

Beata Czarnacka-Chrobot

Szkoła Główna Handlowa w Warszawie

WIARYGODNOŚĆ METOD SZACOWANIA PRACOCHOŁONNOŚCI PRZEDSIĘWZIĘĆ ROZWOJU SYSTEMÓW OPROGRAMOWANIA WSPOMAGAJĄCYCH ZARZĄDZANIE

Streszczenie: Kluczowym wymaganiem wobec metod szacowania pracochłonności przedsięwzięć rozwoju systemów oprogramowania jest ich wiarygodność w określonym obszarze aplikacyjnym. Jednym z zasadniczych obszarów są systemy oprogramowania wspomagające zarządzanie (SOWZ), dlatego wiarygodność metod estymacji pracochłonności ich rozwoju stanowi przedmiot wielu badań. Ich rezultaty są zróżnicowane, niekiedy nawet sprzeczne. Powstaje zatem pytanie, czy na ich tle można dostrzec prawidłowości, które mogą stanowić wskazówkę co do wyboru najwłaściwszego dla przedsięwzięć rozwoju SOWZ sposobu szacowania pracochłonności. Celem artykułu jest próba wyprowadzenia ewentualnych takich prawidłowości na podstawie analizy dostępnych w literaturze i rezultatów badań nad wiarygodnością pięciu zasadniczych metod estymacji pracochłonności rozważanych przedsięwzięć.

Słowa kluczowe: przedsięwzięcia rozwoju systemów oprogramowania wspomagających zarządzanie, metody szacowania pracochłonności, wiarygodność metod szacowania pracochłonności.

1. Wstęp

Koszty i czas realizacji przedsięwzięć informatycznych mających na celu rozwój nowych (budowę od podstaw) lub istniejących (doskonalenie) systemów oprogramowania wspomagających zarządzanie – nazywanych w niniejszym artykule informatycznymi przedsięwzięciami rozwojowymi (IPR) – są silnie uzależnione od pracochłonności działań niezbędnych do dostarczenia produktu zgodnego z wymaganiami zleceniodawcy. Jej właściwe oszacowanie jest tym istotniejsze, że rozważane przedsięwzięcia często stanowią poważne inwestycje: wydatki ponoszone na systemy powstające w ich efekcie mogą znacznie przekraczać nakłady na budowę biurów zajmowanych przez zlecające je firmy, a w skrajnych przypadkach nawet pięćdziesięciopiętrowego drapacza chmur, zadaszonego stadionu czy morskiego statku wycieczkowego o wyporności 70 000 ton [Jones 1999, s. 3]. Tymczasem

kwoty te zleceniodawca nierzadko wydatkuje bez poparcia decyzji o zaangażowaniu w taką inwestycję właściwą analizą kosztów, bazującą na racjonalnych, wystarczająco obiektywnych i wiarygodnych podstawach. Objawem tego stanu rzeczy jest fakt, że koszty ponoszone przez różne organizacje na zbliżone funkcjonalnie aplikacje mogą się różnić nawet piętnastokrotnie [*Acquiring Custom-Build Software...* 2000, s. 1-3].

Nakłady pracy to atrybut niezmiernie istotny także z perspektywy organizacji wytwarzających oprogramowanie. W odniesieniu do przedsięwzięć rozwoju nowych produktów programowych szacuje się, iż koszty osobowe mogą wynosić nawet 80% ich kosztów całkowitych. „Panuje zgodność [...] co do tego, że koszty używanego w procesie produkcji sprzętu i oprogramowania stanowią dla budżetu realizowanego projektu znikome obciążenie. [...] Nawet koszty pozyskania nowych narzędzi są znikome wobec kosztów osobowych” [Gontarz 2003].

Z powyższych powodów trafność prognozy pracochłonności przedsięwzięcia budowy lub doskonalenia systemu oprogramowania powinna być jak największa. W badaniach mających na celu analizę wiarygodności metod estymacji tego atrybutu stosuje się zazwyczaj jedno z następujących kryteriów jej wystarczającego poziomu (por. np. [Abran, Robillard 1993; Kemerer 1987; Verner, Graham 1992]):

- Dla co najmniej 80% przypadków metoda umożliwia uzyskanie takich ocen estymacyjnych, które nie przekraczają założonego względnego błędu, równego 30% uzyskanej w rzeczywistości wartości pracochłonności.
- W co najmniej 75% przypadków metoda umożliwia uzyskanie takich ocen estymacyjnych, które mieszczą się w granicach wyznaczonych przez założony względny błąd, równy 25% rzeczywistej wartości pracochłonności.
- Dla co najmniej 70% przypadków metoda umożliwia uzyskanie takich ocen estymacyjnych, które nie przekraczają przyjętego względnego błędu, równego 20% rzeczywistej wartości pracochłonności.
- W co najmniej 60% przypadków metoda umożliwia uzyskanie takich ocen estymacyjnych, które mieszczą się w granicach wyznaczonych przez założony względny błąd, równy 10% rzeczywistej wartości pracochłonności.

2. Analiza rezultatów badań wiarygodności metod szacowania pracochłonności przedsięwzięć rozwoju systemów oprogramowania wspomagających zarządzanie

W niniejszym artykule analizie poddano rezultaty badań dotyczących zasadniczych metod szacowania pracochłonności, wykorzystywanych wobec informatycznych przedsięwzięć rozwojowych. Do metod tych należą:

- metoda szacowania przez analogię,
- metody dekompozycyjne,
- metody eksperckie,
- metody ekstrapolacji parametrycznej (algorytmiczne) z podziałem na:

- metody oparte na rozmiarze produktu wyrażonym w jednostkach programowych (np. w liczbie linii kodu źródłowego),
- metody oparte na rozmiarze produktu wyrażonym w jednostkach funkcjonalności (np. w liczbie punktów funkcyjnych)¹.

