

**Daniel Czyczyn-Egird**  
**Rafał Wojszczyk**  
Katedra Inżynierii Komputerowej  
Wydział Elektroniki i Informatyki  
Politechnika Koszalińska  
ul. J.J. Śniadeckich 2  
75-453 Koszalin

## **Zastosowanie technik eksploracji danych na przykładzie badania popularności wzorców projektowych w serwisie społecznościowym Stackoverflow.com**

**Słowa kluczowe:** sieci społecznościowe, eksploracja danych, wzorce projektowe

### **1. Wstęp**

Od wieków ludzie poszukiwali sposobów komunikacji na odległość. Każdy krok postępu technologicznego, często związany z wieloma dziedzinami nauki, sprzyjał łatwiejszej wymianie danych i informacji. Sposoby komunikacji zmieniały się z biegiem lat, pierwsze urządzenia elektromechaniczne np. telegraf umożliwiały przesyłanie niewielkiej ilości danych na odległość. Kolejne lata przyniosły wzrost przepustowości i zakresu przesyłanych danych, włączając w to dźwięk i wizję, aż po współczesny Internet. Podobna historia ewolucji spotkała pionierów, którzy dali początek współczesnemu programowaniu. Niewątpliwie pierwszym sposobem komunikacji na odległość pomiędzy praktykami programowania było zwyczajne pisanie listów, ewentualnie wymiana ówczesnych nośników danych. Dopiero dalsze lata i pojawienie się Internetu pozwoliło na komunikację: poprzez pocztę email, grupy usenet oraz publikowanie informacji na stronach internetowych (w przeszłości na stronach domowych, współcześnie popularne są blogi). Wymienione sposoby komunikacji nadal dotyczą wymiany danych, pomijając przy tym szybkość i dostępność łączy internetowych, warto zastanowić się nad sposobem wykorzystania – już nie danych a informacji. W tym przypadku kierunkiem rozwoju wykorzystania Internetu jest tzw. Web 2.0 [12]. Popularne serwisy społecznościowe bazują na założeniach Web 2.0 i znacząco wpływają na szybkość i skalę rozpowszechnianych informacji. Angażując do tego podstawowe założenia sieci semantycznych (poprzez oznaczanie informacji metadanymi

w postaci tagów) można otrzymać połączenie dające więcej niż dostęp do informacji, wynikiem tego połączenia jest wiedza pozyskiwana z eksploracji danych.

Współczesne techniki informatyczne nie istniałyby gdyby nie rozwój rzemiosła programowania. Programowanie nieodłącznie wiąże się z korzystaniem z różnych wzorców: architektonicznych (łączy różne warstwy aplikacji), projektowych (obejmują zakres poziomu klas), implementacyjnych (nazywane również idiomami, występują na poziomie wierszy kodu). Wzorce projektowe przedstawione w [6] powstały na podstawie doświadczenia społeczeństwa, które wówczas zajmowało się zagadnieniami programowania obiektowego i dobrych praktyk. Zasadniczym przeznaczeniem wzorców jest rozwiązywanie problemów projektowania i programowania obiektowego [14]. Wzorce projektowe są uznawane również za pewnego rodzaju język komunikacji. Język ten sprawia, że kod programu, który zawiera wzorce będzie łatwiejszy w poznaniu i zrozumieniu niż ten, który ich nie zawiera. Zatem osoby znające wzorce samoistnie tworzą społeczność porozumiewając się językiem wzorców [1].

Celem artykułu jest zbadanie popularności oraz tendencji zmian wykorzystania wzorców projektowych, w oparciu o wyspecjalizowane sieci społecznościowe. Wyniki badania będą wykorzystane do wskazania kierunku dalszych prac nad jakością implementacji wzorców projektowych, natomiast metodyka przeprowadzonych badań przyczyni się do wybrania kierunku prac w dziedzinie data miningu.

W drugim rozdziale artykułu przedstawiono przegląd wybranych zagadnień związanych z wykorzystaniem sieci komputerowych do wspierania pracy programistów oraz opisu informacji o wzorcach projektowych. Trzeci rozdział zawiera opis środowiska badawczego oraz uzasadnienia dokonanych wyborów. W rozdziale czwartym przedstawiono wyniki badań oraz stosowne wnioski. Ostatni, piąty rozdział, stanowi podsumowanie pracy.

