

České vysoké učení technické v Praze  
Fakulta elektrotechnická

# Household Intelligent Assistant



Autoři: Jiří Burant  
Jakub Drápela  
Martin Klučka  
Petr Kovář  
Jakub Konrád  
Pavel Trutman

Studijní obor: Kybernetika a robotika

Datum vypracování: 21. dubna 2016

## Obsah

<b>1</b>	<b>Úvod</b>	<b>2</b>
<b>2</b>	<b>Aktuální stav projektu</b>	<b>2</b>
2.1	Popis prototypu . . . . .	2
2.2	Stav projektu v rámci časového plánu . . . . .	2
<b>3</b>	<b>Detailní popis projektu</b>	<b>3</b>
3.1	Activation word recognition a Speech-to-text . . . . .	3
3.2	Sentence processing . . . . .	4
3.3	Data query a Answer formulation . . . . .	5
3.4	Text-to-speech . . . . .	5
3.5	Application kernel . . . . .	5
3.6	Řešení pomocí virtuálního stroje . . . . .	6
<b>4</b>	<b>Neočekávané události</b>	<b>6</b>
<b>5</b>	<b>Detailní plán do konce projektu</b>	<b>7</b>

## 1 Úvod

Cílem této zprávy je zhodnotit současný stav projektu *Household Intelligent Assistant* a dosud dosažené výsledky, poskytnout detailní popis jeho jednotlivých součástí a stanovit časový plán, který povede k úspěšnému dokončení projektu v rámci naší časové dotace.

## 2 Aktuální stav projektu

V současnosti existuje funkční prototyp osobního asistenta *Phoenix*. Tento prototyp je umístěn ve virtuálním stroji na systému *Debian*. Vytvořený prototyp dokáže úspěšně vnímat dotazy z oblasti počasí, vyhodnotit je a reagovat na ně.

### 2.1 Popis prototypu

Prototyp systému po spuštění vyčkává v režimu stand-by a čeká na aktivaci oslovením. Prototyp reaguje na oslovení klíčovým slovem (Phoenix) a poté přepne do módu zaznamenávání dotazu. Prototyp zaznamená mluvený dotaz, převede jej do textové podoby, vyhodnotí jej a získá jeho význam. Dotaz je vyhodnocen a odpověď je získána pomocí webové aplikace třetí strany. Odpověď je poté převedena do textové podoby. Systém syntetizuje mluvenou odpověď ze získaného textu a odpoví uživateli.

Prototyp byl důkladně otestován členy projektu a během testování se objevily následující problémy:

- V neoptimálních zvukových podmínkách má systém problém s rozpoznáváním klíčového slova. Systém reaguje i pokud nebyl osloven klíčovým slovem.
- Prototyp má na některých zařízeních problém přijmout zvukový vstup.
- Systém se v některých případech po úspěšném zodpovězení dotazu nevrátí do stand-by režimu.

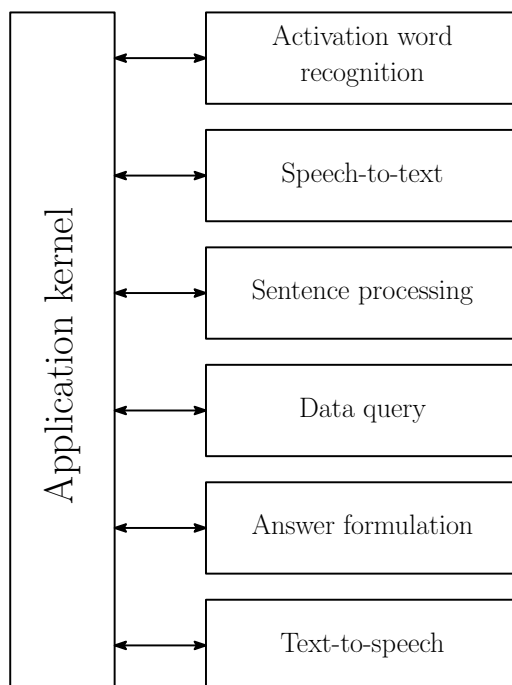
### 2.2 Stav projektu v rámci časového plánu

Dle harmonogramu projektu se nacházíme ve fázi před druhým milníkem projektu, tedy před vytvořením plně funkčního systému. Toto odpovídá skutečnému současnému stavu projektu. Už nyní existuje funkční verze systému. Je nutné odstranit výše zmíněné problémy a jsme připraveni na

vstup do finální fáze. Vytýčený plán se nám dosud podařilo dodržovat bez větších potíží. Věříme, že tomu tak bude i nadále a projekt bude ve stanoveném termínu úspěšně dokončen.