O pozytywnych efektach stosowania metody szacowania przez analogię donoszą M. Shepperd i C. Schofield [1997], według których estymacja pracochłonności IPR zgodnie z tą metodą daje lepsze rezultaty w zestawieniu z ocenami uzyskiwanymi na podstawie statystycznie wyprowadzonych modeli algorytmicznych, co potwierdza wcześniejsze badania F. Heemstra [1992]. Tymczasem J. Maddix [2004] w wyniku późniejszych badań, uwzględniających już rozwój niektórych podejść stanowiących bazę dla metod algorytmicznych, doszedł do wniosku, zresztą wbrew swoim oczekiwaniom, iż oceny szacunkowe uzyskane za pomocą modelu ekstrapolacji parametrycznej opartego na rozmiarze produktu wyrażonym w jednostkach funkcjonalności (tu: w liczbie punktów funkcyjnych metody IFPUG [2004]) są znacznie bliższe rzeczywistym wartościom (i faktycznie im bliskie) niż oceny uzyskane na zasadzie analogii przez inżynierów z co najmniej dziesięcioletnim doświadczeniem. Zwraca jednak przy tym uwagę na bardzo istotny fakt: szacowana wartość musi uwzględniać także dane historyczne, obrazujące zależności specyficzne dla dziedziny zastosowania, rodzaju systemu i danego środowiska projektowego. Toteż z jego badań wynika, iż bazowanie jedynie na danych historycznych, nawet z uwzględnieniem doświadczenia osób oceniających, daje gorsze rezultaty niżli powiązanie takich danych z metodą szacowania rozmiaru produktu w jednostkach funkcjonalności. Konieczność dostosowania empirycznych modeli parametrycznych na poziomie dziedziny zastosowania i organizacji stwierdzili także we wcześniejszych badaniach D. Jeffery i G. Low [Jeffery, Low 1990; Jeffery, Low, Barnes 1993], którzy obliczyli, że bez takiego dostosowania średni absolutny błąd względny w przypadku metod bazujących na rozmiarze produktu wyrażonym w jednostkach funkcjonalności waha się w granicach od 43% do 105%, natomiast po nim może wynosić nawet tylko 12%.

F. Heemstra w przywołanych już badaniach ocenił także, że wykorzystanie metod algorytmicznych nie daje lepszych wyników w zestawieniu z tymi, które używa się za pomocą metod eksperckich. Jednakże S. Vicinanza, T. Mukhopadhyay i J. Prietula [1991] obliczyli, iż prognozy uzyskiwane za pomocą metod eksperckich charakteryzują się średnim absolutnym błędem względnym od 32 do aż 1107%, podczas gdy D. Ferens i R. Gurner [1993], bazując częściowo na tych samych danych, wykazali, że modele szacowania oparte na rozmiarze produktu wyrażonym w jednostkach funkcjonalności nie przekraczają średniego absolutnego błędu względnego równego 105%. Błąd ten jest także wysoki, ale znacznie maleje przy odpowiednim ich dostosowaniu do specyficznych warunków projektowych.

¹ O różnych rodzajach miar produktu programowego różnicujących metody ekstrapolacji parametrycznej Czytelnik znajdzie informacje w [Czarnacka-Chrobot, Kobyliński 2009].

Ciekawe badania na temat wiarygodności metod eksperckich w zestawieniu z metodami algorytmicznymi przeprowadził M. Jørgensen [2007]. Po pierwsze, doszedł on do wniosku, że przeciętna wartość średniego absolutnego błędu procentowego dla ocen szacunkowych dotyczących pracochłonności przedsięwzięć rozwojowych wynosi 30-40%, tj. w rzeczywistości na takie przedsięwzięcia wydatkuje się średnio o 30-40% więcej nakładów pracy w zestawieniu z oszacowaniami. Po drugie, przeanalizował on 16 opublikowanych rezultatów studiów podejmujących problematykę trafności estymacji przy wykorzystaniu wspomnianych metod (z lat 1990-2006), dochodząc do ogólnego wniosku, opatrzonego jednak licznymi zastrzeżeniami, że w 10 na 16 przypadków średnia trafność ocen eksperckich dla pracochłonności była większa (czasami nieznacznie) w zestawieniu z ocenami uzyskanymi za pomocą metod algorytmicznych. Zdaniem autora w uwzględnionych przypadkach wynika to z faktu niedostosowania empirycznych modeli parametrycznych do specyfiki organizacji oraz z nieuwzględniania w nich ważnych kontekstowych informacji, które posiadają eksperci.

M. Jørgensen zastrzega jednak, iż wniosek ów może być zbyt daleko idący z następujących m.in. przyczyn: liczbę przeanalizowanych przypadków uznaje on za małą, nie uwzględnił – z powodu braku informacji w analizowanych badaniach – podejść do szacowania wykorzystywanych przez ekspertów (mogli wspomagać swoją ocenę metodą algorytmiczną), a co najistotniejsze – inne metody szacowania były wykorzystywane do prognozowania odmiennych zadań estymacyjnych, co może prowadzić do zupełnie różnych wniosków. Okazuje się bowiem, że w przypadku porównywania zadań estymacyjnych o podobnym poziomie złożoności metody eksperckie i metody algorytmiczne mogą dawać rezultaty o bardzo zbliżonym poziomie trafności. Natomiast bez uwzględnienia stopnia skomplikowania zadań i przy założeniu, że osoba oceniająca na bazie metody algorytmicznej posiada doświadczenie w realizacji podobnych przedsięwzięć, oceny estymacyjne przez nią wyprowadzone są znacznie lepsze. Co więcej, olbrzymie znaczenie ma wybór ekspertów o odpowiedniej wiedzy, jako że najmniej trafne oceny eksperckie są gorsze w każdym z przeanalizowanych przez niego przypadków niż te uzyskane za pomocą metod algorytmicznych. Dodatkowo autor zauważa, że im później przeprowadzono analizowane przez niego studia, tym częściej wskazują one na przewagę metod algorytmicznych, co może wynikać z ich doskonalenia w miarę upływu czasu. Zwraca także uwagę, że na ocenę ekspertów nierzadko ma wpływ „myślenie życzeniowe”, czego unika się siłą rzeczy przy metodach algorytmicznych.