## **2. Przegląd rozwiązań sieciowych**

### **2.1. Ogólne wsparcie deweloperów**

W dzisiejszych czasach do dyspozycji deweloperów (programistów) oddawane są cały czas nowe platformy z dokumentacją produktowo-techniczną, które charakteryzują się obszerną bazą informacji przydatnych podczas projektowania nowych rozwiązań informatycznych. Wiele firm dostarczających własne rozwiązania informatyczne oraz produkty świadomie i celowo stara się udostępnić wszystkim zainteresowanym dobrze udokumentowane biblioteki czy też interfejsy programistyczne. Takie działanie pozwala osiągnąć pewne korzyści, dokładniej zapewnia produktom odpowiednią reklamę oraz szerokie rozpowszechnienie na rynku, chociażby przez zapewnienie swoim odbiorcom możliwości tworzenia

spersonalizowanych dodatków czy nowych funkcji do już istniejących systemów komputerowych.

W XXI wieku ilość dokumentacji w formie papierowej drastycznie maleje na rzecz dokumentacji online, która dzięki powszechnemu dostępowi do globalnej sieci internetowej, jest rozpowszechniona na cały świat i dostępna przez całą dobę.

Deweloperzy wyposażeni w dostęp do globalnej sieci internetowej są w stanie w łatwy sposób dotrzeć do opisów technicznych, obliczeń konstrukcyjnych, planów, rysunków, harmonogramów i kosztorysów. Dzięki powszechnemu dostępowi do dokumentacji mogą powstawać nowe rozwiązania, bez potrzeby bezpośredniego kontaktu dewelopera z firmą udostępniającą produkt.

Dodatkowym elementem wspierającym pracę deweloperów są sieci społecznościowe, które pozwalają na wspólne analizowanie dokumentacji technicznych a następnie zaproponowanie nowych rozwiązań. Gotowe półprodukty, pakiety oraz biblioteki, udostępnione są w sieciowych repozytoriach, które łączą deweloperów poprzez wymianę zaufanych fragmentów oprogramowania. Jednym z konkretnych narzędzi, o którym warto wspomnieć w kontekście pracy z repozytoriami tego typu, jest NuGet. Jest to tak zwany menedżer pakietów lub też systemem zarządzania pakietami, dla platformy deweloperskiej firmy Microsoft, w tym platformy .NET. Narzędzia klienckie NuGet'a umożliwiają wygodne korzystanie z pakietów lub też przygotowanie własnych, a całość w łatwy sposób integruje się ze środowiskiem programistycznym Visual Studio. Zawartość pakietów zawartych w repozytorium NuGet'a jest stale aktualizowana przez autorów pakietów zrzeszonych w sieci społecznościowej. Podobnymi rozwiązaniami dostępnymi na rynku są na przykład: GetIt dla narzędzi deweloperskich z rodziny Delphi, czy Maven dla Javy.

Sieciami społecznościowymi możemy także określić takie platformy jak Github, CodePlex czy też SourceForge, które są portalami gromadzącymi oraz przechowującymi projekty informatyczne. Udostępniają one darmowy hosting programów open source oraz płatne prywatne repozytoria. Portale tego typu służą deweloperom także do wymiany informacji oraz kodów źródłowych, przyczyniając się tym samym do rozwoju oprogramowania, zwłaszcza na licencji otwartej i bezpłatnej.

Spółeczność połączona w sieć przyczyniła się także do rozwoju i spopularyzowania repozytoriów plikowych dostępnych online, jako systemy kontroli wersji. Takimi narzędziami są przykładowo Git oraz SVN, których przeznaczeniem jest m.in. śledzenie zmian w kodzie źródłowym oraz zapewnienie pomocy programistom w łączeniu różnych wersji plików, edytowanych przez wielu deweloperów w różnych momentach czasowych. Systemy kontroli wersji można podzielić na:

- lokalne, umożliwiające zapisanie danych wyłącznie na lokalnym komputerze (np. SCCS oraz RCS),
- scentralizowane, cechujące się architekturą klient-serwer (np. CVS, SVN),
- rozproszone, oparte na architekturze peer-to-peer (np. BitKeeper, Git).

Pierwszy rodzaj systemów zapisuje jedynie wersje plików na lokalnym komputerze. Jest to bardzo wygodne, jednak mało bezpieczne rozwiązanie (w sensie ochrony przed utratą danych) oraz blokujące możliwość dzielenia się swoimi kodami z innymi deweloperami. W przypadku rozwiązań scentralizowanych istnieje jedno centralne repozytorium, za pomocą którego wszyscy użytkownicy systemu synchronizują swoje zmiany. Rozwiązania rozproszone pozwalają na równoczesne prowadzenie niezależnych, ale też równoprawnych gałęzi, które można wzajemnie synchronizować, np. poprzez pocztę elektroniczną. Typ wykorzystywanego systemu kontroli wersji zależy od założeń projektu i powinien być dopasowany do potrzeb deweloperów.

