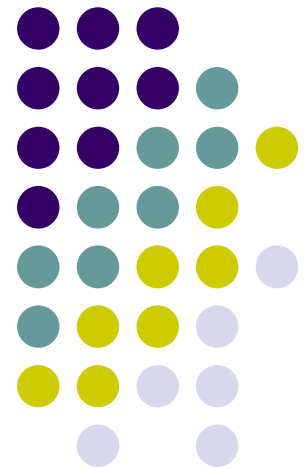
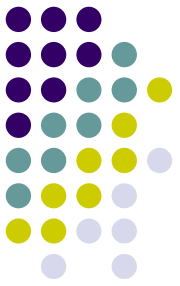


Informatyka w „Pi of the Sky”

Marek Biskup
Uniwersytet Warszawski
mbiskup@mimuw.edu.pl



Informatyka w „Pi of the Sky”



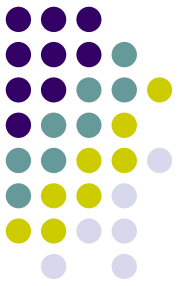
- Przetwarzanie obrazów
 - Obróbka zdjęć
 - Mierzenie jasności i położenia gwiazd
- Sztuczna inteligencja / data mining
 - Wykrywanie błysków i pojaśnień
 - Analiza gwiazd zmiennych
 - Śledzenie satelitów
- Bazy danych / WWW
 - Przechowywanie i udostępnianie informacji

Przetwarzanie zdjęć online



- Surowe zdjęcia
 - 2000x2000 pikseli, 16-bitowa głębokość szarości
- Poprawianie
 - Korekcja czułości pikseli i zmniejszenie szumów
- Szukania błysków
 - Wyostrzanie
 - Szukanie maksimum jasności
 - Porównanie z poprzednią klatką

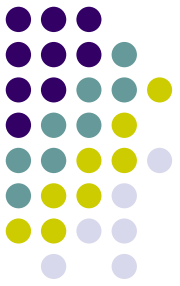
Przetwarzanie zdjęć offline



- Astrometria – pomiar położenia gwiazd
- Fotometria – pomiar jasności gwiazd
- Kalibracja astrometrii i fotometrii
 - Porównanie z katalogiem gwiazd
- Zapisywanie do bazy danych
 - Współrzędne gwiazdy
 - Jasność gwiazdy

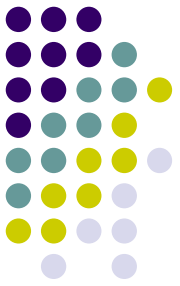


Astrometria i kalibracja

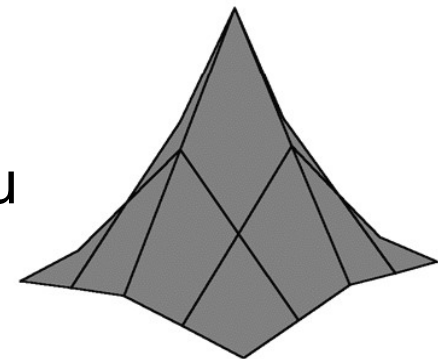
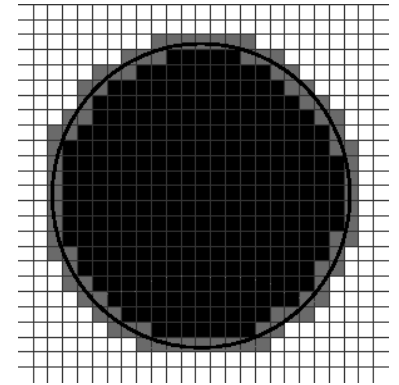


- Nieliniowe odwzorowanie współrzędnych CCD na niebieskie
 - Fluktuacje gęstości atmosfery
 - Optyka
 - Współrzędne sferyczne
- Kalibracja z użyciem katalogu
 - Dopasowywanie powierzchni do pomiarów dla znanych gwiazd

