



„SELIDOR – T. Puza, Ł. Wasilewski”
Spółka Jawna
ul. Łomiańska 20 B
01-685 Warszawa

RAPORT KOŃCOWY Z WYKONANYCH BADAŃ
z dnia 31.03.2022 r.

Raport powstał w wyniku zakończenia prac badawczych w związku z realizacją projektu projektu numer RPMA.01.02.00-14-b405/18-00 pt. „Prace badawczo-rozwojowe nad językiem naturalnym mówionym w spółce Selidor” współfinansowanego z Europejskiego Funduszu Rozwoju Regionalnego w ramach Osi Priorytetowej I „Wykorzystanie działalności badawczo-rozwojowej w gospodarce”, Działanie 1.2 „Działalność badawczo-rozwojowa przedsiębiorstw”, typ projektów „Projekty badawczo – rozwojowe” Regionalnego Programu Operacyjnego Województwa Mazowieckiego na lata 2014-2020.

Dnia 31.03.2022 zakończyły się prace badawcze w wyniku których opracowany został zaawansowany informatyczny system rozpoznawania i rozumienia języka polskiego używanego w rezerwacji terminów, powiązany z internetowym systemem rezerwacji wizyt.

Prace badawcze rozpoczęły się w 1.01.2019 i trwały do 31.12.2021 r. Prace badawcze w zakresie zadania 1 oraz 2-5 prowadzone były przez zespół pracowników badawczych „SELIDOR - T. Puza, Ł. Wasilewski” Spółka Jawna. Natomiast prace badawcze dotyczące zadania 2 zostały wykonane przez Polsko-Japońską Akademię Technik Komputerowych.

Prace badawcze w projekcie podzielone były na 5 zadań:

1. Nazwa etapu: BUDOWA KORPUSU SYSTEMU ROZPOZNAWANIA MOWY

Opis badań i wyniki badań:

W ramach zadania zrealizowano następujące prace:

- Opracowanie metodologii zbierania danych - opracowanie sposobu pozyskiwania danych, kryteriów poprawności wypowiedzi, grup docelowych
- Opracowanie założeń wypowiedzi – czasu trwania dialogów, sposobów jego prowadzenia, liczby wypowiedzi, strategii dochodzenia do końcowej rezerwacji, itp.
- Wyselekcjonowanie i dostosowanie narzędzi wspomagających budowę korpusu (nagrywanie, gromadzenie, annotowanie korpusu)
- Przygotowanie procedur (format danych, sposoby konwersji, metody unikania zakłóceń) i technicznych możliwości gromadzenia i obrabiania dużych porcji informacji dźwiękowych (przestrzeń dyskowa, serwery przetwarzające i zwirtualizowane środowisko, przepływ danych, kopie zapasowe)

Z uwagi na potrzebę zasymulowania dojrzałej pracy systemu uznano, że na początkowym etapie projektu dane zbierane będą w modelu „wizard-of-oz” gdzie człowiek emuluje działanie systemu w sposób niewidoczny dla osoby zewnętrznej, która jest przeświadczona, że rozmawia z maszyną. Na tej podstawie rozpoczęto zbieranie korpusu, oraz jego transliterację.

- zebrano próbki głosowe osób próbujących dokonać rezerwacji (różne kanały dźwiękowe)



- f) przeprowadzono zaawansowaną transliterację próbek
- g) dokonano filtrowania i dodatkowych annotacji dla potrzeb ENIAMA
- h) przeprowadzono wiele badań szczegółowych dotyczących sposobu realizowania dialogu na różnych etapach rozmowy
- i) przetestowano z użytkownikami schematy dialogu tworzone przy konstruowania menedżera dialogu

W toku prac zostały wyróżnione i zaadresowane następujące kwestie:

- sposób rozpoczęcia pracy systemu (pytania zamknięte – dilemma czy ma nastąpić przejęcie inicjatywy przez system na początkowym etapie czy należy pozostawić tę inicjatywę użytkownikowi)
- sposób tworzenia konstrukcji gramatycznych, np. z użyciem przyimków i zaimków (które mają tendencje do zlewania się w jedno słowo z następującymi potem częściami zadania
- - pozajęzykowe nawyki werbalnych, np. przerywniki robione przez rozmówców w procesie wypowiedzi – jako pauzy lub dźwięki typu „hmm”, „eee”, itp.
- użycie liczebników jako potencjalnego obszaru poważnych problemów, gdzie formy poszczególnych liczebników mogą być dosyć zbliżone do siebie, a jednocześnie nie dysponujemy żadnym językowym mechanizmem weryfikacyjnym
- problem wyboru usługi na przykładzie listy usług z branży beauty (salony urody, fryzjerzy, makijażystki, stylistki paznokci, barberzy, masażyści, itp.), gdzie jest ponad 1000 usług, często bardzo skomplikowanych, podzielonych na wiele kategorii
- cechy opisujące intencje użytkownika na etapie wyboru usługi (sposób rezerwowania często wykracza poza użycie nazwy usługi, na rzecz opisania potrzeby za pomocą atrybutów usługi – np. „chciałbym się ostrzyc” zamiast „wybieram strzyżenie męskie” lub „proszę zrobić coś z tymi odrostami” zamiast prośby czy o to o farbowanie czy o skrócenie włosów)
- problem opisu czasu (Czas może być używany w wielu znaczeniach: jako punkt czasowy, w którym wizyta powinna się zacząć (oddzielnie lub łącznie takie parametry jak dzień tygodnia/godzina), wersja o szerszym zakresie - pożądana zakres czasowy (poszerzają się wcześniej używane parametry o wyrażenia „koniec tygodnia/początek tygodnia”, zakres godzin lub pora dnia). Dodatkowo czas jest używany w kontekście długości trwania całej wizyty.

Na tej podstawie wyselekcjonowano ze zgromadzonego korpusu wszystkie wyrażenia czasu, oraz rozpoczęto opis znaczeniowy poszczególnych zwrotów. W przypadku napotkania wieloznaczności prowadzono dalsze prace przy zbieraniu korpusu w celu uzupełnienia materiału źródłowego.

- potrzeba dostarczania rozmówcy informacji o aktualnym stanie rezerwacji i potwierdzenia jego wybórów
- konieczność opracowania mechanizmów sterowania rozmową – czyli pozostawienie swobody użytkownikowi, jednak dbałość by dialog nie trwał zbyt długo i zmierzał do uzupełnienia wszystkich braków danych niezbędnych do rezerwacji
- konieczność opracowania mechanizmów wykrywania wypowiedzi „nie na temat”, a więc spoza ontologii języka rezerwacji
- opracowanie mechanizmów rozumienia wypowiedzi w kontekście poprzednich wypowiedzi – klasyczne parsery językowe rozpoznają jedno zdanie, tymczasem zaobserwowane przez nas zjawisko wskazuje, w przypadku prowadzenia dłuższego dialogu osoby często odwołują się do wcześniejszych wypowiedzi (np. dokończenie myśli lub wypowiedzi w formie równoważników zdań zrozumiałe tylko w zestawieniu z poprzednimi zdaniami).



Kamienie milowe

Stworzono korpus systemu rozpoznawania mowy z ponad 75 godzinami nagrani rozmów.

Rezultat badań

Materiał źródłowy do zasilenia pozostałych zadań (materiał dźwiękowy, materiał tekstowy, dokumentacja wewnętrzna dot. sposobów postępowania w pozostałych zadaniach)

Nowa wiedza

- Wiedza dot. korzystania z modelu swobody wypowiedzi użytkownika vs modelu ścisłego kierowania dialogiem
- Wiedza na temat kierowania dialogiem w interakcjach człowiek-maszyna
- Określono sposoby użycia przez użytkownika podstawowych parametrów rezerwacji (wybór usługi/ wybór terminu/decyzja co do wyboru pracownika)
- Określono sposoby uzyskiwania potwierdzeń od użytkownika

Osiągnięty poziom gotowości technologicznej: poziom III

Osiągnięto założone cele i wykonano całość zaplanowanych prac badawczych.