Doskonałym wykorzystaniem sieci społecznościowych są także fora internetowe zrzeszające użytkowników, których zainteresowania pokrywają się z tematyką owych portali. Przykładowo, największe w Polsce forum programistyczne [4programmers.net](http://4programmers.net) przyciąga programistów, administratorów, webmasterów, słowem – ludzi związanych z branżą IT. Tysiące tematów i komentarzy stanowią o sile tego typu portali, gdzie użytkownicy mogą zakładać nowe tematy a zainteresowani mogą pisać nowe komentarze dotyczące zadanej sprawy. Wraz z upływem czasu pojawiło się zapotrzebowanie na portale jeszcze bardziej ściśle techniczne, gdzie dzielenie się wiedzą przez specjalistów opierało się na schemacie Q&A, czyli „pytanie-odpowiedź”. Jednym z takich portali jest polski serwis [devpytania.pl](http://devpytania.pl), który umożliwia zadawanie pytań czy też udzielanie odpowiedzi do istniejących pytań. Obecnie w bazie portalu jest blisko 4000 pytań. Zdecydowanie lepiej jest skorzystać z portalu o zasięgu światowym jakim jest serwis [stackoverflow.com](http://stackoverflow.com). Baza pytań tego serwisu jest imponująca i sięga blisko 11 milionom, a także odpowiedzi do pytań udzielane są przez specjalistów z całego świata. Serwisy tego typu oferują pomoc w rozwiązywaniu problemów, które są napotykanne przez deweloperów każdego dnia, dlatego ich rosnąca popularność nie powinna dziwić. Niniejsza publikacja skupia się także na analizie i eksploracji danych [4] z wyżej wspomnianego portalu. Dokładniej na przebadaniu popularności wzorców projektowych w odniesieniu do wiedzy zgromadzonej w bazie danych serwisu, który w 2008 roku zaistniał w Internecie. Do tej pory serwis zbudował ogromną sieć społecznościową, która przyczynia się do ciągłego rozwoju portalu – na chwilę obecną pytających i odpowiadających użytkowników jest razem około 4,8 miliona.

## 2.2. Informacje o wzorcach projektowych

Podstawową literaturą dotyczącą wzorców projektowych jest wymienione wcześniej [6]. Na katalogu wzorców [6] bazują często inni autorzy proponując inne warianty lub implementacje dopasowane do konkretnego języka, np. [11], [3]. Forma opisu wzorców projektowych przedstawiona w wymienionej literaturze jest z założenia przeznaczona do nauki ich wykorzystania, zawiera: opis słowny w języku naturalnym, diagramy klas oraz przykładowy kod implementacji oparty o proste przykłady. Podobny cel jest realizowany przez repozytorium wiedzy o wzorcach projektowych przedstawione w [9]. Celem tego repozytorium jest szerzenie wiedzy o wzorcach, nauka oraz wsparcie w ich implementacji. Autorska aplikacja internetowa z [9] umożliwi dialog pomiędzy użytkownikiem a systemem, na zasadzie pytanie-odpowieź. Wiedza o wzorcach oraz pytania są wstępnie predefiniowane z możliwością dodania nowych zasobów.

Niektóre z formalnych sposobów reprezentacji wzorców projektowych opierają się o sieci semantyczne, dokładniej wykorzystują ontologie [8], [2], [5]. W filozofii ontologia opisuje byty świata rzeczywistego [10], natomiast w informatyce jest wykorzystywana do formalnej reprezentacji wiedzy w postaci zbioru pojęć z danej dziedziny i relacji między tymi pojęciami. Bardzo ważną zaletą ontologii jest możliwość transformacji danych „z” oraz „do” różnych źródeł. W [2] oraz [5] zostały zaproponowane ontologie opisujące podstawową wiedzę o wzorcach projektowych. Wiedza ta może być przetransformowana do innych podejść, np. opisane wcześniej [9] czy [8] służącego do wyszukiwania wystąpień wzorców projektowych w kodzie źródłowym.

## 3. Przygotowania do badań

### 3.1. Portal [stackoverflow.com](https://stackoverflow.com)

Serwis [stackoverflow.com](https://stackoverflow.com) jest jednym z największych portali typu Q&A z zakresu szeroko pojętych tematów związanych z inżynierią oprogramowania. Jest to platforma pozwalająca użytkownikom zadawać pytania i uzyskiwać na nie odpowiedzi. Społeczność zgromadzona na portalu ma możliwość oceniania „w dół” lub „w górę” zadawanych pytań i udzielanych odpowiedzi, jest to rodzaj naturalnej selekcji, w której odrzucane są „złe” rozwiązania a wyróżniane „dobre”. Użytkownicy oddają głosy na podstawie posiadanego stanu wiedzy i preferencji, dlatego należy pamiętać, że te wybory mogą nie być obiektywne. Jednakże głos społeczności jest bardzo dobrym wskaźnikiem popularności, co zostało wykorzystane w dalszych badaniach. Użytkownicy za swoją działalność są odpowiednio nagradzani punktami lub odznakami, które służą do wewnętrznej, co ważne – pozytywnej rywalizacji między nimi. Cała zawartość wygenerowana przez użytkowników jest oparta na licencji Creative Commons.

