

Akademia Górniczo - Hutnicza im. Stanisława Staszica w Krakowie
Wydział Elektrotechniki, Automatyki, Informatyki i Elektroniki
Katedra Informatyki



Tytuł pracy

**Integracja systemów przetwarzania mowy w środowisku
zgodnym z paradygmatem SOA**

Konrad Dziedzic, Rafał Fronczyk
rfronczyk@gmail.com, konraddziedzic@gmail.com

Promotor pracy: dr inż. Łukasz Czekierda

Kraków, 2012

Oświadczenie

Oświadczamy, świadomi odpowiedzialności karnej za poświadczenie nieprawdy, że niniejszą pracę dyplomową wykonaliśmy osobiście i samodzielnie (w zakresie wyszczególnionym we wstępie) i że nie korzystaliśmy ze źródeł innych niż wymienione w pracy.

.....

podpis i data

podpis i data

Przedmowa

Rozwój technologiczny, a w szczególności rozwój komputerów i technologii z nimi związanych w ciągu ostatnich lat jest bardzo szybki. Zmieniają się zarówno, podzespoły, moc obliczeniowa, rozmiary, wygląd a nawet interfejsy. Komputery wkroczyły, lub są temu bardzo bliskie, do niemal każdej dziedziny życia. W związku z tym zmienia się sposób komunikacji między użytkownikiem a komputerem. W ostatnich czasach można zaobserwować dążenie inżynierów i projektantów do uczynienia komunikacji z komputerem jak najbardziej naturalną. Pomysły są różne od "kontrolerów bez kontrolera" jak Microsoft Kinect ¹, poprzez rozwiązania żuchowe" podobne do tego na jakie zdecydowało się Sony ² w swoim kontrolerze PlayStation Move ³ poprzez klasyczne jak mysz i klawiatura. Jednak najbardziej naturalnym sposobem porozumiewania się wydaje się głos. Najbardziej znanym, bo na pewno nie pionierskim, systemem który komunikuje się z użytkownikiem za pomocą głosu jest Apple Siri ⁴ - osobisty asystent, "żyjący" wewnątrz systemu, umożliwiający dostęp do jego funkcji za pomocą mowy.

Niniejsza praca przedstawia prototyp systemu UniversalSynthesizer, służącego do zamiany tekstu na dźwięk i mowy na tekst, będącego zaawansowaną, rozproszoną, rozszerzalną, wielojęzyczną platformą/serwisem umożliwiającą łatwe tworzenie zróżnicowanych, wieloplatformowych aplikacji, mających różne zadania. Celem pracy nie było stworzenie gotowego do użytku, kompletnego, w pełni sprawnego produktu. Powstałą aplikację należy traktować bardziej jako punkt wyjścia, prototyp który w przyszłości może być wykorzystany do zbudowania w pełni funkcjonalnego, ogólnie dostępnego,

¹<http://www.xbox.com/en-US/kinect>

²<http://www.sony.com/>

³<http://us.playstation.com/ps3/playstation-move/>

⁴<http://www.apple.com/iphone/features/siri.html>

komercyjnego lub otwartego systemu. W związku z tym niektóre zagadnienia, dość istotne z punktu widzenia potencjalnego odbiorcy, ale nie będące ściśle powiązane z celem pracy zostały pominięty lub też niedopracowane.

Spis treści

1	Wstęp	1
1.1	Definicja problemu	1
1.1.1	Obszar badań	1
2	Aims of the project	3
2.1	Final aim	3
2.2	Preliminary aims	3
3	Discussion	5
4	Materials & methods	7
	Bibliografia	9
	Spis rysunków	11
	Spis tabel	13

SPIS TREŚCI

1

Wstęp

1.1 Definicja problemu

Komputerowe przetwarzanie mowy jest w ostatnim czasie szybko rozwijającą się dziedziną w przemyśle informatycznym. Wykorzystuje się je w wielu różnych dziedzinach zarówno związanych z rozrywką (rozmaite gry polegające na jak najwierniejszym zaśpiewaniu podanego utworu), z życiem codziennym (Apple Siri czyi wirtualny asystent) jak i z pracą (różne systemy telefoniczne działające bez udziału człowieka). W wielu sytuacjach wykorzystanie komputerowo generowanej mowy jest niezastąpione.

Podobnie sytuacja wygląda z wykorzystaniem architektury SOA. Jest to jedno z najważniejszych osiągnięć informatycznych ostatnich lat. Podejście to zdecydowanie dominuje w obecnie projektowanych architekturach. Jednym z głównych powodów takiego stanu rzeczy jest uproszczenie modelu współczesnych dużych aplikacji. Umożliwia ono podzielenie systemu na serwisy które są niezależnymi, samodzielnymi, łatwymi w zarządzaniu, posiadającymi jasno i jednoznacznie zdefiniowane interfejsy encjami, które można wykorzystać w różnej kolejności w celu otrzymania różnych efektów.

Celem tej pracy jest zbadanie możliwości wykorzystania nowoczesnej i prężnie rozwijającej się architektury do stworzenia gotowego systemu ułatwiającego i ujednolicającego proces tworzenia aplikacji przetwarzających mowę.

1.1.1 Obszar badań

W skład zakresu pracy wchodzi zarówno technologie związane z architekturą SOA jak i z przetwarzaniem mowy. W celu użycia odpowiednich narzędzi przeprowadzono wiele

1. WSTĘP

badan i testów w czasie których dużą wagę przykładano do:

- możliwości konfiguracyjnych
- wydajności
- łatwości rozbudowy
- dostępnej dokumentacji i pomocy technicznej

W efekcie przeprowadzonych czynności okazało się, że najlepsze efekty uzyskać można korzystając z szyny ESB Apache ServiceMix, syntezytorów mowy Ivona i FreeTTS oraz innych mniej istotnych narzędzi potrzebnych do stworzenia architektury zgodnej z SOA oraz zdolnej do efektywnego wykonywania postawionych przed nią zadań.

2

Aims of the project

2.1 Final aim

Our ultimate goal is...

2.2 Preliminary aims

There will be several preliminary scientific targets to be accomplished on the way...

2. AIMS OF THE PROJECT

3

Discussion

3. DISCUSSION

4

Materials & methods

4. MATERIALS & METHODS

Bibliografia

[1] LASTNAME. **Title.** *Journal of Sth*, 2007.

[2] NAME. **Title.** *Journal of Sth*, 2006.

BIBLIOGRAFIA

Spis rysunków

SPIS RYSUNKÓW

Spis tabel