**AKADEMIA GÓRNICZO-HUTNICZA**

**im. Stanisława Staszica w Krakowie**



**Informatyczne Narzędzia Pracy Grupowej**

**Projekt programu analizującego dane**

**Konrad Janczarski**

**Michał Pilch**

**Kamil Switek**

**WEAIiIB, AiR**

**Rok I**

Spis treści

[Cel projektu 3](#_Toc452860429)

[Scrum 3](#_Toc452860430)

[Sprint 1 3](#_Toc452860431)

[Cel 3](#_Toc452860432)

[Milestones 3](#_Toc452860433)

[Tickety i subtickety 3](#_Toc452860434)

[Problemy 3](#_Toc452860435)

[Podsumowanie 4](#_Toc452860436)

[Sprint 2 4](#_Toc452860437)

[Sprint 3 4](#_Toc452860438)

[Sprint 4 4](#_Toc452860439)

[Sprint 5 4](#_Toc452860440)

[Cel 4](#_Toc452860441)

[Milestones 4](#_Toc452860442)

[Tickety i subtickety 4](#_Toc452860443)

[Problemy 5](#_Toc452860444)

[Podsumowanie 5](#_Toc452860445)

[Wkład w projekt 6](#_Toc452860446)

[Wkład Konrada 6](#_Toc452860447)

[Sprint 1 6](#_Toc452860448)

[Sprint 2 6](#_Toc452860449)

[Sprint 3 6](#_Toc452860450)

[Sprint 4 6](#_Toc452860451)

[Sprint 5 6](#_Toc452860452)

[Wykresy 7](#_Toc452860453)

[Wkład Michała 8](#_Toc452860454)

[Sprint 1 8](#_Toc452860455)

[Sprint 2 8](#_Toc452860456)

[Sprint 3 8](#_Toc452860457)

[Sprint 4 8](#_Toc452860458)

[Sprint 5 8](#_Toc452860459)

[Wykresy 8](#_Toc452860460)

[Wkład Kamila 8](#_Toc452860461)

[Sprint 1 8](#_Toc452860462)

[Sprint 2 8](#_Toc452860463)

[Sprint 3 8](#_Toc452860464)

[Sprint 4 8](#_Toc452860465)

[Sprint 5 8](#_Toc452860466)

[Wykresy 8](#_Toc452860467)

# Cel projektu

Projekt miał na celu dostarczenie gotowego produktu jakim jest aplikacja wyszukująca podobieństwa w danej bazie danych.

Studenci zostali zobowiązani do wykonania tego zadania korzystając z opracowanej przez firmę NVIDIA technologii CUDA.

Istotnym elementem projektu była praca zespołowa członków zespołu.

Organizacja pracy miała przebiegać w oparciu o metodologię SCRUM oraz KANBAN.

Narzędzia pracy jakie wykorzystał zespół w celu osiągnięcia celu projektu to serwis internetowy GitHub wykorzystujący system kontroli wersji Git oraz Trello – serwis umożliwiający pracę w systemie KanBan.

# Wstęp teoretyczny

Co to jest scrum?

Co to jest KanBan?

Coś o metrykach?

# Scrum

## Sprint 1

### Cel

Celem pierwszego sprintu było opracowanie sposobu pracy nad projektem. Zespół w trakcie trwania sprintu starał się wyszukiwać najlepsze rozwiązania komunikacji między członkami oraz starał się znaleźć najlepszy podział pracy. Najważniejszym elementem pierwszego sprintu było stworzenie środowiska pracy oraz opracowanie najlepszej dla nas tablicy KanBanowej.

### Milestones

Kamieniem milowy tego sprintu było znalezienie pierwszych pięciu metryk i osadzenie ich w środowisku pracy

### Tickety i subtickety

***Scrum Master*** – Scrum Masterem tego sprintu był Konrad.

***Daily Scrum*** .

***Zainstalować Google*** .

***Stworzyć repozytorium GitHub*** – osoba odpowiedzialna - Konrad

***Zapoznać się z poleceniami prowadzącego dotyczącymi pierwszego sprintu.***

***Zainstalować Plus for Trello***

***Znaleźć 5 metryk*** – osoba odpowiedzialna – Konrad

### Problemy

Kamil napotkał problem z instalacją Google Chrome. Problem został szybko rozwiązany.

### Podsumowanie

W czasie sprintu pierwszego zespół wypróbował metodykę „Daily Scrum”. Jednakże, codzienne poświęcanie czasu na projekt nie było możliwe, dlatego przestaliśmy praktykować „Daily Scrum”. Metodę tą zastąpiliśmy w następnych sprintach użyciem dodatkowej listy „Problemy” gdzie każdy wypisywał problemy które napotkał.