Fotometria



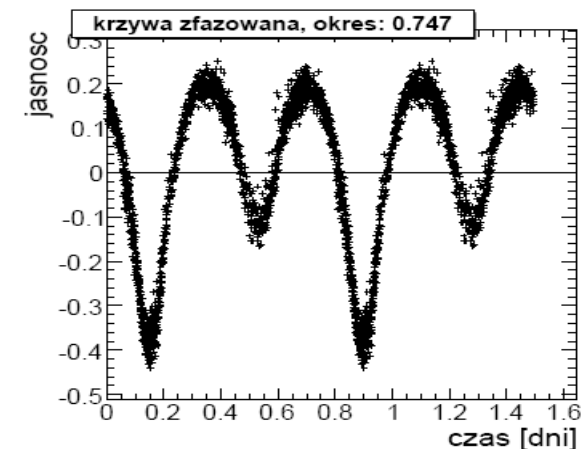
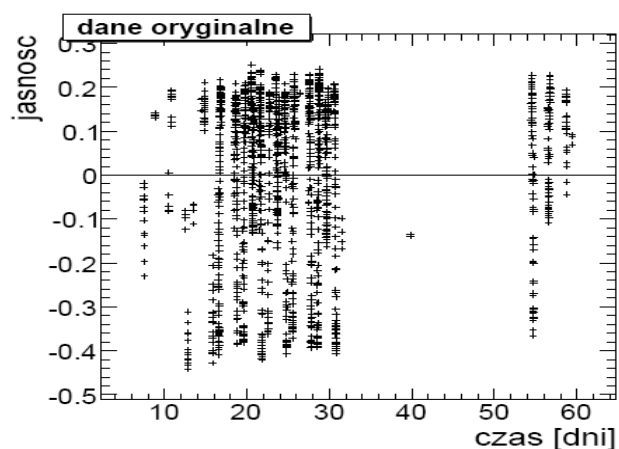
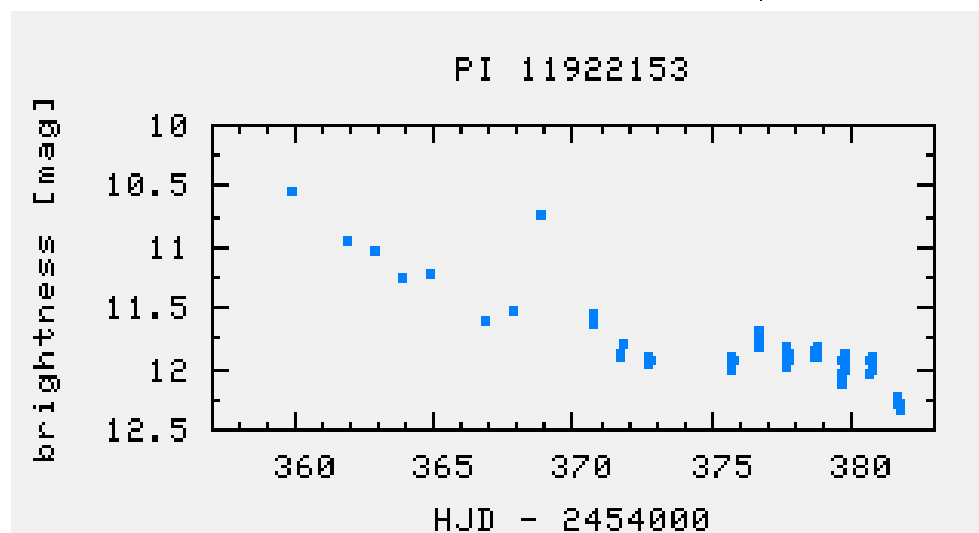
- Aperturowa
 - Sumowanie jasności kilku pikseli
- Profilowa
 - Dopasowywanie profilu (np. Gaussa) do jasności pikseli
- Problemy:
 - Profil gwiazdy zależy od położenia na zdjęciu
 - Fotometria wykazuje duże fluktuacje
 - Kamery czasem pracują z otwartą migawką



Data Mining



- Wyszukiwanie błysków
 - Algorytm online (na obrazkach)
 - Algorytmy offline (na bazie danych)
- Wyszukiwanie gwiazd zmiennych