2. Nazwa etapu: UTWORZENIE PODSTAWOWEGO MODELU ROZPOZNAWANIA MOWY

Opis badań i wyniki badań:

W ramach projektu przygotowano zestaw modeli i narzędzi niezbędnych do rozpoczęcia pracy nad projektem:

- przygotowano model akustyczny dostosowany do rozpoznawania mowy na urządzeniach mobilnych i typu desktop
- przygotowano wstępne modele języka
- przygotowano narzędzie do konwersji zapisu ortograficznego na fonetyczny w celu aktualizacji słownictwa
- przygotowano szereg narzędzi i skryptów pozwalających na aktualizację w/w słownictwa i modeli języka
- przygotowano demonstracje wykorzystania wymienionych narzędzi w celu rozpoznawania mowy w czasie rzeczywistym
- dokonano szeregu spotkań i szkoleń z obsługi w/w narzędzi i procedur

W ramach spotkań rozpoczęto też prace nad aktualizacją modelu akustycznego w celu zwiększenia jego skuteczności w rzeczywistych warunkach produkcyjnych. Proces ten będzie się realizowany na dwa sposoby - poprzez augmentację istniejących zbiorów akustycznych oraz poprzez nagrywanie kolejnych danych w warunkach zbliżonych do systemu produkcyjnego.

Dodatkowo ustalono też wymagania do aktualizacji warstwy językowej poprzez augmentację zbiorów tekstowych o słownictwo występujące w zadaniach rozwiązywanych przez finalny produkt.

- stworzono algorytmy do augmentacji danych językowych o kolejne dziedziny występujące w dialogach związanych z rezerwacją. Zdania podlegające augmentacji opracowano na podstawie przykładowych dialogów zbieranych w trakcie testów platformy. Na podstawie transkrypcji rzeczywistych dialogów opracowano wzorcowe zdania i oznaczono fragmenty, które się nadają do modyfikacji.



- zaprojektowane i stworzone zostały skrypty w języku Python, które na podstawie zdefiniowanych reguł, list wyrazów oraz słowników różnych form gramatycznych wyrazów generują listy pożądanych fraz. Stworzony został program losujący określona ilość fraz z danej puli, a także program wyłuskujący dane z rejestru Teryt (etryt.stat.gov.pl) tj. miasta i ich dzielnice. W wyniku przeprowadzonych prac wygenerowane zostały pliki zawierające następujące ilości fraz:

Location big: 133 tys. fraz na podstawie 32 reguł

Location small: 5 mln fraz na podstawie 32 reguł

Time: 35 tys. fraz

Date: 7,5 tys. fraz

Money: 115 tys. fraz

Service: 2,7 mln fraz usług pojedynczych na podstawie 50 reguł oraz 7 mln fraz usług złożonych na podstawie 12 reguł

Utworzone oprogramowanie pozwala na przekazywanie przez użytkownika dowolnej ilości reguł skonstruowanych na podstawie opracowanych zasad (umieszczonej w oddzielnym pliku), które umożliwiają generowanie fraz w ilości zależnej od zasobności wykorzystywanych słowników, więc zbiory fraz mogą być w przyszłości powiększane.

Wygenerowane dane zostały wykorzystane w procesie trenowania modeli języka polskiego. Dotychczasowe eksperymenty pokazały, że dane temporalne poprawiły współczynnik WER z poziomu 11,24% do poziomu 3,47% (w stosunku do wykorzystywanego wcześniej korpusu zdań związanych z datą i czasem). Poprawie uległy również dwa korpusy związane z liczebnikami (o ok. 15%).

Kamienie milowe

Stworzono model (oprogramowanie, słowniki, modele statystyczne) dostosowany specjalnie do dialogów zbliżonych do specyfiki projektu rezerwacji.

