Warszawa, 17.11.2012

**Projekt WEDT**

Kursy akcji na giełdach a informacje.

Dokumentacja wstępna

Prowadzący:

dr Piotr Andruszkiewicz

Adrian Wiśniewski,

*nr indeksu 206298*

Jan Wilczak,

*nr indeksu 214698*

Spis treści

[Temat projektu 3](#_Toc342065721)

[Propozycja implementacji 3](#_Toc342065722)

[Źródła danych 3](#_Toc342065723)

[Komunikaty 3](#_Toc342065724)

[Wiadomości 3](#_Toc342065725)

[Rekomendacje 3](#_Toc342065726)

[Narzędzia 4](#_Toc342065727)

# Temat projektu

Próba wykorzystania informacji pojawiających się na stronach WWW do prognozowania wyników akcji/indeksów giełdowych bądź innych wskaźników finansowych.

# Propozycja implementacji

Zbadamy, czy zachowanie spółki giełdowej można opisać przy użyciu ukrytych modeli Markowa. Zakładamy, że spółka jest pewną skończoną maszyną stanów, której struktury nie znamy – możemy natomiast się jej domyślać na podstawie obserwacji pewnych sygnałów. Sygnałami tymi mogą być:

* Dzienne wahania kursu
  + Trzy wartości: wzrost, brak zmian (lub zmiana nieznaczna) , spadek
* Czasowniki występujące w wiadomościach dotyczących spółki pojawiających się na portalach internetowych
  + Zastosowanie NLP
* Indeksu grupy wiadomości
  + Po uprzednim pogrupowaniu wiadomości z użyciem metod text-miningu.

Każdy ze stanów maszyny ma określone prawdopodobieństwa wzrostu, utrzymania i spadku kursu, co pozwala w praktyczny sposób wykorzystać model. Ponadto analiza ilości stanów i połączeń między nimi może dodatkowo umożliwić planowanie w długim horyzoncie czasu. Rozważamy możliwość budowania modelu dla jednej konkretnej spółki lub wielu podobnych spółek.

# Źródła danych

## Komunikaty

http://www.money.pl/gielda/spolki\_gpw/grupa;lotos;sa,lts,emitent,1.html

## Wiadomości

http://www.money.pl/gielda/spolki\_gpw/grupa;lotos;sa,lts,artykuly,1.html

## Rekomendacje

http://www.money.pl/gielda/rekomendacje/rekomendacje,457,-1,1.html

# Narzędzia

* Java - jako język implementacji
* Jahmm - biblioteka do ukrytych modeli Markova
  + <https://code.google.com/p/jahmm/>
* Weka - pakiet dataminingowy użyty do grupowania
  + <http://www.cs.waikato.ac.nz/ml/weka/>
* Morfologik -pakiet NLP
  + <http://morfologik.blogspot.com/>
* JSoup - pakiet do manipulacji dokumentami html
  + <http://jsoup.org/>

Update na brudno:

Workflow:

# Konfiguracja.

Wykorzystanym szkieletem tworzenia aplikacji jest Spring Framework. Dzięki implementacji zasady odwrócenia sterowania („Inversion-of-Control”) pozwala na szybkie i prostsze składanie aplikacji poprzez konfigurację w pliku *app.config.xml*.

<bean class="stockfortuneteller.app.Main">

<property name="tasks">

<list>

<ref bean="stockStemmerCSVBuilder"/>

<ref bean="stockWekaARFFBuilder"/>

<ref bean="stockWekaBuilder"/>

</list>

</property>

</bean>

Właściwość „tasks” pozwalaja na wybór tych elementów, które zostaną uruchomione (a dokładnie ich metody „void execute()”). W ten sposób możemy raz pobrać dane, wykonać stemming i zapis do pliku CSV (funkcjonalność” stockStemmerCSVBuilder”), stworzyć model ARFF („stockWekaARFFBuilder”), a następnie testować klasteryzację używając ostatniego składnika („stockWekaBuilder”).

Ponadto w pliku konfiguracyjnym znajdują się właściwości każdego z w/w elementów.