Wobec powyższego M. Jørgensen podkreśla, iż na podstawie rezultatów jego badań nie można wysnuć wniosku o potrzebie zastąpienia metod algorytmicznych metodami eksperckimi ani odwrotnie. Co więcej, wskazuje on na problemy w wyznaczeniu reguł, zgodnie z którymi powinna być wybierana najwłaściwsza w danych warunkach projektowych metoda estymacji pracochłonności. Dlatego autor proponuje wykorzystanie kombinacji dwóch metod: metody eksperckiej oraz metody algorytmicznej. Jest to pogląd zgodny z ogólną zasadą szacowania, postulującą, aby

było ono realizowane w oparciu o co najmniej dwie różne metody. Wniosek M. Jørgensena potwierdza rezultaty badań R. Blattberga i S. Hocha [1990] oraz P. Goodwina [2000]. Jednak badania autorstwa S. Whitecotton i współautorów [Whitecotton, Sanders, Norris 1998] im zaprzeczają: wynika z nich, że uwzględnianie przez ekspertów podczas procesu szacowania nieistotnych dla ostatecznego wyniku informacji prowadzi do przewagi metod algorytmicznych.

Zdaniem autorki niniejszego opracowania jedną z zasadniczych przyczyn tak odmiennych wniosków płynących z badań M. Jørgensena jest różnorodność uwzględnianych w nich modeli ekstrapolacji parametrycznej. W przeanalizowanych przez niego studiach wiarygodność metody COCOMO 81 (metoda algorytmiczna oparta na jednostkach programowych) okazała się być rzeczywiście niska (średni absolutny błąd procentowy od ok. 620% do ok. 760%). Tymczasem jeżeli do estymacji wykorzystano model, w którym rozmiar produktu wyrażony był w jednostkach funkcjonalności, średni absolutny błąd procentowy wahał się w granicach od 10% do 107%. Wyraźnie zauważalne są także dwie tendencje: po pierwsze, im później zrealizowano analizowane studium, tym średni absolutny błąd procentowy dla modelu opartego na rozmiarze produktu wyrażonym w jednostkach funkcjonalności jest mniejszy²; po drugie, jest on tym mniejszy, im wyższy był poziom dostosowania empirycznego modelu parametrycznego estymacji pracochłonności do specyfiki organizacji.

Analiza rezultatów badań M. Jørgensena potwierdza wnioski płynące z porównania dwóch rodzajów metod algorytmicznych szacowania pracochłonności przedsięwzięć rozwojowych przeprowadzonego przez C. Kemerera [1987]. Wynika z nich, że średni błąd względny dla metody SLIM (metoda algorytmiczna oparta na jednostkach programowych) wyniósł ponad 770%, dla COCOMO 81 ok. 600%, a przy zastosowaniu modeli bazujących na rozmiarze produktu wyrażonym w jednostkach funkcjonalności (tu: w punktach funkcyjnych) średni błąd względny równał się nieco ponad 100%. Należy nadmienić, że od czasu przeprowadzenia tych badań metoda COCOMO i metody wymiarowania rozmiaru produktu w jednostkach funkcjonalności uległy znacznemu rozwojowi.

Co ciekawe, niską wiarygodność metody COCOMO 81 potwierdził również jej autor B. Boehm [Abran, Robillard 1993, s. 82]. Zgodnie z jego badaniami model podstawowy tej metody dawał rezultaty, które przy założonym poziomie błędzie $\pm 20\%$ były trafne jedynie dla 25% badanych przedsięwzięć. Dla uwzględnienia 70% analizowanych przypadków poziom błędzie musiałby wzrosnąć do $\pm 100\%$. Z tego przede wszystkim powodu B. Boehm zmodyfikował swoją metodę do wersji COCOMO II, uwzględniającej przy wyrażaniu rozmiaru produktu już nie tylko jednostki programowe, ale także jednostki jego złożoności konstrukcyjnej (tj. punkty obiektowe) i funkcjonalności.

² W ostatnich latach bowiem w efekcie zdobywania kolejnych doświadczeń nastąpił duży postęp w rozwoju metod szacowania rozmiaru produktu w jednostkach funkcjonalności, np. metody IFPUG [2004].

Istnieją badania (por. np. [Devnani-Chulani i in. 1998; Yang i in. 2005]), które rzeczywiście wykazują znacznie większy wskaźnik trafności predykcji dla COCOMO II w zestawieniu z COCOMO 81 – w zależności od wersji metody COCOMO II, fazy szacowania i dostosowania wykorzystywanego modelu do specyficznych warunków projektowych. Inne analizy (por. np. [Chulani, Steece 2000; Clark, Yang 2004]) wskazują nawet na spełnienie przez COCOMO II w zmodyfikowanej wersji z 2000 r. kryteriów, przy których metodę szacowania uznaje się za wiarygodną – dotyczą one jednak dostosowanego do specyfiki organizacji modelu postarchitektonicznego, a zatem ocen estymacyjnych uzyskiwanych stosunkowo późno w cyklu życia przedsięwzięcia. Ogólnie wiarygodność rodziny obecnie istniejących metod COCOMO³ ocenia się przy użyciu dopuszczalnego błędu na poziomie $\pm 30\%$ na 43-69% przedsięwzięć [Garnsey i in. 2006, s. 12]. Jednakże zgodnie z obszernymi analizami D. Ferensa i D. Christensena [2000] wiarygodność ta dla COCOMO II jest znacznie niższa: średni absolutny błąd procentowy dochodzi nawet do 790%, a dostosowanie modeli tylko nieznacznie ją poprawia.

