

POLITECHNIKA ŚLĄSKA
WYDZIAŁ INŻYNIERII MATERIAŁOWEJ I METALURGII

Kierunek: Informatyka Przemysłowa

Rodzaj studiów: stacjonarne

PROJEKT INŻYNIERSKI
Michał Cwiękała

INTEGRACJA JAVA Z MS EXCEL

Kierujący projektem:

Dr inż. Joanna Lisok

Recenzent:

Dr inż. Marcin Blachnik

Katowice, styczeń, 2016 r.

Spis treści

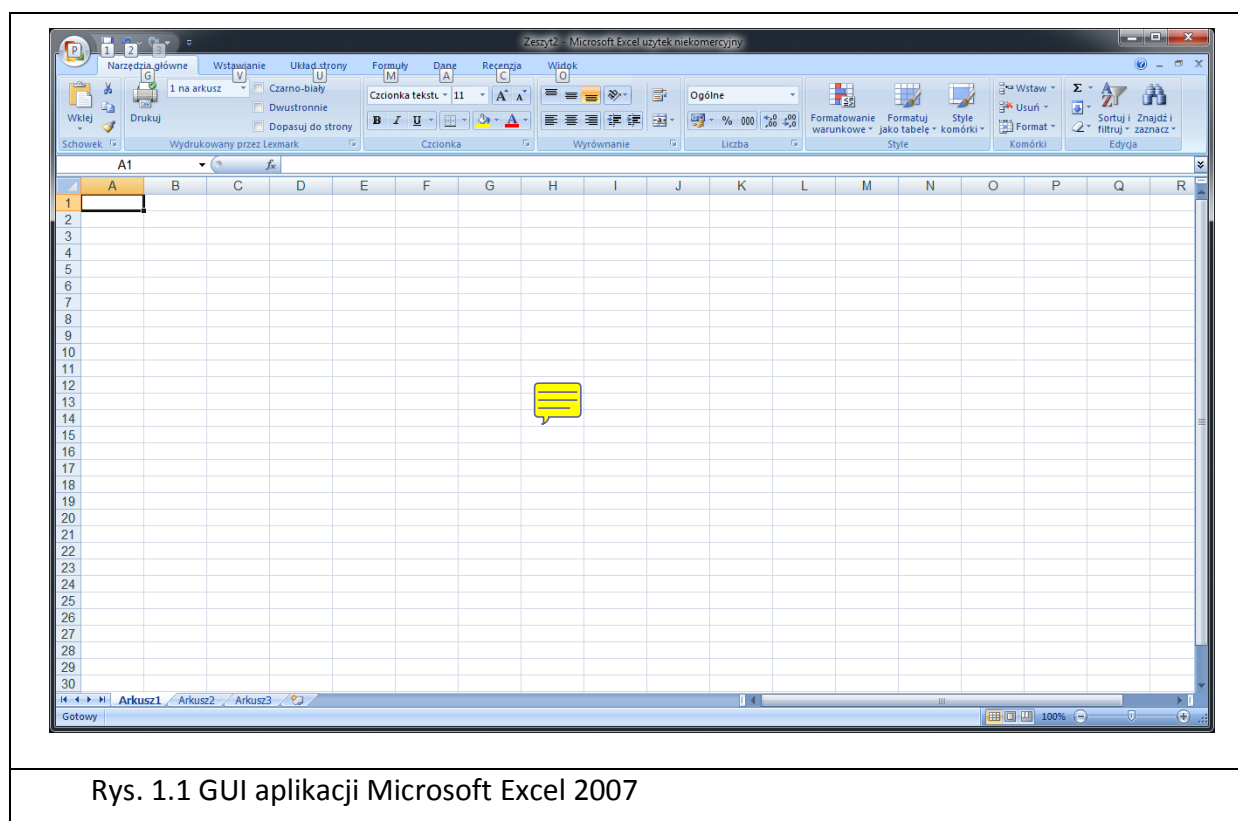
1.	Wstęp	3
2.	Cel i zakres pracy	5
3.	Możliwe narzędzia integracji MS Excel i Java	5
3.1.	Możliwe rozwiązania:	6
3.1.1.	VBA	6
3.1.2.	.Net.....	6
3.1.3.	Java	6
3.2.	Wybór technologii.	7
3.3.	Biblioteka XLLoop[4].....	7
3.3.1.	Zastosowanie	7
3.3.2.	XLLoop składa się z dwóch głównych komponentów:	8
4.	RapidMiner	8
5.	Szczegóły implementacyjne projektu	10
6.	Przykład użycia	10
7.	Wnioski i podumowanie	10
8.	Bibliografia	10