Wybór serwisu stackoverflow.com jako źródła informacji jest uzasadniony popularnością zmierzoną przez Google Trends oraz wykorzystaniem w innych badaniach [13]. Rysunek 1 przedstawia wykres popularności stackoverflow.com w stosunku do bezpośredniego konkurenta experts-exchange.com.



**Rys. 1.** Wykres popularności konkurencyjnych serwisów, linia niebieska to popularność zapytań stackoverflow.com, czerwona experts-exchange.com, stan na 06.01.2016.

### 3.2. Wzorce projektowe jako przedmiot badań

Wzorce projektowe należy traktować jako uniwersalne i sprawdzone w praktyce rozwiązania często pojawiających się problemów projektowych. Pokazują występujące powiązania oraz zależności klas i obiektów, ułatwiają również tworzenie, modyfikację i utrzymanie kodów źródłowych systemów informatycznych. Stosowanie wzorców w małych projektach zazwyczaj przyczynia się zwiększenia pracochłonności procesu wytwórczego. Natomiast w rozbudowanych projektach poprawia ogólną jakość tworzonych systemów oraz ułatwia konserwację. Wnioskując na podstawie korzyści wynikających z wykorzystania wzorców projektowych [15] należałoby założyć, że popularność wzorców powinna cały czas wzrastać. W dalszej części publikacji, zostały przedstawione wyniki badań nad popularnością wzorców projektowych wypromowanych przez tzw. Bandę Czworka (Erich Gamma, Richard Helm, Ralph Johnson oraz John Vlissides) [6]. Badanie zostało zrealizowane poprzez analizę tematów zebranych w portalu stackoverflow.com, na podstawie wystąpień słów kluczowych związanych ze wzorcami projektowymi, z wykorzystaniem technik data mining.

### 3.3. Przebieg badań

Badania zostały przeprowadzone na trzech odrębnych zbiorach danych, które zostały przygotowane przy pomocy autorskiej aplikacji pobierającej informacje. Następnie zbiory zostały odpowiednio przetworzone, w celu uzyskania konkretnych odpowiedzi oraz reprezentacji ich za pomocą określonych miar.

W celu wydobycia konkretnych danych z portalu stackoverflow.com został wykorzystany webowy interfejs programistyczny StackExchange API w najnowszej wersji 2.2. Za pomocą odpowiednich metod interfejsu możliwe było budowanie

sparametryzowanych zapytań do serwera, a następnie uzyskiwanie konkretnych odpowiedzi spełniających zadane kryteria. Rysunek 2 przedstawia przykładowe zapytanie interfejsu, w tym wypadku poszukiwanie informacji dotyczącej wzorca projektowego „singleton”.

### StackExchange 2.2

<https://api.stackexchange.com/2.2/tags?order=desc&sort=popular&site=stackoverflow&inname=singleton>

### Rys. 2. Przykładowe zapytanie interfejsu API

Wszystkie odpowiedzi zwrotne były otrzymywane w formacie JSON [7]. Format ten świetnie nadaje się do wymiany danych pomiędzy usługami programistycznymi, ponieważ jest stosunkowo lekki oraz łatwy do dalszego przetwarzania i obróbki. Rysunek 3 przedstawia przykładowe dane wyjściowe programu w formacie JSON.

#### JSON

```
{
  "tags": ["javascript", "node.js"],
  "owner": {
    "reputation": 374,
    "user_id": 3278897,
    "user_type": "registered",
    "profile_image": "https://www.gravatar.com/avatar/f3b5e891b94e55378996640c52bad84c?s=128&d=identicon&r=PG&f=1",
    "display_name": "user3278897",
    "link": "http://stackoverflow.com/users/3278897/user3278897",
    "is_authorized": true,
    "view_count": 56,
    "answer_count": 2,
    "score": 3,
    "last_activity_date": 1451975389,
    "creation_date": 1451959060,
    "last_edit_date": 1451975389,
    "question_id": 34603040,
    "link": "http://stackoverflow.com/questions/34603040/javascript-node-js-best-way-to-create-singleton-object",
    "title": "Javascript / Node JS best way to create singleton object"
  }
}
```