Sukcesem zakończyła się próba wybrania na sprint pierwszy „Scrum Mastera”, który na celu ma organizację tablicy Trello, rozwiązywanie sporów i innych problemów.

W trakcie trwania pierwszego tygodnia pracy założono repozytorium na GitHub’ie, gdzie znajdują się pliki na których zespół pracował. Dzięki możliwością jakie oferuje serwis GitHub była możliwa szybka wymiana danych i wykonanej pracy.

Dodatkowo każdy z zespołu zainstalował na swoim komputerze przeglądarki Google Chrome, które umożliwiły instalacje dodatku do Trello „Plus for Trello”, który znacznie ułatwił komunikację pomiędzy członkami zespołu i organizację pracy.

Ponieważ członkowie zespołu po raz pierwszy spotkali się z takim trybem pracy, sporo czasu zajęło zapoznanie się z celem projektu i zrozumienie go. To była główna bariera hamująca postępy pierwszego sprintu.

Znaleziono 5 metryk potrzebnych do napisania pierwszych prostych programów.

Celem sprintu pierwszego było zapoznanie się z środowiskiem pracy i uruchomienie odpowiednich środowisk programistycznych na własnych komputerach. Cel sprintu został zrealizowany bez większych trudności. Zadania mniej ważne również.

## Sprint 2

### Cel

Celem sprintu było wykonanie pierwszego właściwego fragmentu wykonywalnego programu. Zespół miał skompilować pierwsze obliczenia metryk

### Milestones

Skonfigurować do współpracy CUDA Toolkit, VisualStudio i emulator

### Tickety i subtickety

**Scrum Master** – Michał Pilch

**Zapoznanie się z technologią CUDA**

**Błąd krytyczny systemu** – Konrad Janczarski

1. **Zainstalować Chrome, Plus for Trello, Tortoise Git** – Konrad – Ticket związany z błędem krytycznym systemu

**Zainstalować NVIDIA CUDA Emulator** – Konrad Janczarski i Michał Pilch

**Zainstalować i skonfigurować do współpracy VisualStudio i CUDA** – Kamil Switek

**Skonfigurować do współpracy CUDA Toolkit, VisualStudio i emulator** – Konrad Janczarski i Michał Pilch

**Skompilować pierwsze obliczenia metryk**- Kamil Switek

**Znaleźć 5 metryk –** Konrad Janczarski

### Problemy

Napotkano istotny problem. Nie udało się skonfigurować emulatora. Uniemożliwia to pracę w architekturze cudy Konradowi i Michałowi

### Podsumowanie

Sprint 2 był już przejściem do poważniejszej pracy. Najpierw zapoznaliśmy się z narzędziami pracy grupowej i dostosowaliśmy je do siebie, zainstalowaliśmy również dodatek PlusforTrello dla przeglądarki Google Chrome. Udało się w pełni skonfigurować środowisko pracy (Nvidia CUDA, Visual Studio 2012). Kompilacja przykładowych programów podanych przez producenta zakończyła się powodzeniem. Następnie zaimplementowaliśmy pierwsze obliczenia metryk na karcie graficznej (tych metryk, które zostały znalezione w poprzednim sprincie). Podczas gdy jedna część zespołu zapoznawała się dalej z technologią CUDA, kompilacją programów i obliczeniami na karcie graficznej, pozostała część zajęła się znajdowaniem kolejnych metryk. Niepowodzeniem zakończyło się skonfigurowanie technologii CUDA na kartach graficznych innych niż Nvidia, ponieważ emulator odmawiał współpracy.

## Sprint 3

### Cel

Celem sprintu jest rozwinięcie fragmentu docelowego programu który do tej pory stworzyliśmy.

### Milestones

-Znaleźć metryki „Razem mamy 57 metryk”

### Tickety i subtickety

-Scrum Master – Kamil Switek

-„Razem mamy 57 metryk”

- Przerobić program z poprzedniego sprintu tak by działał na wielu wątkach

- Napisać program obliczający metryki bez technologii CUDA

### Problemy

-Nadal nie skonfigurowaliśmy emulatora CUDA

### Podsumowanie

Podczas sprintu 3 naszym głównym zadaniem było kontynuowanie zadań zaplanowanych w poprzednim sprincie. Zakończyliśmy znajdowanie metryk, udało się ostatecznie znaleźć ich ponad 50. Kontynuowaliśmy proces implementacji obliczeń metryk, tym razem jednak przydzielając każdej metryce osobny wątek na karcie graficznej. Niestety problem konfiguracji środowiska CUDA na kartach Radeon nie został rozwiązany. Rozpoczęliśmy implementację programu obliczającego to samo, ale na procesorze, aby sprawdzić, czy udało się osiągnąć jakiekolwiek przyspieszenie.

