

Leonard Rozenberg, Magdalena Kieruzel

Projekty informatyczne typu open source jako źródło danych wykorzystywanych przy ocenie ryzyka

Ekonomiczne Problemy Usług nr 104, 401-406

2013

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

LEONARD ROZENBERG, MAGDALENA KIERUZEL

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

PROJEKTY INFORMATYCZNE TYPU OPEN SOURCE JAKO ŹRÓDŁO DANYCH WYKORZYSTYWANYCH PRZY OCENIE RYZYKA

Wprowadzenie

Artykuł wskazuje na możliwość wykorzystania informacji pochodzących z projektów informatycznych typu open source do oceny ryzyka. Sposób pozyskania i przygotowania danych ma charakter autorski i jest wstępem do oceny ryzyka projektu informatycznego w oparciu o metodę wartości ryzykowanej – VaR (ang. *Value at Risk*).

1. Ocena ryzyka w projektach informatycznych

Intensywny rozwój przemysłu z branży IT zwiększa liczbę równocześnie prowadzonych projektów informatycznych, charakteryzujących się coraz większą złożonością. To z kolei zwiększa zapotrzebowanie na metody służące ocenie ryzyka projektów informatycznych, które ograniczą zagrożenia w prowadzonych przedsięwzięciach i/lub zmniejszą skalę negatywnego oddziaływania tych przedsięwzięć na organizacje realizujące projekty. Na starcie, jak i w trakcie realizacji projektu informatycznego ocenie najczęściej poddawane są koszty projektu, jego czas, a także zakres (element jakości) tworzonego oprogramowania.

Ograniczanie ryzyka poprzez zastosowanie odpowiednich narzędzi, które służyć będą jego ocenie, jest wymogiem dynamicznego rozwoju informatyki jako rodzaju biznesu. Dodatkowo rozwój metod statystycznych, ale także metod uczenia maszynowego, pozwala na analizę obszernych zbiorów danych w poszukiwaniu wzorców niedostępnych dla człowieka, z powodu jego ograniczonej percepcji

i możliwości przetwarzania dużych rozmiarowo danych. Wraz z metodami statystyki, narzędziami baz danych czy algorytmami sztucznej inteligencji, tworzą one dział informatyki zwany *computer science*, a szczególnie dotyczą procesów *eksploracji danych*, które (jak wykazują wyniki badań dostępnych w literaturze) pozwalają na poszukanie rozwiązań dla problemów dotychczas słabo zbadanych czy też trudnych w algorytmizacji.

Metody zarządzania projektami informatycznymi muszą zawierać systemową analizę informacji o przebiegu realizacji projektu. Informacje uzyskiwane w wyniku oceny ryzyka projektu mają najczęściej charakter opisowy, zależny od wiedzy osób (ekspertów), które dokonują oceny ryzyka i wyrażają to ryzyko jako zależność prawdopodobieństwa zajścia określonego rodzaju ryzyka i wpływu tego ryzyka na projekt. Wiarygodność sformułowanych przez ekspertów ocen jest kluczowa dla podejmowanych decyzji w procesie zarządzania projektami. Inną klasą metod wykorzystywanych w procesie szacowania ryzyka są metody wywodzące się z teorii decyzji, których przykładem mogą być techniki drzew zdarzeń (ang. *Events Tree Analysis*) czy drzew błędów (ang. *Faults Tree Analysis*)¹.

Powszechnie wykorzystywane metody oceny ryzyka projektów informatycznych charakteryzują się pewnymi niedoskonałościami, wśród których wskazać należy trudności z jednoznaczną interpretacją wyników prowadzonej analizy. Znacznie dokładniejsze są metody wykorzystywane w analizie ryzyk kredytowych, czy całego obszaru ryzyk gospodarczych. Te procedury oceny ryzyk odznaczają się większym sformalizowaniem, ale także zwiększoną dokładnością interpretacji. Przykładem takich metod może być metoda wartości zdyskontowanej netto (ang. NPV – *Net Present Value*), czy coraz częściej wykorzystywana metoda oceny ryzyka, jaką jest analiza wartości zagrożonej VaR (ang. *Value at Risk*).

Ocena ryzyka w wymiarze wartościowym ma doprowadzić do przejścia od metod opisowych do wyznaczania wartości ryzyka w obszarze projektu informatycznego w postaci ilościowej, co prowadzi do możliwości zobjektywizowania oceny ryzyka. Ocena taka może być przeprowadzona z wykorzystaniem znanych z ekonomii i teorii finansów podejść takich jak wspomniana wcześniej wartość narażona na ryzyko (*Value at Risk*) i może być stosowana na wszystkich etapach realizacji projektu informatycznego, to znaczy zarówno przy planowaniu, jak i realizacji projektu. Wartość narażona na ryzyko może zostać wyznaczona dla projektu informatycznego przy pomocy metody symulacji historycznej. Konieczne jest tutaj wykorzystanie historii zmian pracochłonności poszczególnych zadań składających się na projekt i zastosowanie ich do bieżącej oceny ryzyka projektu.