1. Wstęp

W dzisiejszym skomputeryzowanym świecie biznesowym praca z arkuszem kalkulacyjnym to podstawa w każdej firmie i instytucji, niezależnie od branży. Microsoft Excel to najpopularniejszy na rynku program służący do dokonywania obliczeń zestawionych w formie tabelarycznej. [1] Dostarcza on funkcje matematyczne, finansowe, bazodanowe oraz zapewnia półautomatyczne powtarzanie pisanych formuł. Excel jest na dla systemów Windows i MacOS. Program został napisany w 1987 przez Dan'a Bricklin'a, jednak nikt nie spodziewał się jego niesamowitego sukcesu. [2]

Excel został napisany początkowo w języku C, a następnie wprowadzono do niego dużo modyfikacji z C++. Wspiera developerów korzystających z C#, który został również stworzony przez Microsoft. Jednak najbardziej popularnym językiem programowania od wielu lat pozostaje Java.

Zagadnienie integracji języka JAVA i MS Excel jest częstym problemem i trudno jednoznacznie określić który sposób jest najlepszy. Każde rozwiązanie oferuje inne możliwości. Praca, którą opisuje w dużej mierze jest nowatorska, ponieważ istnieje bardzo mało źródeł dotyczących tego zagadnienia.



Technologia Java została stworzona przez firmę Sun Microsystems, jego podstawowe koncepcje zostały przejęte z języka C++ - składnia i słowa kluczowe oraz SamlItalk - maszyna wirtualna i zarządzanie pamięcią. Java jest językiem wysokiego o wysokim poziomie abstrakcji, jest językiem kompilowanym do kodu bajtowego, który wykorzystywany jest przez JVM - wirtualną maszynę Java.



Statystyki[4]

- 97% komputerów lokalnych, używanych w przedsiębiorstwach, korzysta z oprogramowania Java
- 87% komputerów lokalnych, używanych w Stanach Zjednoczonych, korzysta z oprogramowania Java
- 9 milionów programistów rozwija oprogramowanie oparte na technologii Java
- Nr 1 dla programistów
- Nr 1 jako platforma programistyczna
- 3 miliardy telefonów komórkowych korzysta z oprogramowania Java
- 100% odtwarzaczy dysków Blu-ray jest dostarczanych wraz z oprogramowaniem Java
- 5 miliardów kart, będących w użyciu, korzysta z oprogramowania Java
- 125 milionów urządzeń telewizyjnych korzysta z oprogramowania Java
- 5 z 5 czołowych producentów OEM dostarcza oprogramowania Java ME

Architektura Java została napisana tak, by pozwolić na programowanie przenośnych oraz wysoce wydajnych aplikacji dla możliwie najszerszego spektrum platform przetwarzania cyfrowego. Dzięki aplikacjom dostępnym w heterogenicznych środowiskach, firmy mogą dostarczać więcej usług oraz zwiększać produktywność, komunikatywność i współpracę użytkowników końcowych, radykalnie zmniejszając koszty związane z prowadzeniem działalności i utrzymywaniem aplikacji konsumenckich.

