

České vysoké učení technické v Praze

Fakulta elektrotechnická

Household Intelligent Assistant

Závěrečná zpráva



Autoři: Jiří Burant

Jakub Drápela

Martin Klučka

Petr Kovář

Jakub Konrád

Pavel Trutman

Studijní obor: Kybernetika a robotika

Datum vypracování: 19. května 2016

Obsah

1	Úvod	2
2	Finální stav projektu	2
3	Popis aplikace	2
4	Finální řešení	3
4.1	Activation word recognition	3
4.2	Speech-to-text	3
4.3	Sentence processing	3
4.4	Data query a Answer formulation	3
4.5	Text-to-speech	4
5	Stav projektu v rámci časového plánu	4
6	Testy a zhodnocení výkonu	4
6.1	Activation word recognition - testování	5
6.2	Speech-to-text - testování	6
6.3	Sentence processing - testování	7
6.4	Testování celého systému	7
6.5	Shrnutí výsledků testů	8
7	Jak náš projekt vyzkoušet a otestovat	9
8	Závěrečné zhodnocení projektu	9

1 Úvod

Cílem této zprávy je provést závěrečné zhodnocení projektu Household Intelligent Assistant. V rámci zprávy shrneme finální stav produktu a porovnáme dosažené výsledky se zadáním práce. Dále se zaměříme na podrobný popis finálního řešení a zároveň připomeneme jednotlivé milníky vývoje, které nás dovedly k finálnímu produktu.

2 Finální stav projektu

Projekt Household Intelligent Assistant se podařilo dokončit v termínu. V současné době existuje beta verze aplikace Household Intelligent Assistant, která je již plně funkční v rámci zamýšlené domény. Aplikace tedy reaguje na široké spektrum dotazů týkajících se počasí a v omezené podobě také na dotazy z dalších oblastí. Než bude možné vydat oficiální release, je třeba nasbírat více dat od skutečných uživatelů a vyřešit případné dosud neobjevené problémy zjištěné pomocí zpětné vazby. Aktuálně již probíhá uvolnění této beta verze veřejnosti prostřednictvím služby GitHub. Jediným podporovaným systémem, na kterém je aktuálně možné aplikaci spustit je Linux.

3 Popis aplikace

Aplikace po spuštění vyčkává v stand-by režimu na zaznění klíčového slova “Phoenix”. Jakmile aplikace zaznamená klíčové slovo, uživatel je zvukově upozorněn a může položit zamýšlený dotaz. Dotaz uživatele v mluvené podobě je aplikací převeden na text. Text je dále zpracován za účelem získání záměru. Na základě získaného záměru se vyhledá příslušná odpověď prostřednictvím webové služby. Ze získaných informací je sestavena odpověď a ta je následně prostřednictvím mluveného slova předána uživateli. Na konci tohoto cyklu aplikace přechází zpět do stand-by režimu a opětovně vyčkává na zaznění klíčového slova. Prototyp byl důkladně otestován členy týmu a na základě získané zpětné vazby byly jednotlivé moduly laděny tak, aby bylo dosaženo co nejlepších výsledků. Výsledky měření jsou uvedeny v samostatné sekci.

4 Finální řešení

Následující sekce stručně popisuje finální řešení jednotlivých modulů a součástí projektu.

4.1 Activation word recognition

4.2 Speech-to-text

Ve finální verzi jsme pro Speech-to-text modul využily knihovnu `SpeechRecognition`, která je napsaná přímo pro programovací jazyk Python. Knihovna využívá pro záznam zvuku knihovnu `PyAudio` a pro řečovou syntézu může použít více knihoven (CMU Sphinx, IBM Speech To Text, Google Speech Recognition a další). V našem projektu je použita knihovna od společnosti Google. Knihovna `SpeechRecognition` pracuje spolehlivě a řečová syntéza je na velmi vysoké úrovni, která splňuje námi stanovené požadavky.

4.3 Sentence processing

Tento modul je už od první fáze projektu realizován pomocí aplikace *wit.ai*. Modul přijme dotaz v textové podobě a vrátí strukturovaný objekt obsahující význam daného dotazu. Aplikace *wit.ai* porovnává dotaz s námi vytvořenou databází a vrátí význam vyhodnocený jako nejpřesnější shodu spolu s "confidence score", pomocí kterého následující moduly určí, zda je dotaz validní, či ne.