D. Ferens i D. Christensen wykazali także, że jeżeli do szacowania pracochłonności przedsięwzięcia informatycznego wykorzystano model oparty na rozmiarze produktu wyrażonym w jednostkach funkcjonalności, to po jego dostosowaniu do specyfiki projektowania oceny estymacyjne spełniają warunki wiarygodnej metody estymacji. Przy tym jest to najbardziej widoczne dla systemów oprogramowania wspomagających zarządzanie, dla których już przed takim dostosowaniem udział trafnie oszacowanych przedsięwzięć przy założonym poziomie błędu $\pm 25\%$ zbliża się do 70%, jednak dotyczy także przykładowych systemów czasu rzeczywistego. Wniosek ów potwierdza rezultaty jednych z najbardziej znanych studiów dotyczących problematyki wiarygodności metod ekstrapolacji parametrycznej, bazujących na rozmiarze produktu wyrażonym w jednostkach funkcjonalności, przeprowadzonych na początku lat 90. przez C. Kemerera [1993]. Badał on ich wiarygodność z perspektywy zgodności ocen estymacyjnych wyprowadzonych przez różne osoby dla rozmiaru produktów typu SOWZ. Rezultaty zrealizowanego porównania sugerują znacznie większą wiarygodność modeli opartych na punktach funkcyjnych niż można było oczekiwać: średnia różnica między prognozami wyprowadzonymi przez różnych specjalistów wynosi ok. 12%. Dlatego C. Kemerer ocenia, że pomiar rozmiaru systemów oprogramowania wspomagających zarządzanie w jednostkach funkcjonalności daje podstawę do wystarczająco obiektywnej i wiarygodnej estymacji pracochłonności. Co więcej, postuluje ich wykorzystanie do oceny produktywności nie tylko rozwoju, ale i utrzymania oprogramowania oraz do oceny jego jakości, podkreślając jednocześnie, że także Quality Assurance Institute uznał te jednostki za najlepszą z dostępnych miar produktywności dla SOWZ. Zbliżone wnioski płyną z analiz W. Siddiquee'a [1993] i S. Fureya [1997], a co do obiektywności pomiaru

³ Istnieje ponad 15 wariantów tej metody. Aktualne informacje o nich można znaleźć na stronie internetowej University of Southern California, Center for Systems and Software Engineering: <http://sunset.usc.edu/research/cocomosuite>, 31.01.2008.

także z badań T.C. Jonesa [1998]. Ten ostatni podkreśla potrzebę przeprowadzania procesu szacowania zgodnie z ustalonymi standardami i z wykorzystaniem właściwych narzędzi wspomagających, co powoduje, iż odchylenie w otrzymanyach wynikach mieści się w założonych granicach, czyli wynosi ok. 10%.

Autorami równie istotnych dla rozwoju metod estymacji pracochłonności studiów nad wiarygodnością modeli opartych na rozmiarze produktu wyrażonym w jednostkach funkcjonalności są A. Abran i P. Robillard [1993]. Ich rezultaty wskazują, że dla projektów polegających na rozbudowie rozwiązań typu SOWZ wyprowadzony przez autorów model regresji liniowej bazujący na rozmiarze produktu wyrażonym w punktach funkcyjnych spełnia kryterium wystarczająco wiarygodnej metody szacowania: udział trafnie oszacowanych przedsięwzięć przy założonym poziomie błędu $\pm 25\%$ nie jest mniejszy niż 75%, średni zaś absolutny błąd procentowy kształtuje się na poziomie poniżej 30%. Co więcej, autorzy przeanalizowali także wiarygodność prostszego modelu szacowania pracochłonności, opartego jedynie na liczbie punktów funkcyjnych i średnim koszcie jednostkowym. W tym przypadku udział trafnie oszacowanych przedsięwzięć przy założonym poziomie błędu $\pm 25\%$ wynosi niecałe 70%, a średni absolutny błąd procentowy nie przekracza 25%. Na podstawie tych rezultatów autorzy wyprowadzili wniosek, że wiarygodność prostego modelu estymacji, w którym rozmiar produktu wyrażono w tych podstawowych jednostkach funkcjonalności, a koszty jednostkowe odzwierciedlają średnią produktywność dostosowaną do organizacji projektującej, okazuje się być na poziomie zbliżonym do wiarygodności znacznie bardziej skomplikowanych modeli, uwzględniających kosztowe czynniki korygujące (np. COCOMO II).

Od drugiej połowy lat 90. State Government of Victoria (Australia) do estymacji budżetów informatycznych przedsięwzięć rozwojowych realizowanych na swoje potrzeby wykorzystuje stworzoną przez siebie metodykę o nazwie „southernSCOPE” [*SouthernSCOPE...* 2000]. Opiera się ona na szacowaniu kosztów przedsięwzięcia na bazie pracochłonności, do której estymacji wykorzystuje się prognozę rozmiaru produktu wyrażonego w jednostkach funkcjonalności. Po trzech latach jej stosowania stwierdzono, że wykorzystanie tej metodyki znacznie zwiększa prawdopodobieństwo zakończenia IPR w ramach zaplanowanego budżetu: jego średnie przekroczenie w stosunku do prognozy wyprowadzonej pod koniec fazy specyfikacji wymagań wynosi mniej niż 10% i jest znacznie niższe od przeciętnego (por. [*Sage Technology...* 2000; Hill 2007, s. 26-27, 30]).