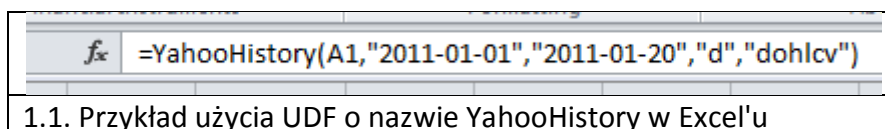
2. Cel i zakres pracy

Moja praca dyplomowa pod tytułem "Integracja Java z MS Excel" ma na celu zbadanie możliwości rozbudowy funkcjonalności programu MicroSoft Excel 2007 o własną implementację UDF tzn. funkcji zdefiniowanej przez użytkownika. Funkcja komunikuje się z Excelem za pomocą biblioteki XLLoop, która tworzy lokalny serwer nasłuchujący protokoły wysyłane z Excela. Następnie wywołuję metodę zaimplementowaną w Javie, która wywołuje proces z RapidMinera zapisany w formacie xml lub rmp. Aplikacja posiada GUI, w którym można dodać własny projekt. Należy ustawić kolejne atrybuty oraz rodzaj danych. Po uruchomieniu programu Serwer lokalny nasłuchuje, uruchomiony dodatek, który przekazuje Excelowi zaimplementowaną funkcjonalność.

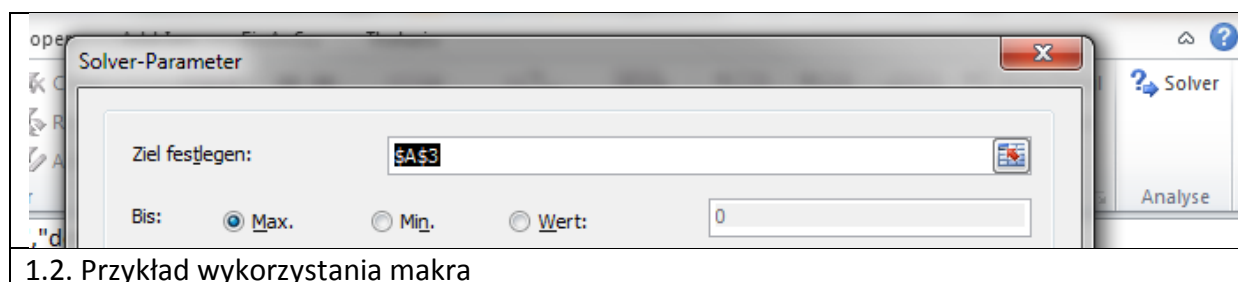
3. Możliwe narzędzia integracji MS Excel i Java

Excel User-Defined Function (UDF) lub makro

Aby wykonać nowe polecenie przez naszą aplikację możemy skorzystać z funkcji napisanej przez użytkownika (UDF) lub z makra. UDF jest to funkcja excelowa, która wykorzystuje komórki excelowe jako argumenty i zwraca jakąś wartość jako wynik obliczenia.



Excelowe makro zachowuje się inaczej - potrafi modyfikować wartości i style pomiędzy wieloma excelowymi komórkami i różnymi arkuszami. Używając makra możemy również tworzyć arkusze, menu, zmieniać kolor itd.



3.1. *Możliwe rozwiązania:*

3.1.1. VBA

Najprostszą drogą do stworzenia UDF'a w MS Excel jest Visual Basic for Applications - w skrócie VBA. Jest to język programowania oparty na Visual Basic'u dodany do aplikacji w pakiecie Microsoft Office, ale także i innych np. AutoCAD. Służy przede wszystkim do automatyzacji pracy z dokumentem. VBA można łatwo używać, lecz nie nadaje się w większych projektach, gdzie naprawialność i debuggowanie jest mocno ograniczone. Przy większych projektach lepiej skorzystać z platformy .Net, łączność z Java jest także możliwa, ale w dużej mierze utrudniona.

3.1.2. .Net

Najbardziej podstawowymi technologiami służącymi do komunikacji z MS Excelem z użyciem C# jest Visual Studio Tools for Office (VSTO) i Primary Interop Assemblies (PIA). Oba te narzędzia wymagają dodatkowego czasu wykonywania programu na komputerze użytkownika. Do tworzenia UDF najbardziej przydatny jest Excel DNA, który również potrafi obsługiwać makra.