¹ P. Sienkiewicz: *Analiza ryzyka w zarządzaniu projektami systemów*, www.witu.mil.pl/www/biuletyn/zeszyty/20050095p/9.pdf.

2. Zbiór informacji o projektach informatycznych

Prowadząc ocenę ryzyka konkretnego projektu informatycznego potrzebujemy z jednej strony informacji o tym projekcie, jak również o projektach wcześniej zrealizowanych. W przypadku gdy organizacja dokonująca oceny ryzyka nie dysponuje danymi historycznymi, może potrzebować zewnętrznych informacji dotyczących zrealizowanych już projektów informatycznych. Informacji można poszukiwać u innych organizacji, jednak najczęściej z uwagi na finansowy charakter danych nie zostają one udostępnione. W takim przypadku rozwiązaniem może być pozyskanie informacji z ogólnie dostępnych źródeł. Przykładem jest archiwum SRDA (ang. *SourceForge Research Data Archive*) platformy <http://SourceForge.net>. Znajdują się tutaj projekty informatyczne typu FLOSS (ang. *A Repository of Free/Libre/Open Source Software Research Data*). Dane z grudnia 2012 wskazują na 386 328 zarejestrowanych projektów, z czego 34 040 to projekty aktywne.

3. Opis statystyczny projektów informatycznych platformy SourceForge.net

SourceForge.net jest jedną z największych platform umożliwiających zarządzanie projektami informatycznymi typu open source. Dane z platformy SourceForge.net są przekazywane w postaci zrzutów bazodanowych do archiwum SRDA z wyłączeniem danych osobowych, danych dotyczących funkcjonowania strony SourceForge.net oraz danych, które nie mogą być przekazane ze względów licencyjnych i ze względów bezpieczeństwa. Zakres informacji udostępnianych w archiwum SRDA jest bardzo obszerny. Każdego miesiąca przekazywane są informacje dotyczące projektów, zadań, użytkowników, dane statystyczne opisujące funkcjonowanie platformy SourceForge.net (np. liczba wejść na stronę, liczba pobranych artefaktów itd.) oraz dane historyczne (np. informacje o projektach nieaktywnych). Miesięczny przyrost informacji wynosi ok. 30 GB². Tak ogromny zakres zobrazować może złożoność bazy danych archiwum SRDA opartej na RDBMS (ang. *Relational Database Management System*) PostgreSQL, na którą składa się ponad 100 tabel.

Pozyskanie danych wymaga budowy zapytań SQL do bazy danych SRDA. Mogą to być zapytania złożone z wykorzystaniem funkcji agregujących. Dane, które możemy pozyskać z repozytorium, to:

- dane dotyczące złożoności projektu, które określą liczbę zadań składających się na projekt;

² *Collection of Activity Data for SourceForge Projects*, Scott Christley and Greg Madey, Technical Report: TR-2005-15, Dept. of Computer Science and Engineering, University of Notre Dame, October 10, 2005.

- dane dotyczące wielkości projektu w czasie, które określają czas trwania projektu i czas trwania poszczególnych zadań składających się na projekt;
- dane dotyczące pełnionej roli w procesie wytwarzania, które wskazują na liczbę różnych ról pełnionych w projekcie przez osoby uczestniczące w realizacji projektu;
- dane związane z udziałem w procesie realizacji, które określają liczbę osób uczestniczących w realizacji projektu oraz w realizacji poszczególnych zadań składających się na projekt.

Liczba projektów aktywnych, posiadających co najmniej jedno zadanie (zgodnie z danymi z grudnia 2012), wynosi 19 713, co stanowi 58% liczebności projektów aktywnych i 5% wszystkich projektów. Natomiast maksymalna liczba zadań w projekcie wynosi 550. Analizując dane dotyczące wielkości projektu w czasie, wyróżnić można 6173 projekty, których czas realizacji nie przekracza 10 dni, co stanowi 31% udział wśród projektów, dla których zostały zdefiniowane zadania. Maksymalny czas realizacji projektu wynosi 5216 dni.

W bazie danych platformy SourceForge.net zdefiniowano 29 ról i funkcji, które można przypisać poszczególnym członkom zespołów projektowych. Największą liczbę stanowią deweloperzy, ich udział w projektach informatycznych stanowił 33,12%, następnie kierownicy projektów – 28,16%. Najmniej liczną grupę stanowią analitycy wsparcia technicznego. Maksymalna liczba różnych ról i funkcji, które przypisane są do osób realizujących projekt informatyczny, wynosi 9. Znaczna większość projektów to te, które angażują jedną lub dwie osoby (łącznie stanowią 90% projektów informatycznych). Projekty, do których przypisanych jest 10 i więcej osób, to zaledwie 0,13%.

4. Struktura zbioru projektów informatycznych wykorzystywana do oceny ryzyka

Ocena ryzyka projektów informatycznych wymaga zgromadzenia, a także ustrukturyzowania danych o projektach informatycznych. Stanowi to początkowy etap oceny ryzyka kalkulowanego przy użyciu VaR. Konieczne jest tutaj wyznaczenie zmienności pracochłonności poszczególnych zadań, a dalej wyznaczenie zmienności pracochłonności w obszarze poszczególnych kategorii (grup) zadań projektowych.