### Rys. 3. Przykładowa odpowiedź serwera w formacie JSON

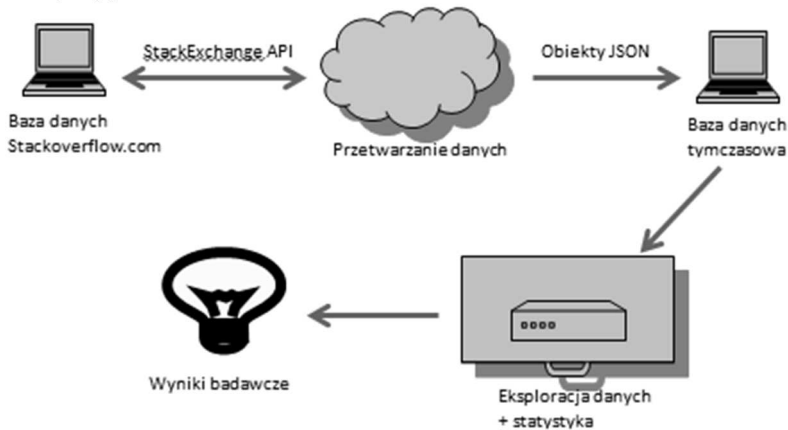
Dane, przetwarzane za pomocą autorskiego oprogramowania, zostały umieszczone jako rekordy w bazie danych, rysunek 4 przedstawia schemat ideowy środowiska badawczego. Warto zauważyć, że API znacząco ułatwia dostęp do bazy informacji w portalu. Niestety posiada również pewne ograniczenia dla deweloperów. Serwis wykorzystuje walidację adresu IP tj. w ciągu jednej doby można wywołać około 300 zapytań, które zwrócą jednorazowo maksymalnie 100 obiektów zwrotnych. W przypadku chęci pobrania jednorazowym zapytaniem zbioru pytań dotyczących wybranego słowa kluczowego, możliwe jest uzyskanie maksymalnie 100 wyników. Samo preparowanie zapytania sprowadza się do odpowiedniego zarządzania parametrami oraz filtrami, np. możliwe jest odfiltrowanie danych po dacie utworzenia, wyrażeniach występujących w tytule pytań czy też jako słowa kluczowe, a także sortowanie według ilości odpowiedzi czy popularności.

Najpoważniejsza niedogodność dotycząca aktualnej wersji API jest związana z brakiem możliwości przeszukiwania bazy serwisu według wystąpienia zadanego

hasła jednocześnie w tytule oraz w treści pytania. Problem jest dość istotny, ponieważ wstępne rozeznanie pokazało, że niektórzy użytkownicy nie zawsze podają konkretne sformułowanie w tytule pytania lub jako słowo kluczowe (tag) – szukana fraza występuje dopiero w treści pytania, do którego nie ma obecnie dostępu z poziomu API. Rozwiązaniem tego problemu jest manualne wykorzystanie wyszukiwarki serwisu, która w wyszukiwaniu uwzględnia treść pytań. Niestety to rozwiązanie jest nie do przyjęcia w postawionym celu, dlatego w ten sposób zostało wykonane tylko jedno badanie. W pozostałych przypadkach wyszukiwanie odbywało się wyłącznie po obecności wybranej frazy w tytule pytania, gdyż okazało się, że wyszukiwanie po słowach kluczowych (tagach) było nieskuteczne – pomijało zbyt dużą ilość pytań.

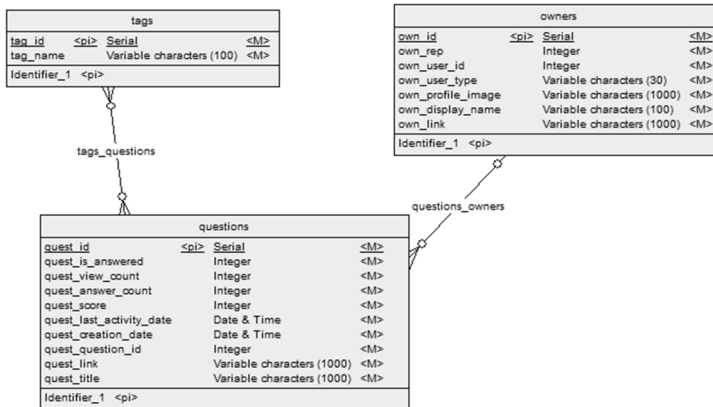
Wszystkie uzyskane wyniki zostały przetworzone i umieszczone w relacyjnej bazie danych PostgreSQL 9.3, dzięki której możliwa jest łatwa eksploracja i klasyfikacja danych [4]. Rysunek 5 przedstawia model danych, który został zaprojektowany w taki sposób, aby zapewnić łatwy dostęp do danych oraz ich bezproblemowe zrozumienie.