Rezultat badań

Oprogramowanie i materiały niezbędne do zasilenia zadania 3

Nowa wiedza

Wyabstrahowano podstawowy model korzystania z języka polskiego, nadający się do dostosowania w zamkniętych, specjalizowanych dziedzinach użycia języka.

Osiągnięty poziom gotowości technologicznej: poziom IV

Osiągnięto założone cele i wykonano całość zaplanowanych prac badawczych.

3. Nazwa etapu: UTWORZENIE SYSTEMU ROZPOZNAWANIA MOWY

Opis badań i wyniki badań:

Przyjęto strategię rozwoju systemu ASR, która założyła początkowo skupienie się na modelowaniu gramatyki i uzupełnieniu słownika o frazy związane z dziedziną projektu, zaś w drugiej fazie skupienie się na dopasowaniu modelu akustycznego do spodziewanych warunków jakościowych sygnału audio od klientów.



- Pracę rozpoczęto od stworzenia korpusu nagrani związanego z dziedziną projektu. Następnie opracowano prototyp aplikacji ASR korzystając z bibliotek Kaldi oraz ogólnego modelu języka polskiego.
- Opracowano narzędzie testowe umożliwiające obliczenie pokrycia słownika korpusu testowego przez słownik modelu języka, oraz skrypty obliczające jakość modelu (WER) oraz jakość krat (oracle WER).
- Następnie stopniowo rozpoczęto rozbudowywanie modelu językowego o zbiory fraz związanych z dziedziną projektu i testowanie powstałych modeli. Model języka został utworzony na podstawie istniejącego ogólnego modelu gramatyki jako bazy i jego ważne połączenie z modelem 3-gramowym, utworzonym z fraz związanych z dziedziną projektu. Jakość rozpoznawania mowy takiego projektu nieco się poprawiła, natomiast jakość generowanych krat poprawiła się znacznie.
- Wykonano szereg prototypowych modeli akustycznych korzystając ze sztucznie zakłóconych nagrani treningowych. Korpus treningowy został wagowo wymieszany ze zbiorem zakłóceń izotropowych oraz został dodany pogłos. Celem było dobranie parametrów tych zakłóceń umożliwiających otrzymanie nagrani najwierniej przypominających nagrane w realnych warunkach, tak aby osiągnąć możliwie mały WER. Wykorzystano publicznie dostępną bazę zakłóceń i charakterystyk odpowiedzi impulsowych pomieszczeń z zasobów OpenSLR.
- Opracowano mowy model gramatyczny języka polskiego. Wykorzystano w dużej mierze korpusy dziedzinowe zebrane w ramach naszego projektu, oraz wygenerowane frazy związane z dziedziną projektu. Model został również wzbogacony o zbiory fraz obejmujące nazwy wszystkich miejscowości gminnych w Polsce. Takie działania poprawiły działanie systemu ASR w zakresie języku wykorzystywanego w procesie rezerwacji usług.
- Wykonano zestaw eksperymentalnych narzędzi do diaryzacji i segmentacji nagrani dialogowych. Narzędzia wspomagają proces przygotowywania poprawnego alignmennu i rozróżniania mówców dla nagrani.
- Zmieniony został rodzaj modelu sieci neuronowej, z dotychczas stosowanego TDNN na TDNNF. Eksperymenty wykazały ponad 2% poprawy WER w porównaniu do modelu wytrenowanego na tych samych danych na sieci TDNN.
- Zmieniony został rodzaj modelu sieci neuronowej, z dotychczas stosowanego TDNN na TDNNF. Eksperymenty wykazały ponad 2% poprawy WER w porównaniu do modelu wytrenowanego na tych samych danych na sieci TDNN.
- Rozszerzenie zbioru korpusów treningowych o nagrania CLARIN Phone i nagrania własne Selidor pozwoliło obniżyć stopę błędów WER z poziomu ok. 10% do poziomu poniżej 6% dla danych dziedzinowych. Wykonano szereg eksperymentów z różnymi poziomami zakłóceń danych treningowych.
- Wykonano eksperymenty nad wykorzystaniem neuronowego modelu języka RNNLM zamiast modelu n-gramowego. Nie udało się poprawić stopy błędów WER a nawet testy wskazywały lekkie pogorszenie rezultatów. Wniosek jest taki, że dane tekstowe zawierały zbyt dużo fraz bezkontekstowych i nie stanowi to dobrego materiały treningowego dla sieci RNNLM.
- Wykorzystano transkrypcje nagrani Selidor w celu rozbudowy modelu języka.
- Przygotowanie modułu detektora mowy (SAD - speech activity detection) dla aplikacji ASR. Moduł zaimplementowano z wykorzystaniem sieci TDNN. Wykorzystano wytrenowany model KALDI ASPIRE SAD Model 1a. Moduł SAD ma za zadanie odciążyć główny potok przetwarzania aplikacji ASR w czasie gdy na wejście podawane są ramki dźwiękowe nie zawierające mowy ludzkiej. Pozwoliło to zredukować obciążenie procesora z 30% do 3%.