### 3 Detailní popis projektu

Systém *Phoenix* je navržen modulárně tak, aby bylo možné jednoduše rozšiřovat a měnit jeho funkcionalitu v závislosti na požadavcích uživatele. Celý systém je naimplementován v jazyce Python a je rozdělen do několika základních modulů, které mezi sebou komunikují pomocí jádra aplikace (obrázek 1). Následuje detailní popis řešení a funkce jádra a jednotlivých modulů:



Obrázek 1: Blokový diagram řešení projektu Household Intelligent Assistant.

#### 3.1 Activation word recognition a Speech-to-text

S využitím knihovny *PocketSphinx* byly v Pythonu implementovány mechanismy pro rozpoznávání klíčového slova (Keyword attention) a převod mluveného slova na text (Speech to text). V obou případech bylo třeba řešit

řadu minoritních problémů - za zmínku stojí například volba a implementace I/O knihovny pro práci s mikrofonním vstupem počítače. Oba výše zmíněné mechanismy jsme dále přepsali do samostatných modulů kompatibilních s jádrem aplikace (rozhraní *Zero MQ*).

Modul pro detekci klíčového slova je od počátku nedílnou součástí navrženého řešení, a tak byl i jako jeden z prvních modulů propojen s jádrem. Za tři týdny testování jsme naměřili vysokou přesnost detekce, která je ale podmíněná vhodnou volbou parametrů pro daný mikrofon. Jako klíčové slovo slouží nadále slovo Phoenix s tím, že po blížícím se testování je možná jeho změna. V průběhu jsme se také potýkali s nestabilitou knihovny *PyAudio*, která zajišťuje práci s mikrofonom. V současné době probíhá hledání a testování alternativních knihoven.

Modul pro převod mluveného slova na text je v současnosti implementován pomocí nástroje *Wit.ai*. Tuto implementaci v blízké době nahradí již připravený a samostatně otestovaný modul pro převod mluveného slova implementovaný pomocí *PocketSphinx*. Od tohoto kroku si slibujeme rychlejší odezvu systému, neboť převod (na rozdíl od předchozího řešení) probíhá lokálně a offline. Přesnost převodu je opět na velmi dobré úrovni, kterou je případně možné dále vylepšit volbou gramatiky (Language modelu) zaměřené na námi podporované domény. Podobně jako u modulu pro detekci klíčového slova, tak i zde se počítá se záměnou knihovny *PyAudio* za některou z alternativ.

### 3.2 Sentence processing

Modul pro parsování a zpracování dotazů byl naimplementován s využitím nástroje *Wit.ai*. Modul přijme dotaz v textové formě a pomocí Pythonovské knihovny se připojí k aplikaci *Wit.ai*, která pomocí námi vytvořené databáze rozebere dotaz a získá jeho význam (intent).

Aplikaci *Wit.ai* jsme zvolili, protože s její pomocí se nám podařilo dosáhnout velmi dobrých výsledků při vyhodnocování dotazu a dále díky jednoduchým možnostem rozšíření její databáze. Databáze je v současné době zaměřena pouze na počasí, avšak do konce projektu ji plánujeme rozšířit i o další oblasti.

Za zmínku stojí, že databáze se kterou aplikace *Wit.ai* pracuje je volně přístupná. Uživatel má možnost vytvořit si její kopii (fork) a modifikovat/doplnit ji tak, aby lépe vyhovovala jeho potřebám.

### 3.3 Data query a Answer formulation

Modul, který obstarává vyhledávání informací a logiku, která má být vykonána na základě intentu se nazývá „*querylogic*“. Tento modul dostane od jádra jako vstup zpracovaný dotaz z *Wit.ai* a podle jeho významu vykoná žádané úkony a vyhledá potřebné informace. Následně vrátí zpět do jádra sestavenou odpověď ve formě textového řetězce.

V současné době je modul *querylogic* schopen obsluhovat základní dotazy z oblasti počasí. K tomu využívá internetové API *Forecast.io*, kterému posílá webový požadavek. Odpovědí je JSON objekt, ze kterého jsou vyčtena potřebná data, jako teplota, síla větru apod. Dále tento modul používá knihovnu *Geopy*, která slouží k získání geografických souřadnic daného místa, pakliže ho tazatel specifikuje.

Modul je schopen poskytnout informace o počasí na týden dopředu pro různá místa na zemi a je schopen poskytnout i historické údaje za posledních šedesát let, pokud pro dané místo existují.