Szerokie badania na temat wiarygodności metod estymacji atrybutów IPR zrealizowała również instytucja International Software Benchmarking Standards Group [*The ISBSG Report...* 2005]. W wyniku przeanalizowania przez nią 439 zakończonych projektów – przy założeniu uznania oceny estymacyjnej za wiarygodną, jeśli jej maksymalne odchylenie od wartości rzeczywistej nie przekracza 10% – ustalono, że:

- w granicach ocen szacunkowych zarówno dla pracochłonności, jak i czasu trwania projektu mieści się 25% przeanalizowanych przedsięwzięć, niemal 40%

z nich przekracza jedną z tych ocen, przy czym częściej dotyczy to nakładów pracy (25%), około 35% projektów przekracza obie te oceny;

- małe przedsięwzięcia, z krótkim czasem trwania i małymi zespołami projektowymi częściej charakteryzują się trafnymi szacunkami; nieprawidłowe oceny estymacyjne, zwłaszcza co do czasu trwania projektu, są zauważalne przede wszystkim w przypadku przedsięwzięć realizowanych przez niedoświadczony zespół projektowy na platformie klient-serwer i dla dużej liczby użytkowników;
- przeszacowania są rzadkie, zwykle niewielkie i występują zazwyczaj w małych przedsięwzięciach;
- z badań jasno wynika, że całkowite koszty IPR są zdeterminowane jego kosztami osobowymi, a zatem pracochłonnością: wielkość błędu w prognozie kosztów IPR stanowi ewidentne odzwierciedlenie wielkości błędu w prognozie nakładów pracy;
- wbrew oczekiwaniom nie istnieje związek pomiędzy wielkością produktywności a trafnością ocen estymacyjnych.

W badaniu ISBSG stwierdzono wykorzystywanie różnych technik estymacyjnych dla pracochłonności przedsięwzięcia w następujących proporcjach: w 38% przedsięwzięć zastosowano jedynie metodę dekompozycyjną bazującą na podziale pracy; w 16% – tylko model oparty na rozmiarze produktu wyrażonym w punktach funkcyjnych; w 27% – oba powyżej wymienione podejścia; w 18% – nie wykorzystywano żadnego z tych podejść, opierając się jedynie na modelach cyklu życia projektów lub na narzędziach. W 17% tych przypadków przedsięwzięcia były sterowane czasem (ze względu na dyrektywy zarządu i/lub klienta, wymagania prawne, wymagania okolicznościowe), toteż miały one z góry wyznaczoną datę zakończenia. Na tej podstawie ustalono, że:

- nie ma związku pomiędzy wykorzystywanymi technikami szacowania a rodzajem organizacji realizującej projekty ani rodzajem IPR;
- jeżeli znane są prognozy co do rozmiaru produktu, to są one niemal w każdym przypadku wykorzystywane do oszacowania pracochłonności, kosztów i czasu trwania przedsięwzięcia; gdy zaś nie są one dostępne, szacunki bazują zwykle na podziale pracy lub z góry wyznaczonej dacie zakończenia przedsięwzięcia. Odnośnie estymacji kluczowych atrybutów IPR z raportu ISBSG wynika, że:
 - pracochłonność przedsięwzięcia jest zwykle niedoszacowana (58% przedsięwzięć); atrybut ten bywa także trafnie szacowany (23%) lub przeszacowywany (19% – dotyczy z reguły małych projektów), a ponadto:
 - dla wszystkich przeanalizowanych przedsięwzięć średni błąd szacunku wskazuje na niedoszacowanie o wielkości 60%;
 - nie zauważono ewidentnych związków pomiędzy trafnością ocen estymacyjnych dla nakładów pracy a różnymi rodzajami przedsięwzięć oraz wykorzystywaną techniką estymacji;

- koszty realizacji przedsięwzięcia są estymowane bardziej trafnie w zestawieniu z ocenami szacunkowymi dla pracochłonności: prawie połowa (49%) analizowanych przedsięwzięć charakteryzuje się trafnymi prognozami; zdarzają się także projekty z nieoszacowanymi kosztami (35% z medianą błędu 25%) lub pod tym względem przeszacowane (16% – dotyczy z reguły małych przedsięwzięć); z badań wynika ponadto, że:
 - widoczna jest silna zależność poziomu trafności ocen estymacyjnych od wykorzystywanych metod szacowania: dla technik opartych jedynie na podziale pracy oceny estymacyjne są trafne w 40% przypadków, a niedoszacowania występują trzy razy częściej niż przeszacowania, tymczasem modele oparte na rozmiarze produktu wyrażonym w jednostkach funkcjonalności wykazują trafność na poziomie 70%, a dalsze 20% mieści się w granicach 20% błędu, tylko w 10% przypadków poziom błędu wynosi co najmniej 20% – dla 90% przypadków zatem oceny estymacyjne kosztów uzyskane na bazie modeli wykorzystujących rozmiar produktu wyrażony w punktach funkcyjnych wykazują odchylenie nieprzekraczające 20% ich rzeczywistej wartości;
 - różnice między poziomem trafności ocen estymacyjnych dla pracochłonności i kosztów mogą być spowodowane faktem, że dodatkowe nakłady pracy wydatkowane na przedsięwzięcie pozostają nieopłacone;
- czas trwania przedsięwzięcia jest szacowany stosunkowo dobrze: ponad połowa projektów kończy się w przewidywanym czasie (ok. 51%) lub wcześniej (ok. 4%), a ponadto:
 - dla wszystkich przeanalizowanych przedsięwzięć średni błąd szacunku wskazuje na niedoszacowanie o wielkości 40%;
 - trafność ocen estymacyjnych dla czasu realizacji zależy przede wszystkim od planowanego czasu trwania przedsięwzięcia: im jest on dłuższy, tym oceny estymacyjne są bardziej trafne; najniższy poziom trafności zanotowano przy projektach planowanych na mniej niż 3 miesiące, a najwyższy dla przedsięwzięć planowanych na co najmniej 12 miesięcy (są one z reguły realizowane na czas);
 - trafność ocen estymacyjnych dla czasu trwania przedsięwzięcia zależy także od wykorzystywanej techniki estymacji: najgorzej kształtują się one w przypadku wykorzystania podejść opartych jedynie na modelu cyklu życia projektu (średnie niedoszacowanie wynosi ok. 70%), lepiej w przypadku metod dekompozycyjnych opartych na podziale pracy (średnie niedoszacowanie równe ok. 40%), natomiast modele bazujące na rozmiarze produktu wyrażonym w jednostkach funkcjonalności są w ponad połowie przypadków dokładnie zgodne z wartościami rzeczywistymi (choć reszta z nich wykazuje niedoszacowanie o średnim błędzie ok. 45%); zauważono także, że trafność ocen estymacyjnych znacznie się poprawia, jeżeli wykorzystuje się prognozę rozmiaru produktu w jednostkach funkcjonalności w połączeniu z narzędziem wspomagającym estymację;