3.1.3. Java

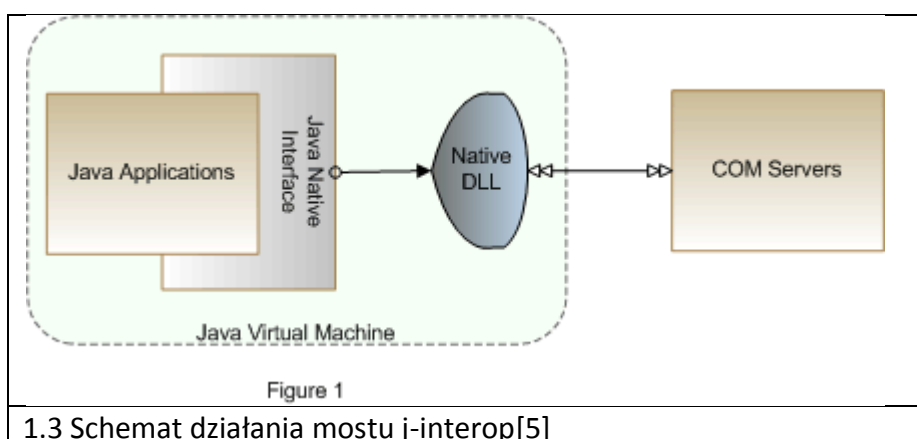
Integracja Java z produktami Microsoft zawsze była problematyczna i wymagała zawiłych rozwiązań. W przypadku generowania plików excelowych przydatna jest biblioteka Apache POI, która pozwala na tworzenie nowych arkuszy, edycję stylów, kalkulacje itd. Natomiast aby osiągnąć komunikację on-line w arkuszu Excela, możemy wykorzystać:

- integrację na poziomie makr - wykorzystuje Excel COM interface. W internecie dostępnych jest wiele bibliotek wspierających mosty COM np. JACOB, com4j i J-Interop. COM (Component Object Model) jest standardem do tworzenia i projektowania interfejsów programistycznych na poziomie binarnym dla każdego modułu wprowadzonym przez Microsoft.
- integrację na poziomie UDF - jest o wiele bardziej skomplikowana niż wezwanie COM. Dobrym open source'owym projektem jest XLLoop. Pozwala na integrację Excela z Java, C, Python, Ruby i wiele więcej. Minusem tego rozwiązania jest konieczność włączenia serwera XLLoop przez użytkownika.
- komunikację za pomocą protokołu SOAP, który pozwala przekazać .Net'owe dodatki w postaci protokołu do kodu Java.

3.2. Wybór technologii.

Projekt powstał z myślą integracji najbardziej popularnego języka programowania na świecie Java z jednym z najczęściej używanych aplikacji w środowisku biznesowym. Java jest podstawą projektowania praktycznie każdego rodzaju aplikacji sieciowych jak również desktopowych. Jest ogólnosięciowym standardem z którego korzysta się do produkcji aplikacji mobilnych, wbudowanych, gier, zawartości i treści internetowych. Ponad 9 milionów developerów Java przyczyniło się do nowoczesnego tworzenia, implementacji oraz wykorzystywania tych aplikacji i usług.[4]

Z początku próbowałem komunikować się z Excelem za pomocą COM bridge, gdyż jest najprostszą i najbardziej "naturalną" metodą z możliwych. Wybrałem najczęściej wykorzystywanym i zarazem najnowszy most COM - j-Interop (<http://www.j-interop.org>). Niestety ~~ale~~ po pierwszych próbach wystąpiła dużo ilość problemów związanych z bezpieczeństwem. Korzystałem z systemu Windows 7 Home Premium i używałem MS Excel 2007. Próbowałem wielu rozwiązań m.in. konfigurację serwisów, komponentów, ustawień domeny i serwera, modyfikacje firewalla. Biblioteka była rozwijana najintensywniej w latach 2006-2007 i ze względu na politykę bezpieczeństwa Microsoft wykorzystanie jej jest bardzo problematyczne. Po wielu konsultacjach wspólnie z promotorem zdecydowaliśmy porzucić to rozwiązanie na rzecz właściwego rozwiązania - XLLoop.