Bazy danych: zawartość



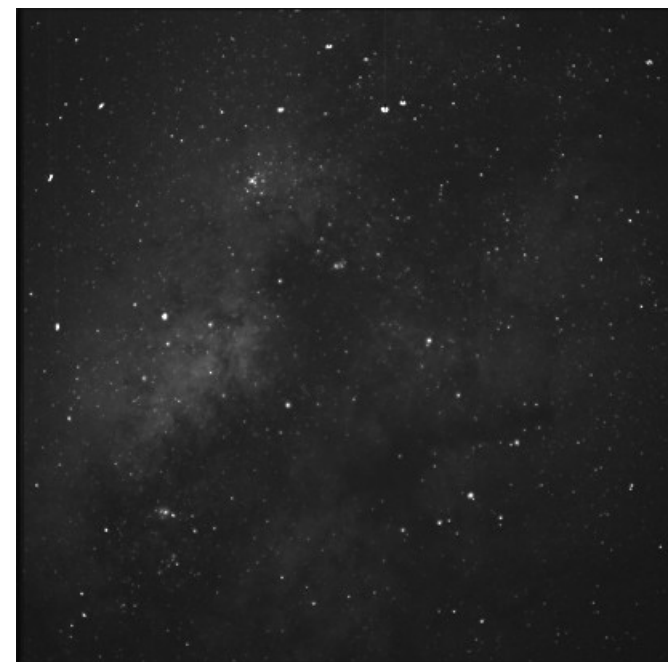
- **Pomiary jasności gwiazd**
 - Czas wykonania pomiaru
 - Jasność
 - Dodatkowe parametry
 - Numer kamery
- **Uśrednione informacje o gwiazdach**
 - Średnia jasność
 - Średnie położenie
 - Nazwa (astronomiczna)



Rozmiar bazy danych



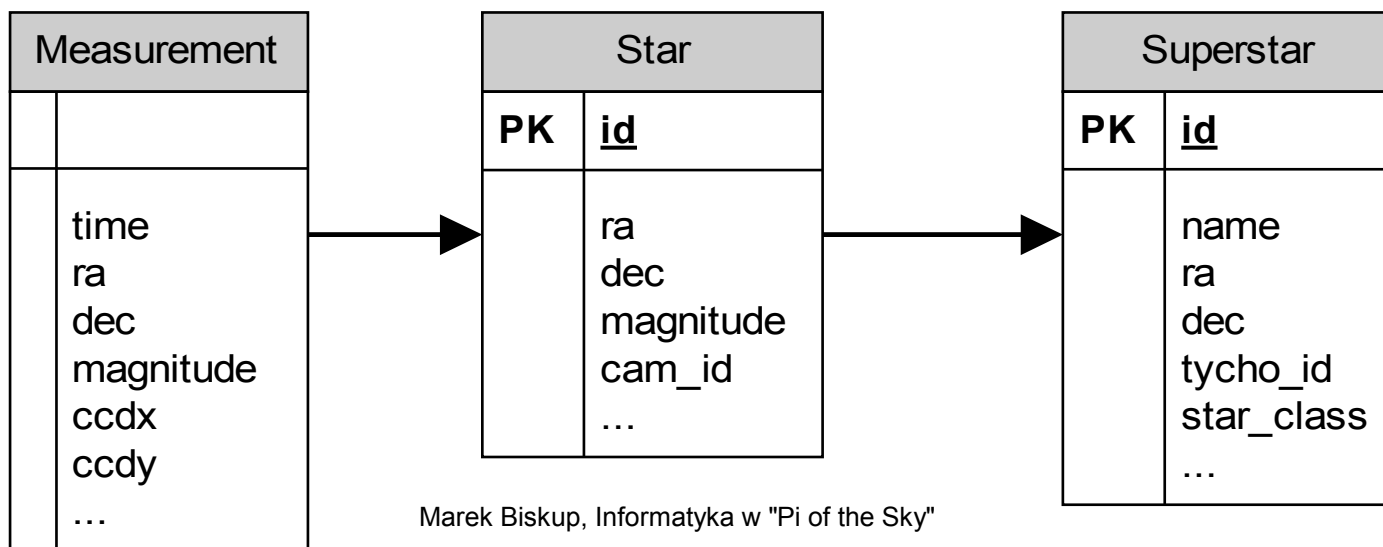
- Około 6 mln pomiarów na dobę
 - Zdjęcia co 12 s., sumowane po 20
 - ~20000 gwiazd / zdjęcie
 - ~10 godzin obserwacji / noc
 - 2 kamery
- Ponad 1 mld pomiarów na rok
 - ~200GB
- Docelowo 32 kamery
 - 100 mln pomiarów na dobę
 - 20 mld pomiarów na rok
 - Nawet 20x więcej jeśli obrabiane będą pojedyncze klatki