- lepsze oceny estymacyjne dla czasu realizacji przedsięwzięcia mogą być tłumaczone faktem wykorzystania niemal w każdym przypadku do jego oszacowania prognozy rozmiaru produktu, o ile tylko jest ona dostępna.

Rezultaty badań ISBSG prowadzą do następujących zasadniczych wniosków:

1. Ogólnie rzecz ujmując, przedsięwzięcia, w których do estymacji atrybutów wykorzystuje się modele szacowania bazujące na rozmiarze produktu wyrażonym w jednostkach funkcjonalności, charakteryzują się najbardziej trafnymi prognozami.

2. Ze wszystkich trzech analizowanych atrybutów przedsięwzięcia najniższa trafność ocen estymacyjnych występuje przy szacowaniu pracochłonności; dla pozostałych dwóch atrybutów mniej więcej połowa przedsięwzięć jest właściwie szacowana, podczas gdy dla nakładów pracy odsetek ten wynosi mniej niż 25%.

3. Zasadnicze czynniki wpływające na trafność prognoz dla atrybutów IPR to:

- wielkość podejmowanego przedsięwzięcia: małe projekty sprzyjają trafnym ocenom estymacyjnym, czasami też są przeszacowywane, duże projekty zaś są często niedoszacowane (wyjątkiem jest tu czas realizacji przedsięwzięcia),
- rodzaj IPR: przedsięwzięcia polegające na doskonaleniu produktu programowego są łatwiejsze do trafnego oszacowania niż projekty budowy nowego systemu (zwłaszcza co do czasu realizacji),
- technologia realizacji: oceny estymacyjne są mniej trafne dla projektów opartych na stosunkowo nowej technologii,
- liczebność i zróżnicowanie użytkowników: im są one mniejsze, tym oceny estymacyjne są bardziej trafne.

Przy interpretacji rezultatów analizy przeprowadzonej przez ISBSG należy mieć na uwadze, że dane zgromadzone przez tę instytucję przypuszczalnie nie są reprezentatywne dla ogółu informatycznych przedsięwzięć rozwojowych, a raczej dla projektów ponadprzeciętnych. Organizacje bowiem same wybierają projekty, o których dane przekazują do repozytorium ISBSG – mogą to być zarówno przedsięwzięcia dla nich typowe, jak i te o najlepszej trafności prognoz. Co jednak najistotniejsze, dane o nich pochodzą z organizacji charakteryzujących się na tyle dojrzałym procesem wytwarzania oprogramowania, że realizowany jest w nich program wymiarowania produktów programowych.

3. Wnioski

Powyżej przeanalizowano rezultaty 25 badań dotyczących problematyki wiarygodności metod szacowania pracochłonności informatycznych przedsięwzięć rozwojowych. W ich doborze kierowano się przede wszystkim dostępnością, starając się przy tym, aby dobór ten był możliwie obiektywny. Kluczowe wnioski wynikające z ich analizy zamieszczono w tab. 1.

Jak wynika z przedstawionych analiz, wiarygodny proces estymacji pracochłonności informatycznych przedsięwzięć rozwojowych stosunkowo wcześniej w cyklu

Tabela 1. Podsumowanie rezultatów badań dotyczących wiarygodności metod szacowania IPR

Przedmiot badań	Zasadnicze wnioski	Autorzy badań (rok)
Porównanie wiarygodności metod ekstrapolacji parametrycznej z wiarygodnością innych metod szacowania: eksperckich, dekompozycyjnych i szacowania przez analogię	przewaga metod szacowania przez analogię i/lub eksperckich nad metodami ekstrapolacji parametrycznej	Hemstra (1992) Shepperd i Schofield (1997)
	– przewaga metod ekstrapolacji parametrycznej nad pozostałymi metodami szacowania – rozmiar produktu wyrażany w jednostkach funkcjonalności lub postulat wyrażania rozmiaru produktu w takich jednostkach – potrzeba dostosowania empirycznych modeli parametrycznych do specyficznych warunków projektowych	Jeffery i Low (1990, 1993) Vicinanza i inni (1991) Ferens i Gurner (1993) Whitecotton i inni (1998) Maddix (2004) ISBSG (2005)
	– potrzeba wykorzystywania metod ekstrapolacji parametrycznej wraz z metodami eksperckimi – w miarę upływu czasu coraz większa wiarygodność metod ekstrapolacji parametrycznej	Blattberg i Hoch (1990) Goodwin (2000) Jørgensen (2007)
Porównanie wiarygodności metod ekstrapolacji parametrycznej – analiza metod z rodziny COCOMO	niska wiarygodność metody COCOMO 81	Boehm (1991)
	w rodzinie metod COCOMO przewaga metody COCOMO II ze wskazaniem na wersję z 2000 r.	Devnani-Chulani i inni (1998) Chulani i Steece (2000) Clark i Yang (2004) Yang i inni (2005)
	w rodzinie metod COCOMO wskaźniki wiarygodności poniżej granicy uznawanej za konieczną dla wystarczająco wiarygodnej metody szacowania	Garnsey i inni (2006)
Porównanie wiarygodności metod ekstrapolacji parametrycznej – jednostki programowe a jednostki funkcjonalności	przewaga metod ekstrapolacji parametrycznej opartych na rozmiarze produktu wyrażonym w jednostkach funkcjonalności w zestawieniu z metodami opartymi na rozmiarze produktu wyrażonym w jednostkach programowych	Kemerer (1987) Abran i Robillard (1993) Ferens i Christensen (2000)
Wiarygodność metod ekstrapolacji parametrycznej opartych na rozmiarze produktu wyrażonym w jednostkach funkcjonalności	metody ekstrapolacji parametrycznej oparte na rozmiarze produktu wyrażonym w jednostkach funkcjonalności są wystarczająco obiektywne i wiarygodne, zwłaszcza dla systemów klasy SOWZ oraz po dostosowaniu modeli parametrycznych do specyficznych warunków projektowych	Kemerer (1993), Siddiquee (1993) Furey (1997) Jones (1998) State Government of Victoria (2000)