## Sprint 4

### Cel

Celem sprintu było napisanie programu który będzie liczyć metryki nie używając technologii CUDA oraz rozwijanie programu docelowego.

### Milestones

Napisać program liczący metryki nie używając CUDY

### Tickety i subtickety

**Scrum Master** – Konrad Janczarski

**Uporządkować plik „Metryki”** – Konrad Janczarski

**Znaleźć źródła metryk** – Konrad Janczarski

**Napisać dokumentację sprintu 1** – Konrad Janczarski

**Dodać do programu kolejne metryki (36)** – Kamil Switek

**Napisać program liczący metryki nie używając CUDY (11 metryk)** – Michał Pilch

### Problemy

Nie znaleziono źródła „metryki w X^N” która została już skompilowana.

Nie udało się również naprawić problemów związanych z błędnymi wprowadzeniami czasów do Plus for Trello (Konrad Janczarski).

### Podsumowanie

Sprint 4 był rozwinięciem poprzednich sprintów. Udało nam się przerobić program na zwykły język C, bez użycia CUDY, żeby móc później sprawdzić czy CUDA przyspiesza nam czas obliczeń. Do programu w CUDZie udało się dołożyć kolejne metryki, program działa obecnie na 36 metrykach. Zrobiliśmy również dokumentację poprzednich sprintów. Udało się także znaleźć źródła uprzednio znalezionych metryk. Nie udało się nam niestety znaleźć źródła metryki X^N. Kolejny problem który wynikł w tym tygodniu to naprawienie błędów przepracowanego czasu w Trello, niestety nie udało się go rozwiązać.

## Sprint 5

### Cel

Celem sprintu piątego była ogólnie pojęta finalizacja projektu. Składało się na to doprowadzenie aplikacji do ostatecznego kształtu oraz sporządzenie dokumentacji projektu oraz opracowanie prezentacji.

### Milestones

Napisać program który porówna szybkość wykonywania analizy danych za pomocą technologii CUDA i bez użycia tej technologii.

Uruchomić działający program analizujący dane w technologii CUDA.

Napisanie dokumentacji.

Napisanie newsletter’a

### Tickety i subtickety

***Dokumentacja Sprintu 2 – Kamil Świtek***

***Dokumentacja Sprintu 3 – Kamil świtek***

***Dokumentacja Sprintu 4 – Michał Pilch***

***Dokumentacja Sprintu 5 – Konrad Janczarski***

***Uzupełnienie kodu metryk nie wykorzystującego CUDy***

***Napisanie programu porównującego czas wykonywania analizy***

***Stworzenie bazy danych do analizy***

***Uruchomienie finalnego produktu***

***Udział w sprintach***

***Udział w sprintach***

***Udział w sprintach***

***Newsletter***

***Przygotować prezentację***

### Problemy

W sprincie tym nie zgłaszano żadnych problemów

### Podsumowanie

Sprint piąty to ostatni sprint w naszym Scrumie. W sprincie tym dokonaliśmy ostatnich prac finalizujących produkt.

Sprint piąty był dwa razy dłuższy od pozostałych sprintów. Mimo tego, że mieliśmy na wykonanie tego sprintu więcej czasu niż na inne to nakład pracy na jednostkę czasu w tym sprincie był największy spośród całego Scruma. Było to spowodowane złym oszacowaniem pozostałego czasu pracy.

Podczas tego sprintu sporządzono podsumowanie pracy jaką wykonał zespół w Sprintach 2,3,4 i 5.

Zdecydowanie były to najłatwiejsze zadania tego sprintu. Dzięki prowadzonym tablicom KanBanowym z łatwością można było przypomnieć sobie cele stawiane w tamtych okresach, wykonane zadania oraz występujące problemy.

W sprincie tym został uzupełniony kod programu wykonującego analizę bazy danych nie używając technologii CUDy. Należało przepisać napisane już kody w architekturze CUDY dokonując koniecznych zmian.

Istotnym elementem było stworzenie pliku tekstowego zawierającego przykładową bazę danych którą będzie analizował nasz program.