Encje odpowiadają obiektom takim jak: pytania (question), autorzy pytań (owners) oraz słowa kluczowe (tags). Każda encja zawiera odpowiednie pola, które ją charakteryzują.



Rys. 4. Schemat środowiska badawczego



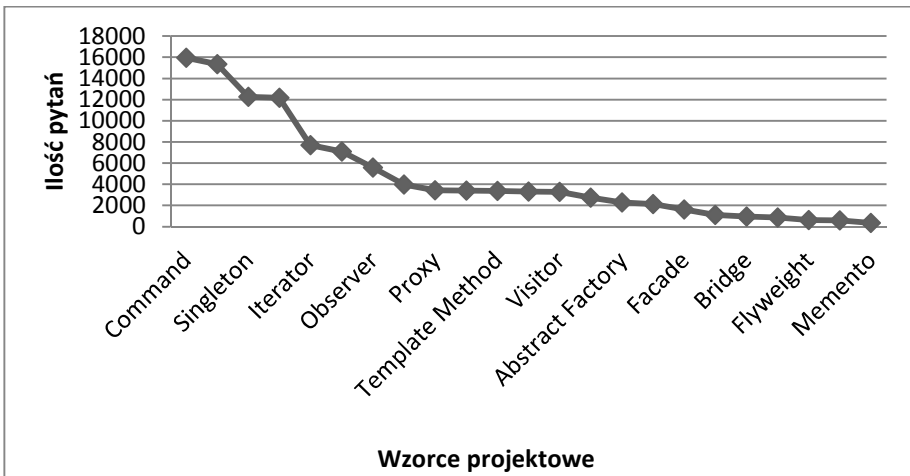


Rys. 5. Model encja-związek bazy danych dla obiektów z portalu stackoverflow.com

## 4. Wyniki badań

### 4.1. Ogólna popularność wzorców projektowych

Pierwszy zbiór danych został zasilony poprzez manualne zadanie pytań w wyszukiwarce serwisu. Uzyskany wynik, wyrażony w postaci wykresu liniowego, przedstawia rysunek 6.



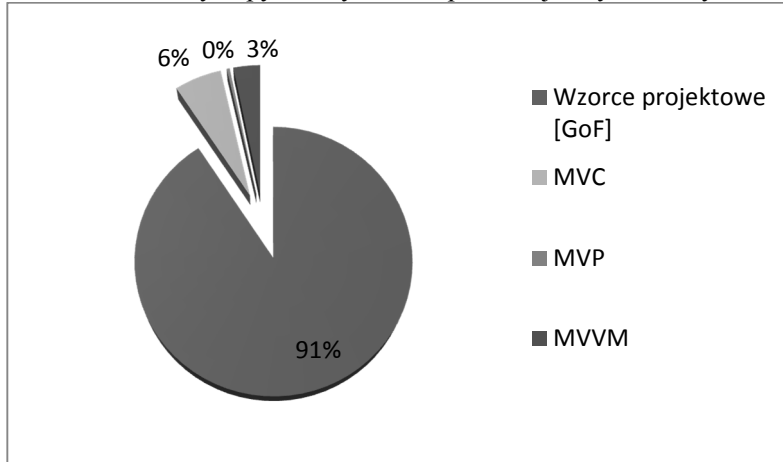
Rys. 6. Wykres liniowy pokazujący ilości pytań dla poszczególnych wzorców projektowych w serwisie stackoverflow.com, wraz z linią trendu (średnia ruchoma), stan na okres grudzień 2015

Analizując otrzymane wyniki można zauważyć dominującą grupę czterech wzorców: Command, State, Singleton oraz Factory. Badanie ogólnej popularności nie dostarcza dodatkowych informacji, które wynikałyby z kontekstu zadawanych pytań w serwisie stackoverflow.com. Wysoka popularność oznacza, że programiści podejmują próby implementacji danych wzorców projektowych, jednakże napotykają w tym problemy lub ich wiedza jest niewystarczająca. Niska popularność wskazuje na mniejszą potrzebę stosowania danych wzorców lub też łatwą implementację. Aczkolwiek zaprzecza temu popularność wzorca Singleton, który jest uznawany za jeden z najłatwiejszych. W celu doszczegółowienia wniosków zostały wykonane następne badania.

#### 4.2. Względna popularność wzorców projektowych

Wzorce projektowe należą do tzw. dobrych praktyk, które często są tworzone na podstawie doświadczenia społeczeństwa. Toteż uzasadnione jest porównanie aktualnej popularności wzorców projektowych [6] względem podobnych dobrych praktyk, w tym przypadku do wzorców architektonicznych [9] takich jak MVC, MVP oraz MVVM.