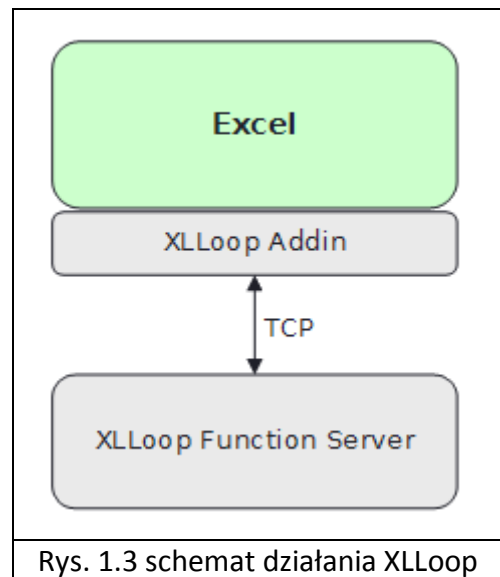
3.3. Biblioteka XLLoop[4]

XLLoop to open sourceowa biblioteka pozwalająca implementować Excelowe funkcje zdefiniowane przez użytkownika (UDF - User-Defined Functions) na zcentralizowanym serwerze (serwerze funkcyjnym).

3.3.1. Zastosowanie

- Funkcje mogą być napisane w każdym języku (w tym Java, języki skryptowe itd.),

- Funkcje mogą być dodane szybko i dynamicznie bez konieczności restartowania Excela,
- Funkcje mogą być zarządzane oddzielnie i centralnie, co zapobiega kosztów zarządzania wieloma XLL'ami i gwarantuje wszystkim użytkownikom używanie tych samych funkcji,
- Dane mogą być dzielone pomiędzy excelowymi sesjami/użytkownikami (np. obecne ceny giełdowe mogą być przechowywane na pojedynczym serwerze i wszystkie sesje Excela mogą uzyskiwać te dane za pomocą przykładowej funkcji GetStock).



Rys. 1.3 schemat działania XLLoop

3.3.2. XLLoop składa się z dwóch głównych komponentów:

- Excelowy dodatek (XLL napisany w C++)
- Serwer i biblioteka napisana w Java (i w wielu innych językach)

Dodatek i serwer komunikują się poprzez zwykłe oparte o schowek protokoły, wysyłające i odbierające (binarnie) zserializowane excelowe obiekty (nazywane xloper).