Najważniejszym i najtrudniejszym elementem Backlogu Produktu w tym sprincie było uruchomienie programu w finalnej formie. Należało napisać fragmenty kodu wykorzystujące metryki do analizy pliku.

Następnie każdy z członków zespołu dokonał sprawozdania swojego udziału w dokonywaniu przyrostów każdego sprintu.

Newsletter

Prezentacja

# Wkład w projekt

## Wkład Konrada

### Sprint 1

W sprincie pierwszym pełniłem funkcję Scrum Mastera. Zadanie to zabrało mi najwięcej czasu spośród powierzonych mi na ten sprint zadań. Jako Scrum Master zajmowałem się organizacją tablicy Kanbanowej. Dokonywałem przypisywania osób do danych ticketów, nadawałem ticketom odpowiednie etykiety, zarządzałem Daily Scrumami.

W sprincie tym powierzono mi również zadanie założenia repozytorium GitHub. Zainstalowałem przeglądarkę Google Chrome i dodatek Plus for Trello.

### Sprint 2

W sprincie tym napotkałem problem w postaci błędu krytycznego systemu. Musiałem przeinstalować system na komputerze na którym pracuję. Oczywiście po reinstalacji systemu musiałem znowu zainstalować programy które instalowałem w sprincie pierwszym – Google Chrome, Plus for Trello, TortoiseGit.

Gdy uporałem się już z systemem operacyjnym napotkałem kolejny, bardzo istotny problem – moja karta graficzna nie obsługuje technologii CUDy.

Problem ten miał rozwiązać NVIDIA CUDA Emulator który zainstalowałem na swoim komputerze.

Następnie zainstalowałem Visual Studio oraz CUDA Toolkit. Bardzo dużo czasu zajęła mi próba skonfigurowania do współpracy CUDA Toolkit, Visual Studio i emulator. Niestety próba zakończyła się niepowodzeniem.

W Sprincie tym znalazłem 5 metryk.

### Sprint 3

Sprint trzeci rozpocząłem ponowną próbą skonfigurowania emulatora na swoim komputerze. Ponownie zajęło mi to dużo czasu i zakończyło się niepowodzeniem. Doszedłem do wniosku, że nie warto poświęcać w tym celu więcej czasu. Zająłem się szukaniem metryk.

W sprincie tym udało mi się znaleźć 47 metryk.

### Sprint 4

Sprint ten rozpocząłem od porządkowania pliku z metrykami. Należało sprawdzić czy któreś metryki się nie powtarzają oraz znaleźć źródła metryk. Ponieważ na bieżąco nie były notowane źródła, poszukiwania zajęły sporo czasu.

Nie udało mi się znaleźć źródła jednej z metryk która już została zaimplementowana.

Napisałem dokumentację sprintu 1.

Niestety w trakcie poprzednich sprintów wprowadziłem nieprawidłowe szacunki czasów oraz nieprawidłowe czasy spędzone nad konkretnymi zadaniami do Plus for Trello. W trakcie tego sprintu próbowałem naprawić błędne wprowadzenia czasów jednakże komendy które miały naprawić złe wpisy nie działały.

### Sprint 5

W sprincie piątym zajmowałem się sporządzaniem dokumentacji projektu.

Sporo czasu zajęło mi sporządzenie sprawozdania mojego wkładu w projekt .

Napisałem również dokumentacje sprintu 5.

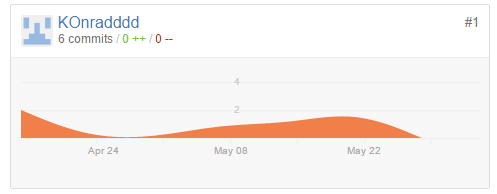
### Wykresy

#### Trello

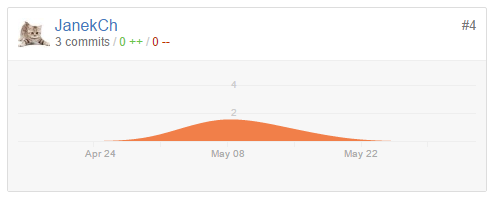
Niestety, najprawdopodobniej podczas próby naprawy błędnych wprowadzeń szacunków czasu, usunąłem wszystkie szacunki. Plus for Trello nie wygenerowało żadnego wykresu.

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

#### GitHub



Wykres commitów w czasie trwania scrum’a.



Commitów dokonywałem również z pracowni komputerowej, gdzie na stanowisku pracy TortoiseGit zalogowany był na konto JanekCh. Mimo że commitów dokonałem ja (logując się uprzednio) zostały zarejestrowane jako commity Janka.