Dále je rozpracován submodul obecné komunikace, který bude obstarávat některé běžné konverzační otázky, například: “What is the time?“, “Tell me a joke.“, apod.

### 3.4 Text-to-speech

Při realizaci tohoto modulu se jako největší problém ukázalo nalezení vhodné knihovny pro syntézu řeči. Z počátku jsme využívali knihovnu *Pyttssx* v kombinaci s linuxovým ovladačem zvuku *Espeak*. Zde se nám však nepodařilo dosáhnout požadované kvality syntetizované řeči a proto jsme se rozhodli od použití této knihovny odstoupit.

V současné době využívá náš modul syntetizátor řeči *Mimic* od společnosti *Mycroft*. Tato knihovna má oproti předchozímu řešení výhodu jednoduššího nastavení parametrů. S kvalitou syntetizovaného hlasu pomocí knihovny *Mimic* jsme spokojeni. Z řeči je sice patrné, že je počítačové povahy, avšak je srozumitelná a jedná se o výrazné zlepšení oproti předchozímu řešení.

### 3.5 Application kernel

Jakmile byly vytvořeny všechny prototypy modulů, bylo třeba je zakomponovat do sebe tak, aby tvořily funkční aplikaci. K tomu slouží jádro aplikace, které s jednotlivými moduly komunikuje pomocí definovaného API. Jednotlivé moduly jsou jádrem spouštěny jako samostatné procesy, což dává velkou svobodu pro návrh a implementaci jednotlivých modulů, protože nejsou s jádrem vůbec svázány například vybraným programovacím jazykem.

K meziprocessorové komunikaci mezi jádrem a moduly jsme použili knihovnu Zero MQ, která je rychlá a má jednoduché rozhraní pro většinu programovacích jazyků.

Skutečnost, že jednotlivé moduly komunikují s jádrem přes API, je výhodná i v tom, že bylo-li potřeba některý modul nahradit za jiný, například z důvodu špatné funkčnosti použité knihovny, tak se vytvořil nový modul a v nastavení jádra aplikace stačilo změnit cestu k novému modulu. Žádné složité úpravy jádra nebyly třeba. Tato velká modularita umožňuje případnému zájemci si kterýkoliv modul jednoduše upravit či vytvořit nový, aniž by musel podrobně zkoumat strukturu ostatních modulů nebo jádra aplikace.

### 3.6 Řešení pomocí virtuálního stroje

Protože jednotlivé moduly používají knihovny, jejichž instalace není snadná a liší se v závislosti na druhu operačního systému, vytvořili jsme virtuální stroj pro nástroj VirtualBox. Tento virtuální stroj běží na operačním systému Debian Jessie a obsahuje všechny nainstalované knihovny a balíčky potřebné pro běh naší aplikace, která je ve virtuálním stroji připravena ke spuštění.

Vytvoření virtuálního stroje nám umožnilo vytvořit jednotné testovací rozhraní pro naši aplikaci a odbourali jsme tím mnoho problémů spojených s instalací potřebných knihoven na různých operačních systémech. Každý člen týmu tedy nemusel ztrácet čas zdlouhavou instalací všech knihoven na svůj počítač, ale stačilo mu jen si stáhnout obraz virtuálního stroje a ten spustit. Myšlenka virtuálního stroje nám také umožňuje jednoduše aplikaci nasdílet dalším osobám mimo náš tým, které mohou aplikaci jednoduše testovat a nám tak poskytovat cennou zpětnou vazbu.

Pro sestavení virtuálního stroje používáme nástroj Vagrant, který velmi usnadňuje tvorbu a práci s virtuálními stroji. Pomocí Vagrantu si stáhneme již předinstalovaný operační systém, na který jednoduše nainstalujeme potřebné knihovny a naši aplikaci pomocí konfiguračního nástroje Ansible. Díky použití těchto dvou nástrojů můžeme jednoduše změnit operační systém na kterém aplikace běží přičemž nás to bude stát jen minimální úsilí. Takto tedy můžeme velmi efektivně testovat chování aplikace na různých operačních systémech.

## 4 Neočekávané události

V průběhu projektu došlo k následujícím neočekávaným událostem:

- Změna rozhraní Pythonovského modulu třetí strany *Pywit* pro použití aplikace *Wit.ai*.

Jak již bylo zmíněno, náš projekt využívá *Wit.ai* k parsování dotazu a získání jeho významu. Původně však *Wit.ai* zajišťoval také převod řeči na text. S vydáním nové verze *Pywit* byla tato funkcionality omezena. V rámci našeho projektu bylo tedy nutno zareagovat a najít alternativní řešení.