Źródło: opracowanie własne na podstawie rezultatów badań przywoływanych w rozdz. 2.

ich życia stanowi ciągle złożone zagadnienie i przez to jedno z największych wyzwań inżynierii oprogramowania, wymagające dalszych badań. Rezultaty dostępnych analiz są zróżnicowane, jednak nawet ich ostrożna interpretacja w odniesieniu do przedsięwzięć mających na celu rozwój systemów oprogramowania wspomagających zarządzanie pozwala stwierdzić, iż przy uwzględnieniu ewolucji podejść wykorzystywanych do estymacji **spośród wszystkich analizowanych metod szacowania pracochłonności najmniejsze podstawy do odrzucenia istnieją obecnie w stosunku do metod ekstrapolacji parametrycznej opartych na rozmiarze produktu IPR wyrażonym w jednostkach jego funkcjonalności** – szczególnie w przypadku dostosowania modeli parametrycznych do specyficznych organizacyjnych warunków projektowych, do czego niezbędne jest gromadzenie odpowiednich danych historycznych.

Niektóre z przytoczonych badań sugerują wykorzystywanie takich metod wraz ze stosunkowo często spotykanymi w praktyce metodami eksperckimi (te ostatnie jednak nierzadko bywają mylone z szacowaniem przez analogię), co wobec charakterystycznego dla IPR dużego ryzyka związanego z dostarczaniem produktów w ramach prognozowanych kosztów i planowanego czasu realizacji stanowi zasadny postulat. Jednak trzeba mieć przy tym na uwadze, że metody eksperckie charakteryzują się bardzo różną wiarygodnością – stosunek pomiędzy uzyskanymi przez ekspertów ocenami dla tego samego przedsięwzięcia może wynosić 1:6, a w skrajnych przypadkach nawet 1:12⁴. Są one także niepowtarzalne, czasami nieobiektywne, a sposób przeprowadzenia procesu szacowania nie jest wyrażony *explicite*, co nie sprzyja zaangażowaniu zleceniodawcy w IPR, a to stanowi jeden z zasadniczych czynników sukcesu tego typu przedsięwzięć [CHAOS Summary... 2008, s. 2].

Reasumując należy nadmienić, iż autorka niniejszego artykułu podjęła próbę przeprowadzenia badań, których jednym z zasadniczych celów było przeanalizowanie metod estymacji pracochłonności IPR wykorzystywanych przez polskich wykonawców (wewnętrznych i zewnętrznych) dedykowanych systemów oprogramowania wspomagających zarządzanie oraz przyczyn istniejącego stanu rzeczy. Zgodnie z zamierzeniami przedmiotem analizy miała być zarówno popularność poszczególnych metod, jak i ich wiarygodność. Uzyskane odpowiedzi pozwoliły jednakże osiągnąć założony cel tylko połowicznie, a mianowicie w części dotyczącej popularności metod estymacji. Wynika to z faktu, że posiadanie własnych danych wystarczających do analizy wiarygodności deklaruje znikomy odsetek badanych, dodatkowo jedynie w odniesieniu do niektórych z zasadniczych metod szacowania⁵.

⁴ International Software Benchmarking Standards Group (ISBSG), <http://www.isbsg.org/Isbsg.Nsf/weben/Functional%20Sizing%20Methods> (21.04.2008).

⁵ Rezultaty tych badań zostały przedstawione we wcześniejszej pracy autorki [Czarnacka-Chrobot 2009].