Drugi zbiór danych składał się z 13000 pytań, które pojawiły się w grudniu 2015 roku. Pytania zostały pobrane z serwisu za pomocą API. Bazy danych została zasilona z pominięciem duplikatów, co poskutkowało zmniejszeniem rozmiaru zbioru do 12095 unikalnych pytań. Rysunek 7 prezentuje uzyskane wyniki.



Rys. 7. Diagram kołowy obrazujący rozkład pytań dla wzorców projektowych [6] oraz architektonicznych

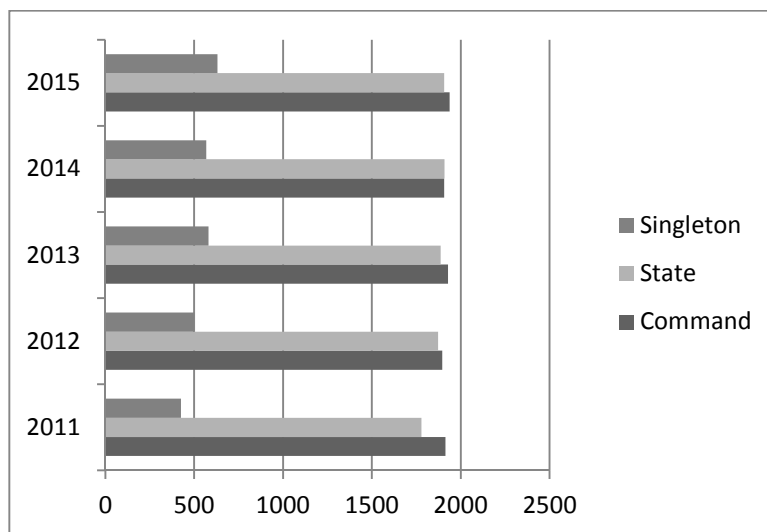
Otrzymany wynik wskazuje na znacząco większą popularność względem podobnych dobrych praktyk. Przyczyną tego może być fakt, że wzorce projektowe przy każdej implementacji wymagają zrozumienia oraz własnego wkładu od

programisty. Popularne środowiska programistyczne nie oferują automatycznej implementacji wzorców. Przeciwna sytuacja występuje wobec wzorców architektonicznych, w przypadku których zautomatyzowane narzędzia integrują się ze środowiskami programistycznymi, często niezauważalnie dla dewelopera. Zatem brak narzędzi automatyzujących implementację badanych wzorców wzmagą zainteresowanie tym tematem.

### 4.3. Popularność wzorców projektowych w funkcji czasu

Celem ostatniego badania opierającego się o miary popularności było wykazanie tendencji zmian dla wybranych wzorców w zadanej funkcji czasu.

Zbiór danych, przygotowany do tego badania, składał się z pytań pobranych dla trzech najpopularniejszych wzorców projektowych wykazanych we wcześniejszych badaniach, tj. dla: Command, State, Singleton. Dla każdego wzorca zostały pobrane pytania od początku roku 2011 aż do końca 2015 roku. Początkowa suma wynosiła zatem blisko 25000 pytań, jednak po odrzuceniu duplikatów pozostało 21640 pytań w bazie danych. Wyszukiwanie odbywało się ponownie według kryterium obecności wybranej frazy w tytule pytania. Uzyskane wyniki przedstawia rysunek 8.



Rys. 8. Popularność wybranych wzorców w funkcji czasu

Analiza uzyskanych wyników pozwala na wyciągnięcie następujących wniosków:

- wybrane do badania wzorce utrzymują stałą tendencję popularności, nieznaczne wahania wynikają z różnych czynników losowych, których występuje bardzo wiele w aktywnych społecznościach,

- stosunek popularności pomiędzy poszczególnymi wzorcami jest również stały, co potwierdza wynik badania przedstawiony przez miarę ogólnej popularności,
- można przewidywać, że popularność wybranych wzorców projektowych nie zmniejszy się w 2016 roku.

## 5. Podsumowanie

W pracy przedstawiono wyniki badań popularności wzorców projektowych. W opracowaniu wyników zostały wykorzystane techniki data miningu, wspierane dodatkowo przez autorskie oprogramowanie. Dane do badań zostały pozyskane z serwisu stackoverflow.com, który jest jednym z przykładów sieci społecznościowych. Ważną zaletą wykorzystania tego rodzaju sieci jest fakt, że społeczność programistów obligatoryjnie dokona selekcji treści pojawiających się w serwisie.