Tento problém jsme přednesli na pravidelné týdenní schůzce a po dohodě jsme přerozdělili plánované úkoly na další týden tak aby došlo k vyřešení problému a zároveň nedošlo ke zpoždění projektu oproti harmonogramu.

## 5 Detailní plán do konce projektu

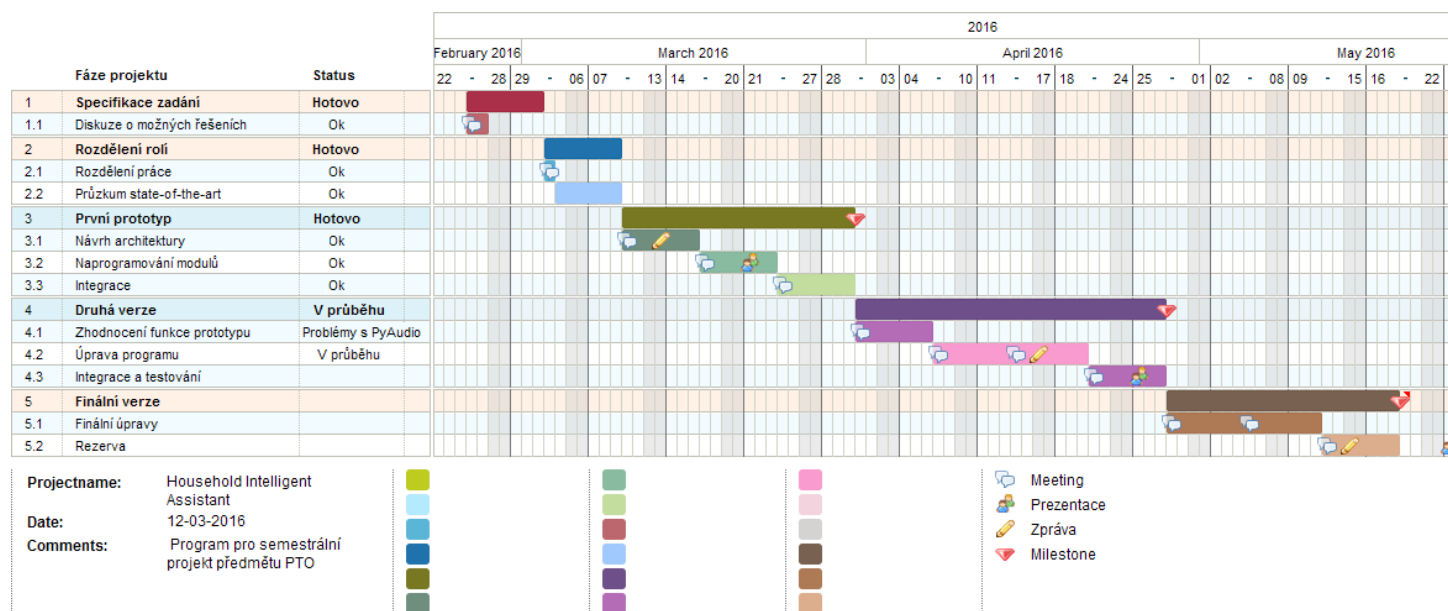
Nyní se nacházíme ve druhé fázi projektu (viz harmonogram na obrázku 2), na jejímž konci bychom měli mít plně funkční systém. Tato fáze dle plánu končí 28. 4. 2016. Do tohoto data se tedy budeme věnovat především odstranění již zmíněných problémů prototypu. Jmenovitě: odstranění problémů s nahráváním zvuku pomocí *PyAudio*, zajištění spolehlivosti při rozpoznávání klíčového slova a zajištění přechodu systému do režimu standby po zodpovězení dotazu.

Po vyřešení těchto problémů a přechodu do finální verze projektu, která má probíhat od 29. 4. 2016 do 19. 5. 2016 nás čekají dva hlavní úkoly:

- Rozšíření funkcionality a přidání jednoduchých “obecných” modulů pro zajištění větší flexibility systému (např.: time modul - sdělí uživateli čas, joke modul - řekne uživateli vtip z databáze, atd.).
- Důkladné testování systému. Systém bude podroben důkladnému testu, a to i uživateli mimo skupinu vývojářů projektu tak, aby došlo k zachycení a opravení jeho případných chyb a nedostatků.



8



Obrázek 2: Harmonogram projektu graficky znázorněn formou Ganttova diagramu s klíčovými milníky.