Struktura bazy danych



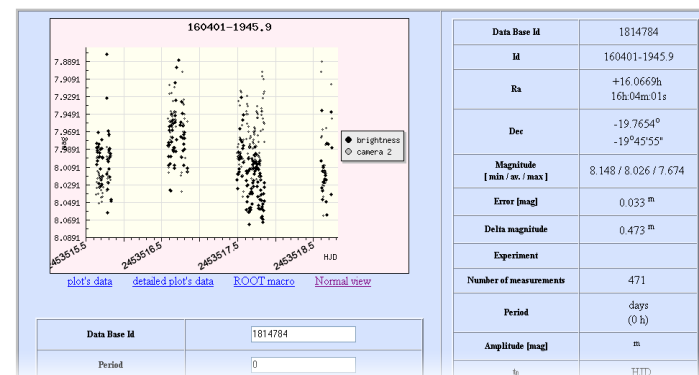
- Podstawowe tabele
 - **Measurements** – pomiary jasności
 - **Stars** – uśrednione pomiary z jednej kamery dla każdej gwiazdy
 - **Superstars** – uśrednione pomiary ze wszystkich kamer dla każdej gwiazdy



Implementacja bazy danych



- Engine: PostgreSQL
- Kilka baz danych gwiazd
 - Oddzielne bazy danych dla różnych okresów
 - Bazy danych z informacjami z katalogów gwiazd
 - Robocze bazy danych dla aktualnej fazy eksperymentu (fragmenty danych)
- Skrypty do ładowania z plików z danymi
- Interfejs WWW (php)

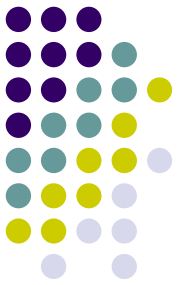


Problemy do rozwiązania



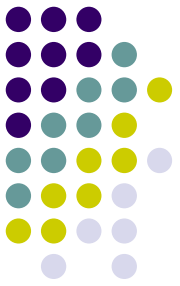
- Wiele baz danych
 - bałagan (trudności w utrzymaniu)
 - brak możliwości łączenia danych do zapytań SQL
- PostgreSQL – brak:
 - narzędzi dostępnych dla komercyjnych baz danych
 - ręcznej optymalizacji zapytań
 - rozproszenia bazy danych
- Niedoskonane „środowisko pracy naukowca”
- Chwiejna fotometria

Tematy z baz danych



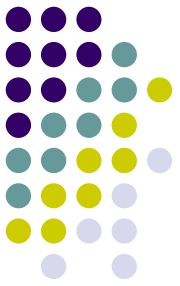
- **Migracja bazy** do DB2 + integracja baz danych (L, M)
- Opracowanie strategii backupów, fail-over, partycjonowania, utrzymania środowiska roboczego
- Stworzenie bazy buforowej dla nowych danych
- Migracja oprogramowania do DB2 (L)
 - Biblioteki C++, interfejsy PHP, procedury PL/SQL
- Analiza kosztów zapytań, klastrowanie, materialized views, **optymalizacja** bazy i zapytań SQL (M)

BD – tematy licencjackie



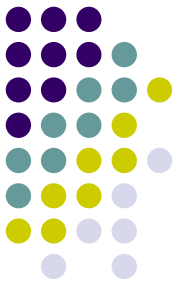
- Stworzenie biblioteki C++ będącej interfejsem do bazy DB2 (L)
- Optymalizacja i porównanie kosztów zapytań SQL w DB2 i PostgreSQL (L)
- Analiza wydajności partycjonowania danych w DB2 (L)
- Stworzenie GUI do analizy gwiazd zmiennych (L)
- Migracja oprogramowania PHP z PgSQL do DB2 (L)