Przeprowadzone badania dotyczą: ogólnej popularności wzorców projektowych, popularności wzorców względem podobnych dobrych praktyk oraz próby przewidywania popularności w 2016 na podstawie wcześniejszych tendencji. Uzyskane wyniki pokazują, że najpopularniejszymi wzorcami projektowymi są: Command, State, Singleton oraz Factory. Na przykładzie wymienionych wzorców wykazano stałą popularność na przestrzeni 5 ostatnich lat, co pozwala przewidywać, że popularność pozostanie bez zmian. Również w przypadku popularności względem wzorców architektonicznych wykazano, że wzorce projektowe [6] są popularniejsze.

Dalsze prace w zakresie technik eksploracji danych przewidują rozbudowę metody i wykorzystanie narzędzi Business Intelligence. Następnie przewidywana jest zmiana przedmiotu badań na inne zagadnienia inżynierii oprogramowania. Wyniki otrzymane w powyższych badaniach przyczyniają się do wyboru kierunku dalszych prac w zakresie oceny jakości implementacji wzorców projektowych.

## BIBLIOGRAFIA

1. Alexander C.: *Język wzorców. Miasta, budynki, konstrukcja*, GWP Gdańskie Wydawnictwo Psychologiczne, Sopot 2008
2. Alnusair A., Zhao T., Yan G., *Rule-based detection of design patterns in program code*, International Journal on Software Tools for Technology Transfer, Springer Berlin Heidelberg, 2013
3. Bernd Bruegge, Allen H. Dutoit, *Inżynieria oprogramowania w ujęciu obiektowym*. UML, wzorce projektowe i Java, Helion

4. Czyczyn-Egird D.: *Eksploracja danych na podstawie szacowania informacji z wykorzystaniem regresji liniowej*, w: Modele inżynierii teleinformatyki 10, Wydawnictwo Uczelniane Politechniki Koszalińskiej, ISBN 978-83-7365-365-8, Koszalin 2015
5. Dietrich J., Elgar C., *A formal description of design patterns using OWL*, Software Engineering Conference, Australian, 2005
6. Gamma E. i inni, *WZORCE PROJEKTOWE Elementy oprogramowania wielokrotnego użytku*, Helion, Gliwice, 2010
7. Kasprowski P.: *Choosing a persistent storage for data mining task*, Studia Informatica, Vol. 33, No. 2B, 2012
8. Kirasić D., Basch D., *Ontology-Based Design Pattern Recognition*, Knowledge-Based Intelligent Information and Engineering Systems, Zagreb, Croatia, 2008
9. Pavlic L. i inni, *Improving design pattern adoption with Ontology-Based Design Pattern Repository*, Informatica An International Journal of Computing and Informatics, Vol 33, Ljubljana, Slovenia, 2009
10. Plecka P., Bzdyra K.: *Wykorzystanie ontologii w wymiarowaniu projektów informatycznych*, w: Innowacje w Zarządzaniu i Inżynierii Produkcji, Tom I, strony od 230 do 240, Politechnika Opolska, Opole 2014
11. Steven John Metsker, *C#. Wzorce projektowe*, Helion
12. Szewczyk A., *Popularność funkcji serwisów społecznościowych*, Studia Informatica 2011
13. Wang S., Lo D., Jiang L.: *An Empirical Study on Developer Interactions in StackOverflow*, Research Collection School Of Information Systems, Singapore 2013
14. Wojszczyk R.: *Porównanie sposobów reprezentacji wzorców projektowych*, w: Modele inżynierii teleinformatyki 9, strony od 133 do 145, Wydawnictwo Uczelniane Politechniki Koszalińskiej, ISSN 2353-6535, Koszalin 2014
15. Wojszczyk R.: *The model and function of quality assessment of implementation of design patterns*, w: Applied Computer Science, Vol. 11, No. 3, strony od 44 do 55, Politechnika Lubelska, ISSN 1895-3735, Lublin 2015

## Streszczenie

Idea sieci społecznościowych jest znana od wielu lat. Jednak dopiero od niedawna nabrały nowego znaczenia, do czego przyczyniła się popularność współczesnych serwisów społecznościowych. Generowana treść przez społeczności jest ogromnym zasobem wiedzy do przeanalizowania. W artykule przedstawiono wyniki badań nad popularnością wzorców projektowych w oparciu o dane zgromadzone w wyspecjalizowanych sieciach społecznościowych. Wyniki badań uzyskano poprzez wykorzystanie technik eksploracji danych.

**Abstract**

The idea of social networks has been known for many years. However, only recently took on a new meaning, which was due to the popularity of today's social networks. Social service user-generated content constitutes tremendous stores of knowledge to be analysed. The article presented results of research on the popularity of design patterns on the basis of data gathered in the specialised social networks. The research results were obtained thanks to using data mining techniques.

**Keywords:** social networks, data mining, design patterns