4.4 Data query a Answer formulation

Modul, který obstarává vyhledávání informací a logiku, která má být vykonána na základě intentu (záměru) 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.

Hlavní částí modulu *querylogic* je obsluha dotazů z oblasti počasí. Modul je schopen poskytnout historické informace o počasí a týdenní předpověď počasí pro různá místa na zemi, pokud má potřebná data k dispozici. Ve finální verzi je implementován souhrn počasí za určitý časový interval, který v předchozí verzi chyběl. Nyní je možné ptát se nejen na předpověď pro daný den a čas, ale také na časový interval a modul shrne celou předpověď do stručné odpovědi.

Byl dokončen submodul obecné komunikace, který je schopen obstarávat dotazy na čas a vybírat vtipné fráze z vlastní databáze.

Dále byl vytvořen jednoduchý submodul, který je schopen vyhledat aktuální novinky ze světa.

Zároveň jsou v modulu *querylogic* odfiltrovány dotazy, ke kterým *Wit.ai* přiřadí nepodporovaný význam, a na které tedy modul není schopen správně odpovědět, a také dotazy s "confidence score" pod nastaveným prahem. Prahové hodnoty pro rozhodnutí byly získány pomocí testování sady žádoucích a nežádoucích dotazů.

4.5 Text-to-speech

V modulu *text-to-speech* nedošlo v poslední fázi projektu k výrazným změnám. Modul využívá syntetizátor řeči *Mimic* od společnosti *Mycroft*. Kvalita syntetizované řeči není sice srovnatelná s komerčním softwarem a je patrné, že řeč je strojového původu, ale je plně srozumitelná a pro naše použití tak dostačující.

5 Stav projektu v rámci časového plánu

Jak již bylo řečeno v úvodu, projekt se podařilo dokončit dle sestaveného harmonogramu. V současné době tedy máme plně funkční aplikaci testovanou interně mezi členy týmu, kterou od oficiálního releasu dělí už jen plánované testování komunitou uživatelů. Vzhledem k neočekávaným problémům týkajících se modulu pro převod řeči na text, které si vyžádaly plné nasazení všech členů týmu, bylo nutné zkrátit testovací dobu z plánovaných tří týdnů na dva. Díky reorganizaci týmu se nám povedlo zamýšlené testování provést i v tomto zkráceném čase.

6 Testy a zhodnocení výkonu

Projekt prošel několika testovacími fázemi. Kromě běžného testování funkcionality jednotlivých programových součástí všech modulů, jsme navrhli experimenty pro otestování každého z modulů *Activation word*, *Speech-to-text* a *Sentence processing*. K samostatnému testování byly vybrány tyto moduly, protože právě na nich závisí, zda náš systém zpracuje dotaz uživatele, a zda-li má smysl u nich mluvit o úspěšnosti. Dále byl systém otestován jako celek, tedy od zavolání systému pomocí "attention word" až po jeho odpověď.

6.1 Activation word recognition

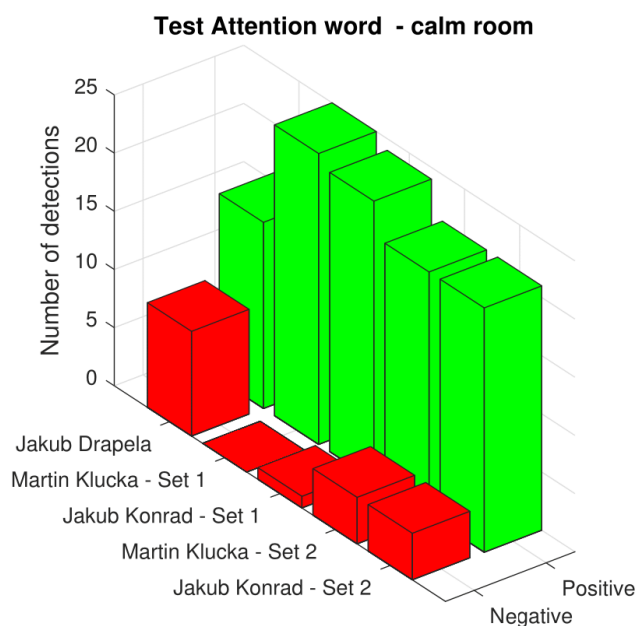
Moduly závisující na speech recognition, tedy *Activation word* a *Speech-to-text* byly testovány ve dvou prostředích, a to v místnosti bez echa a běžného akustického pozadí (klidná místnost) a poté v běžném prostředí s akustickým pozadím (hlučná místnost).

Modul byl testován ve dvou fázích. Nejprve byl vždy osloven pomocí klíčového slova, a bylo zaznamenáno, zda úspěšně zareagoval, či ne a poté byl ponechán aktivní a byla testována jeho falešně pozitivní reakce.

V blízkosti mikrofonu jsme se bavili 15 minut v anglickém jazyce a 15 minut v českém jazyce. Dále jsme zkoušeli systém oslovovat slovy podobnými aktivačnímu "*Phoenix*".

Testování se účastnili vždy 3 muži ve věku 20-25 let a proběhlo 125 dotazů na klíčové slovo pro klidnou místnost a 60 dotazů pro hlučnou místnost.

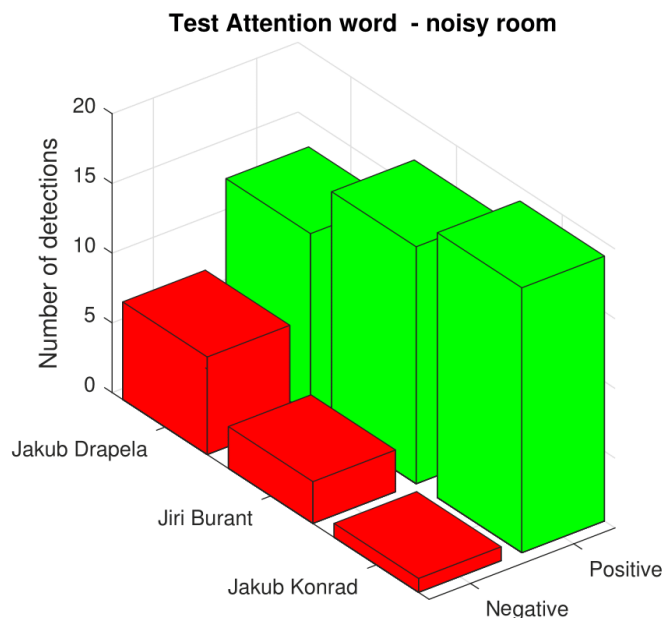
Při testech rozpoznání klíčového slova v klidné místnosti jsme měli celkovou úspěšnost 86 %. K falešně pozitivní detekci došlo pouze jednou. Výsledek testu je znázorněn na grafu 1.



Obrázek 1: Výsledky rozpoznání klíčového slova na 3 respondentech v klidné místnosti.

V hlučné místnosti s ozvěnou jsme byli schopni naším systémem

detekovat 80 % dotazů s klíčovým slovem. Objevili se také falešné detekce, ke kterým došlo celkem 4-krát během celého testování. Výsledek testu je znázorněn na grafu 2.



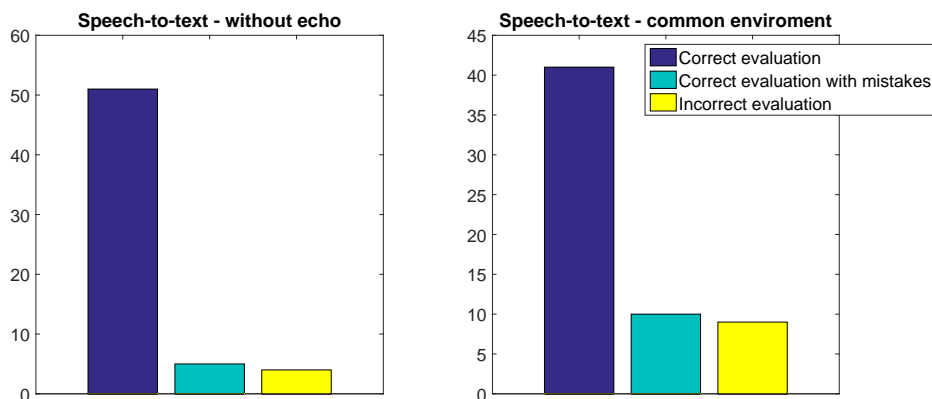
Obrázek 2: Výsledky rozpoznání klíčového slova na 3 respondentech v hlučné místnosti.

6.2 Speech-to-text

Modulu bylo v obou verzích testu předloženo 60 dotazů a ze získaných odpovědí byla vyhodnocena jeho přesnost. Výsledky byly rozděleny do tří kategorií - správně vyhodnocené dotazy, dotazy s drobnými odchylkami (rozpoznáno "is" místo "was", špatně rozpoznáný místní název jako Liberec nebo Krkonoše), které lze pro naše účely s ohledem na možnosti systému stále považovat za správně vyhodnocené a dotazy vyhodnocené nesprávně. Výsledky testů je možno vidět v grafu na obrázku.

3.

Pro testování v běžném prostředí (s mírnou ozvěnou a hlukem v pozadí) lze pozorovat nepatrné zhoršení výsledků oproti nezašuměnému prostředí, tedy 51 dotazů bylo správně vyhodnoceno a 9 dotazů bylo špatně vyhodnoceno oproti 56 a 4. Tuto úspěšnost považujeme za velmi dobrý výsledek pro oba experimenty.



Obrázek 3: Výsledky testování modulu Speech-to-text.

Při testování navíc pravděpodobnost rozpoznání dotazu zvyšuje správná artikulace a dostatečná hlasitost řeči.

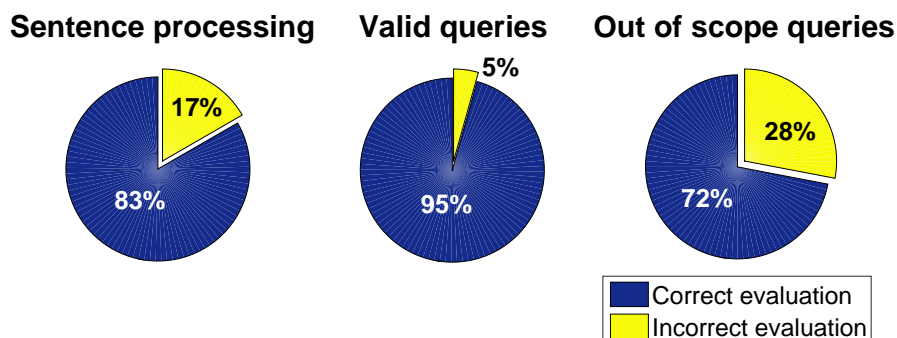
6.3 Sentence processing

Modul *Sentence processing* byl otestován pomocí vytvořené sady testovacích dotazů. Výsledky modulu byly porovnány s očekávanými výsledky sady a následně vyhodnoceny. Testovací sada se skládala z 70 dotazů a obsahovala jak dotazy na které má systém umět odpovědět, tak dotazy které jsou mimo scope systému. Pokud systém odpověděl dle očekávání, byl dotaz vyhodnocen jako správně zodpovězený. Pokud se systém odlišil, odpověď byla vyhodnocena jako špatná. Výsledky testu systému lze vidět na níže uvedeném grafu (obrázek 4).

Z výsledků lze pozorovat, že pro validní dotazy dosahuje náš systém velmi vysoké úspěšnosti - až 95 procent. Pro dotazy, na které systém neumí odpovědět lze pozorovat zvýšenou míru falešně pozitivních vyhodnocení. Toto je způsobeno zatím neimplementovanou podporou pro filtraci nepodporovaných významů v nástroji *wit.ai*. I přes tento nedostatek se nám podařilo dosáhnout správného rozpoznání nevalidního dotazu v 72 procentech případů.

6.4 Testování celého systému

Následující test byl proveden v běžném prostředí s akustickým pozadím. Systém byl testován stejným způsobem, jakým by jej používal běžný



Obrázek 4: Výsledky testování modulu sentence processing.

uživatel, tedy byl zavolán, byla mu položena otázka a byla zaznamenána odpověď. Celý průběh jednoho dotazu byl poté vyhodnocen. Systém byl dotázán celkem 45-krát.

V grafu 6 vidíme, že systému se podařilo odpovědět na položený dotaz v 62 procentech případů. Tato úspěšnost je poměrně nízká, je však potřeba vzít v úvahu, že z 17 neúspěšných testů jich 8 selhalo na oslovení systému, v tomto případě uživatel systém pouze osloví znovu a po úspěšné registraci klíčového slova může pokračovat. Pokud bychom vypustili testy, kde bylo první oslovení systému neúspěšné, dostali bychom se s úspěšností na 75 procent, což už se jeví jako přijatelnější výsledek.

6.5 Shrnutí výsledků testů

Souhrnně lze říci, že výsledky testů systému byly pozitivní. Jednotlivé moduly si vedly velmi dobře. Systém jako celek má výsledky horší, ale stále přijatelné.

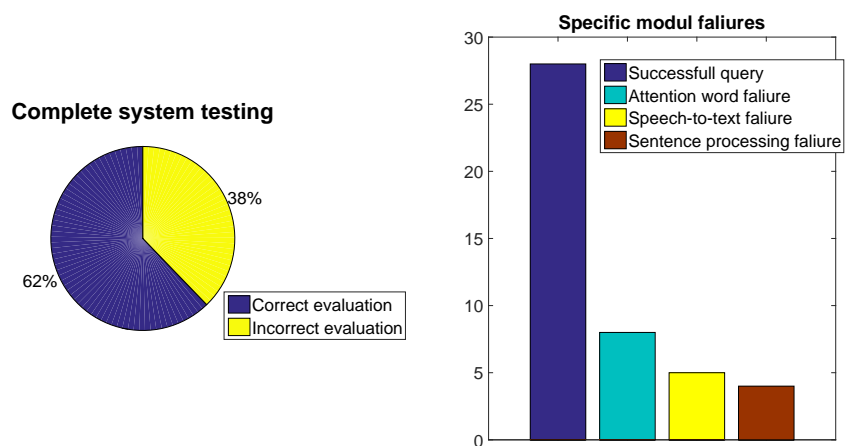
Věříme, že s dostatkem času je možné výsledky systému zlepšit a systém zrobusnit tak, aby se jeho selhání stalo pouze ojedinělým jevem.

7 Jak náš projekt vyzkoušet a otestovat

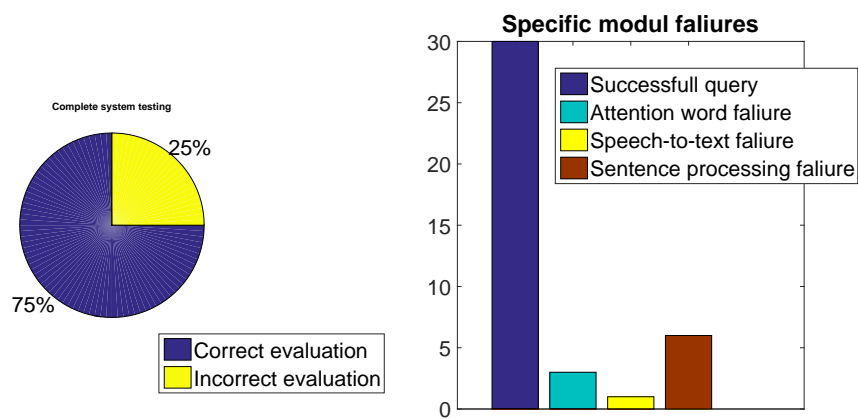
8 Závěrečné zhodnocení projektu

V rámci tohoto projektu byl sestaven plně funkční hlasově ovládaný asistent schopný vyhodnocování několika jednoduchých dotazů, převážně z oblasti předpovědi počasí. Reakce na "Activation word", schopnost přeložit dotaz do textové podoby i "Sentence processing" mají poměrně dobrou úspěšnost, jak je vidět z testů. Logika "Query module" spolehlivě pokrývá základní informace o počasí a dokáže zodpovědět převážnou většinu běžných dotazů na toto téma. Zde představený asistent pochopitelně nemůže konkurovat komplexním komerčním řešením, ovšem úspěšně demonstruje, že s dnešními volně dostupnými technologiemi je možné vybudovat v krátkém čase funkční, byť zjednodušené řešení. Vzhledem k modulární a univerzální struktuře projektu je možné v budoucnu lehce přidávat funkcionalitu z dalších znalostních oblastí, či rozšiřovat komunikační dovednosti samotného asistenta. Demonstrace tohoto rozšíření je demo submodulu "agenda" a demo submodulu "news", první dokáže říci vtip, a aktuální čas, druhý přečte tři vybrané nadpisy článků novin The Guardian.

Závěrem je možno říci, že zadání projektu, stejně jako cíle stanovené v první zprávě byly splněny. V průběhu semestru se objevilo několik problémů, ovšem projekt jimi nebyl příliš zdržen vůči plánu a v daném termínu se podařilo sestavit fungující prototyp, který snad bude v budoucnu rozšířen o další funkce.

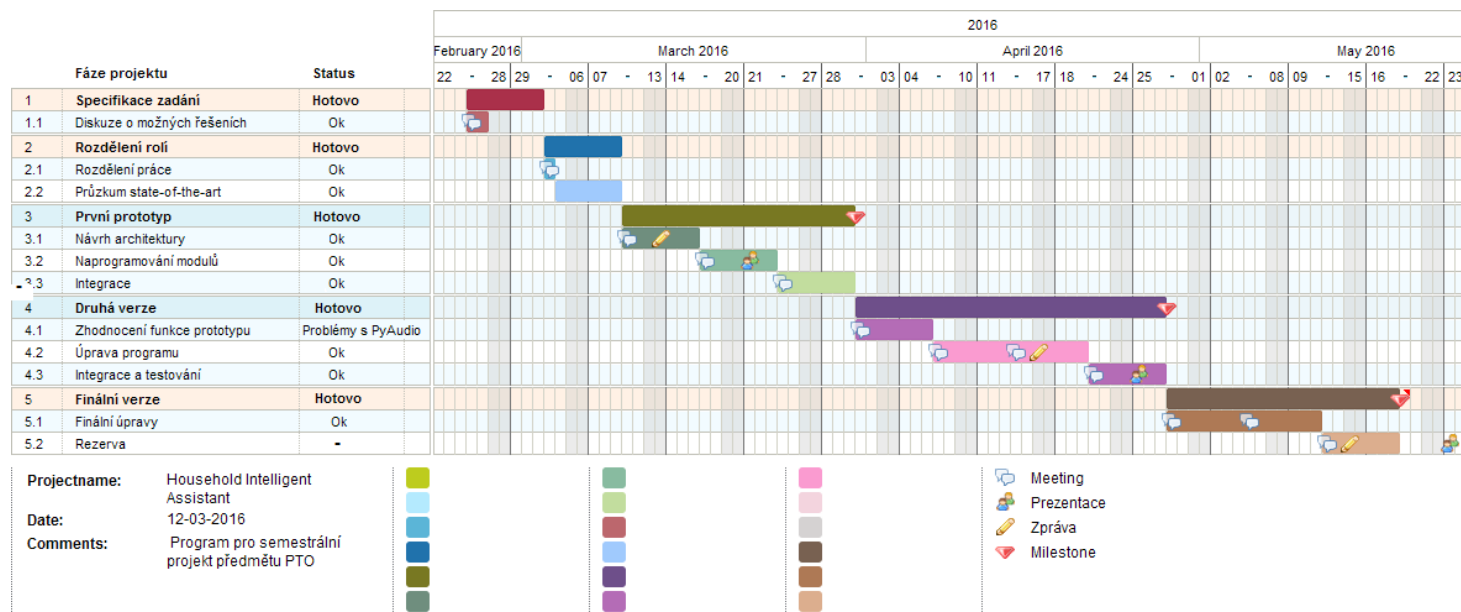


Obrázek 5: Výsledky testování celého systému v klidném prostředí.



Obrázek 6: Výsledky testování celého systému v hlučném prostředí.

11



Obrázek 7: Harmonogram projektu graficky znázorněn formou Ganttova diagramu s klíčovými milníky.