Tematy z WWW



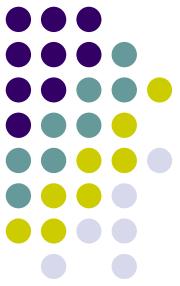
- **System kontroli eksperymentu** (L, M)
 - Parametry pracy urządzeń zapisywane do bazy danych
 - Interfejs do kontroli
 - Analiza danych – możliwość pisania pluginów
 - Technologie: np. Ruby on Rails, GWT, Ajax, Django
- Interfejs **przeglądania gwiazd** w stylu „Google Maps” (płynne przewijanie, zoom, zdjęcia, etc.) (M)
- Przeglądarka WWW do **zdjęć** astronomicznych w formacie FIT (65536 odcieni szarości) (L)

Tematy z przetwarzania obrazów



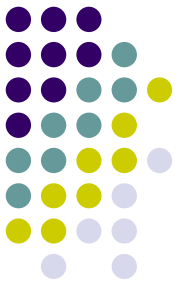
- Analiza i poprawa **fotometrii** (M)
 - Profil gwiazdy zależny od położenia na chipie
- Stworzenie programu do **astrometrii** (M)
 - Perspektywy: szukanie śmieci kosmicznych
- Stworzenie **mapy nieba** ze zdjęć (M)
 - Uśrednienie i sklejanie zdjęć, usunięcie obiektów ruchomych, korekcja jasności
- Kompresja obrazów astronomicznych (L, M)

Tematy ze sztucznej inteligencji

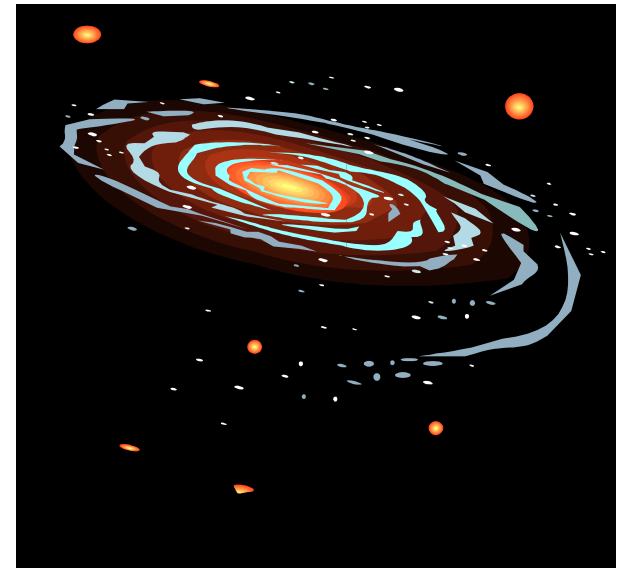


- Automatyczna **klasyfikacja gwiazd** zmiennych (M)
 - Sieci neuronowe / analiza fourierowska
- Odróżnianie **błędnych pomiarów** od poprawnych (M)
- Znajdowanie **śmieci** kosmicznych (M)

Możliwości



- Uczestnictwo w **eksperymentach naukowym**
- Praca z „produkcyjną” bazą danych średniego i dużego (w przyszłości) rozmiaru
- Poznawanie **nowych technologii**
 - w szczególności DB2 i jej narzędzi
- Starcie z typowymi problemami **inżynierii oprogramowania**
- Dużo danych do analizy

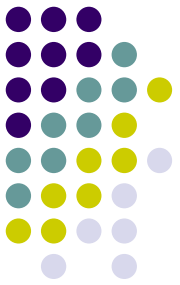


Podsumowanie



- „Pi” jest otwarte na współpracę ze studentami
- W „Pi of the Sky” pracuje się z rzeczywistymi i sporymi bazami danych
 - „Pi” planuje migrację do DB2 i integrację baz danych
 - Przy tym potrzebna jest pomoc
- Pi dysponuje sporą ilością danych do których automatyczna analiza jest konieczna
 - Zastosowania znajduje sztuczna inteligencja i przetwarzanie obrazów

Dodatkowe informacje



- Strona projektu „Pi of the Sky”:
 - <http://grb.fuw.edu.pl>
- Kontakt: mbiskup@mimuw.edu.pl