Ze względu na dużą liczbę projektów z platformy <http://SourceForge.net> opracowano kryterium wyboru projektów, a następnie przeprowadzono analizę pracochłonności projektów informatycznych.

Spośród wszystkich dostępnych na platformie projektów wybrano te, na które³:

- składało się co najmniej 10 zadań;

³ M. Kieruzel: *Wartościowa ocena ryzyka projektów informatycznych*, praca doktorska, Wydział Informatyki ZUT, Szczecin 2012.

- wszystkie zadania w projekcie były przypisane do poszczególnych osób (członków zespołu);
- w projekcie można było wyróżnić co najmniej cztery role.

56 projektów posiadało powyższe cechy. Próba ta może być wykorzystana jako źródło danych przy bieżącej ocenie ryzyka projektu w dowolnej firmie z branży IT. Przed oceną ryzyka projektów informatycznych w oparciu o wartość ryzykowaną można dokonać pogrupowania pracochłonności poszczególnych zadań projektowych na cztery przykładowe kategorie:

- wytwarzanie,
- analiza i projektowanie,
- zarządzanie,
- wdrożenie i wsparcie.

W każdej z kategorii należy zsumować czasy realizacji poszczególnych zadań, co prowadzi w dalszej kolejności do wyznaczenia struktury każdego z 56 projektów. Następnym krokiem jest wyznaczenie średniej struktury projektu (tabela 1), co z kolei pozwala na określenie odchylenia od tej struktury każdego z wyodrębnionych projektów. Na tej podstawie w dalszej kolejności można również (przy VaR liczoną metodą wariancji–kowariancji) wyznaczyć rozkłady prawdopodobieństwa każdej z czterech kategorii zadań.

Tabela 1

Struktura średniego projektu

	Wytwarzanie	Analiza i projektowanie	Zarządzanie	Wdrożenie i wsparcie
Średni projekt	54%	11%	31%	4%

Źródło: opracowanie własne.

Wyodrębnione w ten sposób dane dają podstawę do zastosowania wartościowej oceny ryzyka projektu informatycznego z wykorzystaniem wartości narażonej na ryzyko VaR.

Podsumowanie

Wybór metody oceny ryzyka projektu jest zależny od danych wejściowych, którymi dysponuje decydent. Kalkulacja ryzyka przy użyciu metody VaR metodą symulacji historycznej czy metodą wariancji–kowariancji wymaga posiadania danych o zrealizowanych już projektach. Nie zawsze jednak organizacja dysponuje danymi historycznymi, stąd analiza ogólnie dostępnych repozytoriów projektów informatycznych wydaje się być skutecznym rozwiązaniem. Ciężar analizy danych leży po stronie

kierownika projektu, który zwykle odpowiedzialny jest za proces zarządzania ryzykiem niezależnie od metodyki, według której projekt jest prowadzony.

Baza projektów informatycznych zarejestrowanych na platformie SourceForge.net daje podstawę do zgromadzenia danych niezbędnych przy ocenie ryzyka w organizacji realizującej projekty informatyczne, chociaż rozwiązaniem lepszym jest stworzenie własnej bazy projektów zrealizowanych. Zaprezentowana analiza repozytorium platformy SourceForge.net jest rozwiązaniem autorskim. Sposób kategoryzacji zadań projektowych może być oczywiście przeprowadzony według innych kryteriów niż proponowane w artykule.

Literatura

1. Balamut T.: *Metody estymacji Value at Risk*, Materiały i Studia NBP, Zeszyt 147, Warszawa 2002.
2. Best P.: *Wartość narażona na ryzyko. Obliczanie i wdrażanie modelu VAR*, Kraków 2000.
3. *Collection of Activity Data for SourceForge Projects*, Scott Christley and Greg Madey, Technical Report: TR-2005-15, Dept. of Computer Science and Engineering, University of Notre Dame, 2005, October 10.
4. Dowd K.: *Beyond Value at Risk. The New science of risk management*, Wiley, Chichester 1998.
5. Jorion P.: *Value at Risk: The new benchmark for managing risk*, third edition, McGraw-Hill 2007.
6. Sienkiewicz P.: *Analiza ryzyka w zarządzaniu projektami systemów*. www.witu.mil.pl/www/biuletyn/zeszyty/20050095p/9.pdf.
7. SourceForge (2012). <http://www.sourceforge.net>.
8. Kieruzel M.: *Wartościowa ocena ryzyka projektów informatycznych*, praca doktorska, Wydział Informatyki ZUT, Szczecin 2012.

OPEN SOURCE PROJECTS AS A SOURCE OF DATA NEEDED IN RISK ASSESSMENT

Summary

The article presents the possibility of using data from Open Source IT projects for risk evaluation. The way of preparing the data was proposed by an author and it is only an introduction to risk assessment based on the Value at Risk method.

Translated by Magdalena Kieruzel