Literatura

- Abran A., Robillard P.N., *Reliability of function points productivity models for enhancement projects (a field study)*, [w:] Conference on Software Maintenance 1993-CSM-93, Montreal, IEEE Computer Society Press, Los Alamitos 1993, s. 80-97.
- Acquiring Custom-Build Software Policy*, Government of Victoria, Melbourne, Australia, August 2000.
- Blattberg R.C., Hoch S.J., *Database models and managerial intuition: 50% model + 50% manager*, „Management Science” 1990, vol. 36, no. 8, s. 887-899.
- CHAOS Summary 2008*, Standish Group, West Yarmouth, Massachusetts, 2008.
- Chulani S., Steece B., *COCOMO II calibration*, [w:] Proceedings of the Fifteenth International Forum on COCOMO and Software Cost Modeling, University of Southern California, Los Angeles, CA, October 24-27, 2000, s. 3-19.
- Clark B., Yang Y., *COCOMO II calibration status*, [w:] Proceedings of the 19th International Forum on COCOMO and Software Cost Modeling, University of Southern California, Los Angeles, CA, October 26-29, 2004, s. 1-19.
- Czarnacka-Chrobot B., *Usage of the software projects scope estimation methods by Polish MIS Providers*, [w:] *Information Management*, B. Kubiak (ed.), The 9th International Conference on Information Management, Gdańsk University Press, Gdańsk 2009, s. 229-238.
- Czarnacka-Chrobot B., Kobyliński A., *Ocena miar zakresu produktu programowego*, [w:] *Informatyka Ekonomiczna. Aspekty informatyzacji organizacji*, red. A. Nowicki, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 55, Wrocław 2009, s. 76-89.
- Devnani-Chulani S. i in., *Calibration Approach and Results of the COCOMO II Post-Architecture Model*, International Society of Parametric Analysts, June 1998.
- Ferens D.V., Christensen D.S., *Does calibration improve predictive accuracy?*, „CrossTalk. The Journal of Defence Software Engineering”, April 2000, s. 14-17.
- Ferens D.V., Gurner R.B., *An evaluation of tree function point models for estimation of software effort*, [w:] Proceedings of the IEEE National Aerospace and Electronics Conference (NAECON), vol. 2, 1993, s. 635-642.
- Furey S., *Why we should use function points*, „IEEE Software” vol. 14, no. 2, March 1997, s. 28-30.
- Garnsey M. i in., *COCOMO and SCORM: Cost estimation model for Web based training*, [w:] Interservice/Industry Training, Simulation, and Education Conference, Orlando, Florida, December 4-7, 2006.
- Gontarz A., *Ile kosztuje program?*, „Computerworld” 2003, nr 11 (wydanie elektroniczne).
- Goodwin P., *Improving the voluntary integration of statistical forecasts and judgment*, „International Journal of Forecasting” 2000, vol. 16, no. 1, s. 85-99.
- Heemstra F.J., *Software cost estimation*, „Information and Software Technology” 1992, vol. 34, no. 10, s. 627-639.
- Hill P.R., *Some practical uses of the ISBSG history data to improve Project Management*, ISBSG, Hawthorn VIC, Australia, 2007.
- IFPUG Function Point Counting Practices Manual, Release 4.2, Part 1-4*, International Function Point Users Group, Princeton Junction, NJ, January 2004.
- Jeffery D.R., Low G.C., *Calibrating estimation tools for software development*, „Software Engineering Journal”, July 1990, s. 215-221.
- Jeffery D.R., Low G.C., Barnes M., *A comparison of function point counting techniques*, „IEEE Transaction on Software Engineering” 1993, vol. 19, no. 5, s. 529-532.
- Jones T.C., *Sizing up software*, „Scientific American Magazine” 1998, vol. 279, no. 6, s. 104 -111.
- Jones T.C., *Software Project Management in the Twenty-first Century*, Software Productivity Research, Burlington 1999.

- Jørgensen M., *Estimation of software development work effort: Evidence on expert judgment and formal models*, „International Journal of Forecasting” 2007, no. 23(3), s. 449-462.
- Kemerer C.F., *An empirical validation of software cost estimation models*, „Communications of the ACM” 1987, vol. 30, no. 5, s. 416-429.
- Kemerer C.F., *Reliability of function points measurement: A field experiment*, „Communication of the ACM” 1993, vol. 36, no. 2, s. 85-97.
- Maddix J.L., *A comparison of function point analysis versus Engineer Experience in Predicting Semiconductor FAB Integration Software Project Durations*, [w:] 20th Computer Science Seminar, SE3-T4-1, Hartford 2004, s. 1-6.
- Sage Technology. Report on the SCUD Methodology Review*, June 2000.
- Shepperd M., Schofield C., *Estimating software project effort using analogy*, „IEEE Transaction on Software Engineering” 1997, SE-23:12, s. 736-743.
- Siddiquee W., *Function point delivery rates under various environments: Some actual results*, [w:] Proceedings of the Computer Management Group’s International Conference, San Diego, CA, December 5-10, 1993, s. 259-264.
- SouthernSCOPE, Reference Manual, Version 1*, Government of Victoria, Melbourne, Australia, September 2000.
- The ISBSG Report: Software Project Estimates – How accurate are they?*, International Software Benchmarking Standards Group, Hawthorn VIC, Australia, January 2005.
- Verner J., Graham T., *A software size model*, „IEEE Transactions on Software Engineering” 1992, vol. 18, no. 4, s. 265-278.
- Vicinanza S., Mukhopadhyay T., Prietula J., *Software-effort estimation: an exploratory study of expert performance*, „Information Systems Research” 1991, vol. 2, no. 4, s. 243-262.
- Whitcotton S.M., Sanders D.E., Norris K.B., *Improving predictive accuracy with a combination of human intuition and mechanical decision aids*, „Organizational Behaviour and Human Decision Processes” vol. 76, no. 3, 1998, s. 325-348.
- Yang Y. in., *Effect of schedule compression on project effort*, [w:] 27th Conference of the International Society of Parametric Analysts, Denver, June 2005, s. 1-14.

RELIABILITY OF THE BSS DEVELOPMENT AND ENHANCEMENT PROJECTS EFFORT ESTIMATION METHODS

Summary: The key requirement for software development and enhancement projects effort estimation methods is reliability of such methods in a specified application area. Business Software Systems (BSS) are the fundamental area of employing information technology. That is why the reliability of methods used for BSS development and enhancement projects effort estimation is the subject of a lot of studies. Their results are different, sometimes even conflicting. So the question arises whether there is a possibility to find on their basis any regularities for BSS development and enhancement projects, which could be a conclusive suggestion for the selection of the most adequate effort estimation method for such projects. This article is aimed at investigating such possible regularities on the basis of the analysis of the studies’ results presented in the subject literature, concerning the reliability of five essential effort estimation methods used for such projects.