej nazwie do kolejnych obliczeń i zapisując te obliczenia do pdf-a.

Trzeba było zmodyfikować te algorytmy tak, aby zapisywać do nowego pliku informacje o numerze wykładowcy, numerze pytania i obliczonej średniej, zanim informacja zostanie nadpisana nowymi obliczeniami.

Wystarczyło w pętli zapisującej do pdf-a dane o średnich dopisać linijki wymuszające zapis do tego pliku. Oto one:

```
\immediate\openout11=srednie.vvv}
```

⁶Przypominamy pytania: **1. Klarowność prowadzenia zajęć; 2. Życzliwość wobec studentów; 3. Przejrzystość zasad zaliczania; 4. Atrakcyjność zajęć; 5. Przydatność materiałów dydaktycznych.**

⁷Zarówno na zajęciach ze studentami, jak i w artykułach należy zdaniem autora tam, gdzie to się da, unikać oprogramowania komercyjnego, aby nie narażać studentów na koszty lub łamanie praw autorskich.

```
\immediate\write11{\string\sred{\the\ilep}}
\immediate\openout8=\the\ilep_sr.uuu
\whiledo{\ile<5}{\advance\ile1
\immediate\write18{call pojed 1f\the\ilep w\the\ile p}
{\par\bf Parametry \the\ilep\ wykładowcy oraz \the\ile\ pytania:\\}
\input{pojed.tex}\pizspojed
\immediate\write11{{\sred}}
```

Linie dotyczące tego zapisu to 1,2 i ostatnia. Linia druga zapisuje numer wykładowcy, linia 5 wymusza wyjście „na zewnątrz” i obliczenie średniej (jak również innych danych, ale do celów ekonometrycznych z nich nie korzystamy) z pytania numer `\ile` z zapisem do pliku `pojed.tex`, a przedostania zapisuje do pliku `srednie.vvv` danych potrzebnych do analizy ekonometrycznej.

A oto fragment tego pliku:

```
\sred{1}
{5.84}
{5.94}
{5.84}
{5.81}
{5.77}
\sred{2}
{5.35}
{5.81}
{5.83}
{5.32}
{5.58}
```

Teraz przekształcamy ten plik w postać macierzową, gdzie kolumny odpowiadają za numer pytania, a wiersze za numer ocenianego wykładowcy przy pomocy algorytmu:

```
\def\sred#1#2#3#4#5#6{\immediate\write11{#3;#4;#5;#6;#2}}
\immediate\openout11=ekonometria.vvv
\input{srednie.vvv}
\immediate\closeout11
```

6. Podsumowanie

Dzięki algorytmom zaprezentowanym w pracy i bogatej historii przeprowadzania ankiet oceniających wykładowców w WSM w Warszawie można prawie automatycznie uzyskać olbrzymią liczbę danych statystycznych i ekonometrycznych. Dość powiedzieć, że ankiety w tej formie prowadzone są od 9 lat,

każdego roku oceniają studenci z około 60 grup, przeciętna liczba wykładowców w grupie jest ponad 10. Może to dać w przybliżeniu od 20000 do 30000 stron danych. Na każdej stronie danych jest kilkadziesiąt.

Analiza tych danych może dać ciekawe wnioski dydaktyczne, a nawet psychologiczne i mamy nadzieję, że zostanie przeprowadzona w najbliższym czasie.⁸

Bibliografia

- 1 Carlisle D. (1994), *The ifthen package*
<http://texdoc.net/texmf-dist/doc/latex/base/ifthen.pdf>
- 2 Knuth D. E. (2005); *TeX Przewodnik użytkownika*, WNT
- 3 Rusinek J. (2007); *Algorytm permutowania w TeX-u zastosowany do informatyzacji procesu egzaminacyjnego*, „Rocznik Naukowy Wydziału Zarządzania w Ciechanowie”, 1-4 (I), (153–174)
- 4 Rusinek J. (2009); *Testy egzaminacyjne z matematyki*, „Rocznik Naukowy Wydziału Zarządzania w Ciechanowie”, 3-4 (III), (101–111)
- 5 Rusinek J., (2010); *Obliczeniowe możliwości TeX-a. Zastosowanie wspomagające dydaktykę*, „Rocznik Naukowy Wydziału Zarządzania w Ciechanowie”, 1-4 (IV), (119-131)
- 6 Sobczyk M., (2013). *Ekonometria*, Wydawnictwo C.H. Beck
- 7 Sobczyk M., (2006). *Statystyka aspekty praktyczne i teoretyczne*, Wydawnictwo UMCS Lublin 2006

⁸Ciekawe byłoby np. porównanie wyniku wykładowcy, który z danego przedmiotu wypadł nienajlepiej z wynikiem tego wykładowcy z tego samego przedmiotu w kolejnych latach (poprawa tego wyniku pokazywałaby na wyraźne korzyści z ankiety), ale to wymagałoby „odtajnienia” wyników, które znane są tylko ocenianemu wykładowcy i władzom uczelni.