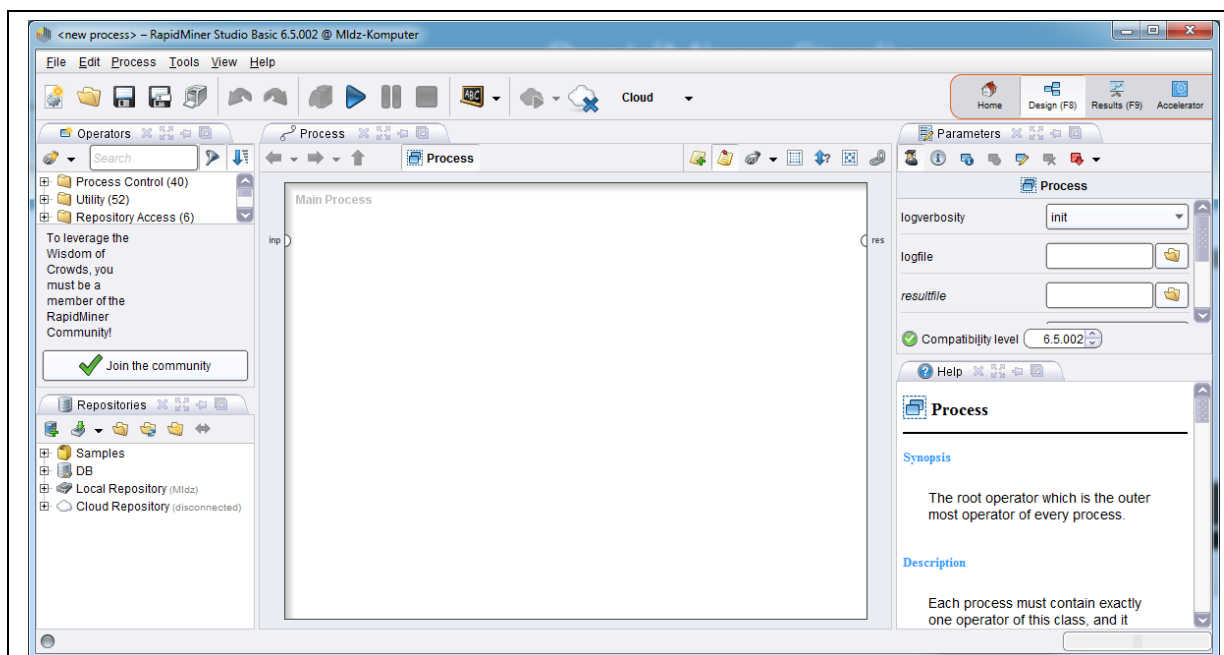
W modyfikacji JSON dodatek i serwer komunikują się ponad protokołami http(s), wysyłając i otrzymując JSON'owe zserializowane excelowe obiekty. XLLoop jest pod licencją
Common Public License (CPL).

4. RapidMiner

RapidMiner to najczęściej używana open sourcowa platforma do predykcyjnej analizy oraz eksploracji danych. Tworzony w nim kod generuje się za pośrednictwem intuicyjnego graficznego interfejsu użytkownika, który pomaga zaprojektować analizowany proces. RapidMiner posiada upublicznione biblioteki służące do rozszerzenia jego funkcjonalności, można je znaleźć na serwerze GitHub.com. Platforma jest otwarta do współpracy z mnó-

stwem źródeł danych w tym: Excel, Access, Oracle, IBM DB2, Microsoft SQL, Netezza Teradata, MySQL, Postgres, SPSS, Salesforce.com. RapidMiner umożliwia nowoczesną analizę w każdej skali, jest wręcz idealne dla Big Data.

RapidMiner jest idealnym narzędziem do współpracy z MS Excel. Jest napisany w Java i posiada możliwość wczytywania arkuszy Excel. W moim projekcie najważniejsze jest by z poziomu Excela - który potrafi obsługiwać przeciętny użytkownik - można było wywołać zaawansowane obliczenia (procesy) z wykorzystaniem funkcjonalności RapidMinera.



Rys. 1.4 GUI Rapidminera

5. Szczegóły implementacyjne projektu

6. Przykład użycia

7. Wnioski i podsumowanie

8. Bibliografia

1 - Computerworld: [Polski Word 2.0 i Excel 4.0 już w sprzedaży](#). [dostęp 2009-01-31]

2 - <http://naukaexcel.pl/podstawy/kurs-o-historii-excela-jak-to-sie-stalo-ze-excel-zostal-krolem-na-ryнку-arkuszy-kalkulacyjnych/>

3 - <http://computeraidedfinance.com/2011/11/26/what-shall-i-choose-for-implementing-an-excel-plugin-and-udf/>

4 - <https://www.java.com/pl/about/> [dostęp 2015-12-27]

5 - <http://www.j-interop.org/introduction.html>

1 - Computerworld: [Polski Word 2.0 i Excel 4.0 już w sprzedaży](#). [dostęp 2009-01-31]

2 - https://pl.wikipedia.org/wiki/Microsoft_Excel#cite_note-1 [dostęp 2016-01-01]

3 - <https://www.java.com/pl/about/> [dostęp 2015-12-27]

4 - <http://xlloop.sourceforge.net/> [dostęp 2015-12-22]

5 - <https://rapidminer.com/> - zebrane i opracowane materiały przez autora pracy i przetłumaczone [dostęp 2016-01-02]