## Wkład Michała

### Sprint 1

### Sprint 2

### Sprint 3

### Sprint 4

### Sprint 5

### Wykresy

#### Trello

#### GitHub

## Wkład Kamila

### Sprint 1

### Sprint 2

### Sprint 3

### Sprint 4

### Sprint 5

### Wykresy

#### Trello

#### GitHub

# Podsumowanie

Zespołowi udało się osiągnąć cel projektu.

Zespół stworzył program który analizuje bazę danych zawartą w pliku tekstowym. Do tego celu została użyta równoległa architektura obliczeniowa CUDA.

W trakcie trwania projektu stworzono również aplikację która porównuje szybkość wykonywania analiz szeregowo i równolegle – za pomocą cudy. Okazało się, że program wykonywany w cudzie działa dużo szybciej. Cuda zapewnia radykalny wzrost wydajności.

W celu wykonania przykładowej analizy danych wygenerowano plik wypełniony tablicami losowych liczb.

Program działa w te sposób, że dokonuje losowych wyborów wierszy zbudowanej macierzy. Następnie dla danych wierszy losowana jest liczba analizowanych współrzędnych i dane współrzędne.

Wybrane współrzędne są poddawane obliczeniom ich metryk, dzięki czemu można określić stopień ich korelacji.

Metodologia organizacji pracy „Scrum” zasadniczo usprawniła wykonanie projektu. Dzięki jednoznaczności obieranych celów udawało nam się osiągać systematyczne i efektywne przyrosty pracy, jednocześnie nie tracąc czasu na brak organizacji.

Scrum bardzo dobrze sprawdza się w dużym natłoku pracy. O ile w poprzedzających sprint piąty sprintach nie były zauważalne zalety Scruma, to w sprincie piątym Scrum sprawdził się wyjątkowo dobrze. Umożliwił skuteczną organizację pracy. Scrum sprawdza się bardzo dobrze w obszernych projektach, które należy wykonać w krótkim czasie. Znacząco zwiększa wydajność pracy.

W poszczególnych sprintach zespół posługiwał się metodą tablicy Kanbanowej. W celu organizacji sprintu tą metodą użyliśmy serwisu Trello.

Adaptacja do systemu Kanbanowego zajęła zespołowi jeden tydzień. W tym czasie zespół praktykował Daily Scrumy które okazały się nieskuteczne. Daily Scrum to bardzo dobra metoda podnoszenia wydajności pracy wykonywanej ciągle. Nasz zespół musiał zajmować się także innymi obowiązkami związanymi ze studiami, dlatego praca nad projektem nie była systematyczna i Daily Scrum nie miał dużego sensu stosowania.

Pracując nad programem korzystaliśmy z narzędzia GitHub wykorzystującego system kontroli wersji Git. W trakcie trwania całego Scruma nie skorzystaliśmy z możliwości systemu kontroli wersji. Pracą dzieliliśmy się w taki sposób, że nie było konieczne pracowanie na tych samych plikach. Nie wystąpiła żadna kolizja wprowadzenia danych.

Bardzo przydatnym okazał się dostęp do repozytorium projektu. Brak konieczności wymieniania się plikami z wynikami pracy poprzez rozsyłanie każdorazowo do członków projektu plików znacząco podnosi komfort pracy i pozwala uniknąć wielu nieporozumień.

Projekt umożliwił członkom zespołu nabyć wiele umiejętności związanej z pracą zespołową. Bardzo ważne jest zdobyte doświadczenie pracy nad wspólnym celem z innymi osobami.

Dzięki projektowi studenci poznali wiele metod i narzędzi których wcześniej nie znali, a które znacznie poprawiają organizację i wydajność pracy zespołowej.

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

# Źródła

## Metryki

### Wikipedia.pl

1. **Artykuł:” Przestrzeń metryczna”** (https://pl.wikipedia.org/wiki/Przestrze%C5%84\_metryczna), 22.05.201616r.)
2. **Artykuł „Odległość Hamminga”** (https://pl.wikipedia.org/wiki/Odleg%C5%82o%C5%9B%C4%87\_Hamminga, 22.05.2016r.:)

### MathDotNet

1. **Artykuł: Distance Metrics** (http://numerics.mathdotnet.com/Distance.html, 22.05.201616r.)

### MATHWORLD– WOLFRAM

1. **Artykuł: French Metro Metric** (http://mathworld.wolfram.com/FrenchMetroMetric.html), 22.05.2016r.:

Sung-Hyuk Cha**,**

1. **” *Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions”*** Issue 4, Volume 1, 2007: