Semestrální práce

Filip Polák

Akademický rok 2020/2021

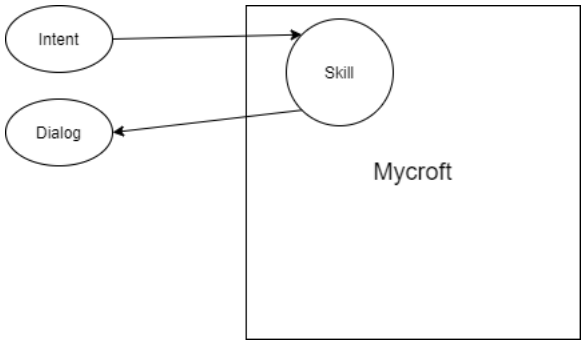
1 Mycroft

Mycroft, pojmenovaný podle počítače ze sci-fi románu *Měsíc je drsná milenka*, je open source hlasový asistent napsaný v jazyce Python společností Mycroft AI, který je možné spustit na počítači či notebooku s operačním systémem Linux, v autě nebo na mobilu s operačním systémem Android. Hlavním tahákem jsou ovšem Mark I a Mark II, které představují alternativu k Google Home či Amazon Echo a slouží k ovládání domácnosti. Ve svých počátcích byl projekt podpořen kampaní na Kickstarteru, kde na něj uživatelé přispěli přes 120 000$. Jedním z hlavních důvodů úspěchu této kampaně byly skandály velkých hráčů v oboru hlasových asistentů se sdílením dat uživatelů třetím stranám, které otevřely cestu menším společnostem, které si zakládají na ochraně uživatelských dat a uživatelích jako takových.

Obrázek 1: Mark II

1.1 Skills

Každý hlasový dialogový systém musí mít definovanou doménu, ve které se bude konverzace provádět. Tato definice je v Mycroftu elegantně provedena pomocí tzv. Skilly (dovedností). Skilly jsou vytvářeny techniky a programátory přímo v Mycroftu AI, ale možnost přispět k rozšíření novými či zlepšení již existujících Skillů mají i samotní uživatelé. Skilly mohou obhospodařovat jednoduché funkce jako je hod kostkou, vyprávění vtipů nebo křišťálová koule (náhodně řekne ano/ne), ale lze díky nim nastavit budík, zjistit situaci na burze nebo pustit hudbu. Skilly tedy zajišťují specializaci na různá odvětví, zpracování a generování řeči obstarává Mycroft jako takový.



Obrázek 2: Zakomponování Skillů do Mycroftu

1.2 Demo Skill

Pro seznámení s prostředím Mycroft byl vytvořen Skill **Poker Hands**, který uživateli vysvětlí, co znamená určitá kombinace karet, jako např. pár, dva páry, postupka atd.

1.2.1 Intent

Každý Skill je rozdělen do dvou částí, kopírující klasickou definici vstupu a výstupu hlasového agenta. Intent (úmysl) představuje, jak název napovídá, vstup uživatele a tedy cíl, se kterým konverzaci s Mycroftem začal. Ve Skillu Poker Hands je Intent jednoduše vytvořen (TODO DODĚLAT VÍCE INTENTŮ) jako:

What is a {combination}

Mycroft Intent rozeznává tak, že zpracuje část *What is a*, projede všechny nainstalované Skilly a snaží se najít ten, který promluvě nejlépe odpovídá, a pokud jich najde více, hledá Skill vyhovující kombinaci *What is a* + {*combination*}.

1.2.2 Dialog

Dialog definuje, jakým způsobem bude hlasový asistent odpovídat. Hlasoví asistenti by měli mít možnost odpovídát vícero způsoby, proto je dobrou praxí Skillu předepsat několik možných variant, aby konverzace zněla více "life-like"(živěji???). (TODO DODĚLAT VÍCE DIALOGŮ)

It is a {result}

Pokud Mycroft neporozumí promluvě (například "pair"(pár) převede na "bear"(medvěd)), řekne uživateli, aby svoji promluvu opakoval ještě jednou.

1.2.3 Tesotování

V Mycroft AI jsou velmi pečliví ve vybírání Skillů, které zveřejní na svém Skills Marketplace (trhu), ze kterého si můžou uživatelé stahovat libovolné Skilly podle své potřeby. V samotném Mycroftu je tedy zakomponován testovací algoritmus, s

pomocí kterého může uživatel jednoduše ověřit funkčnost svého vytvořeného Skillu. Testy mohou být nadefinované v *.json* formátu, kde návrhář nadefinuje, kterému *intent* má určitá *utterance* (promluva) odpovídat a jaký *dialog* by na ní měl reagovat. Může se zde také definovat maximální *timeout*, jestli má Mycroft žádat uživatele o *confirmation* (potvrzení) atd., což bylo z důvodu pouhého seznámení s funkčností Mycroftu vynecháno.
Návrhář má také možnost nadefinovat testy pomocí tzv. *test steps*. V Mycroftu existují čtyři základní steps, které obsáhnou naprostou většinu možných Intents a Dialogs, i tak má ale technik možnost vytvořit si své vlastní steps, pokud mu pomohou se zachycením kýžené funkčnosti Skillu. Prvním ze základních steps je Given (předpokládající), který Mycroftu předkládá skutečnosti, které platí pro určitého uživatele. Pokud se předpokládá, že uživatel hovoří anglicky a žije v Londýně, lze tuto podmínku napsat jako: ***Given*** *an english speaking user* ***And*** *user is located in London*. Dalšími důležitými steps jsou *When* (pokud, když) a *Then* (pak), jež představují jednu dialogovou obrátku a jsou napsané ve stylu: ***When*** *the user says "how hot will it be today",* ***Then*** *"mycroft-weather"should reply with dialog from "current.high.temperature.dialog"*, kde "mycroft-weather"odpovídá Skillu, který má zpracovat "how hot will it be today"pomocí dialogu "current.high.temperature.dialog".
Posledním základním step je *And, But*, který pouze upravuje čitelnost více po sobě jdoucích *Then* steps.
Pro Skill Poker Hands by tedy měl jeden test vypadat jako:
Feature: poker-hands
Scenario: flush
Given an english speaking user
When the user says what is a flush
Then poker-hands should reply with dialog from poker.hands.dialog

1.3 Jak Mycroft funguje

1.3.1 Wake Word

Wake Word ("probouzecí"slovo) je fráze používaná u všech hlasových asistentů na světě. Jednoduché vysvětlení je takové, že hlasoví asistenti pořád poslouchají a průběžně zprácovávají zvuky, které zachytili, ale žádné z těchto zvuků si neukládají nebo s nimi dále nepracují. Skutečná konverzace začíná až potom, co uživatel pronese *Wake Word*. "Hey Siri"nebo "Hey Google"jsou těmi nejznámějšími, v Mycroftu si může uživatel vybrat z několika variant, nejpoužívanější ovšem je pro zachování jednotnosti "Hey Mycroft".

1.4 ASR

Mycroft se datuje do roku 2016, kdy nebyly ještě hluboké neuronové sítě rozvinuté jako dnes, proto byly pro automatické rozpoznávání řeči použité skryté Markovovy modely, které i v Mycroftu v následujících letech nahradily neuronové sítě. ASR je v Mycroftu zajištěna **Deep Speech** speech-to-text enginem, který je založen na **Deep**

Speech: Scaling up end-to-end speech recognition paperu (TODO NAPSAT NĚCO O DEEP SPEECH)

1.5 SLU

SLU je v Mycroftu obstaráno **Adaptem** a **Padatiusem** (TODO NAPSAT O OBOU NĚCO). Pro udržení konvencí zavedené Mycroftem je SLU nazýváno jako **Intent parsing**.....

1.6 TTS

Pro převod textu do řeči a generování hlasu využívá Mycroft **Mimic**, který je založen na **Festival Lite** systému syntézi řeči.