- Przeprowadzono szereg eksperymentów mających na celu poprawienie stopy błędów modelu akustycznego. Wprowadzono dalsze zakłócenia izotropowe i punktowe oraz pogłosowe do wykorzystywanych zbiorów danych, tak aby zwiększyć liczbę nagrań podczas treningu sieci neuronowej. Modyfikując rozmiarem sieci TDNNF dopasowano go do używanego obecnie zbioru danych. Eksperymentalnie dobrano optymalne parametry uczenia sieci. Obecnie stopa błędów sięga 4,5% podczas testów na korpusach, co w pełni odpowiada założeniom badawczym i jest jednym z najlepszych wyników w tego systemach w kraju.

Kamienie milowe

Stworzono system przetwarzający w czasie rzeczywistym wypowiedzi mówione w języku polskim o wysokim stopniu trafności transliteracji.

Rezultat badań

Zrealizowano oprogramowanie dokonujące w czasie rzeczywistym transliteracji języka polskiego używanego w rezerwacjach

Nowa wiedza

Opracowanie sposobu przetwarzania w czasie rzeczywistym dziedzinowego języka z szerokim słownictwem i konstruktami gramatycznymi języka polskiego, o bardzo niskim poziomie błędów, nadającym się do zastosowań komercyjnych.

Osiągnięty poziom gotowości technologicznej: poziom V

Osiągnięto założone cele i wykonano całość zaplanowanych prac badawczych.

4. Nazwa etapu: ANALIZA ROZUMIENIA JĘZYKA MÓWIONEGO

Opis badań i wyniki badań:

Prace obejmowały:

- Przyjęto koncepcję przetwarzania języka mówionego w oparciu o koncept krat – tekstowej reprezentacji (macierzy) możliwych interpretacji wypowiedzi głosowej. Z uwagi na to, że prawdopodobieństwo wygenerowania całkowicie poprawnej interpretacji przez moduł rozpoznawania mowy jest niewielkie, przyjęto założenie, że danymi wejściowymi dla modułu rozumienia języka jest krata, czyli zbiór możliwych interpretacji przetwarzanej wypowiedzi. Prawdopodobieństwo tego, że jedna tych interpretacji jest poprawna, jest wysokie (w eksperymantach udało się uzyskać wartość 97%).
- dostosowanie parsera ENIAM do przetwarzania danych wejściowych zapisanych w postaci krat
- wytworzenie leksykonów zawierających słownictwo wykorzystywane w wypowiedziach poświęconych wyborowi usług,
- opracowanie gramatyki języka występującego w wypowiedziach poświęconych wyborowi usług,
- opracowanie reguł generujących reprezentację semantyczną dla wypowiedzi poświęconych wyborowi usług,
- implementację narzędzia odmieniającego frazy przez przypadki (na potrzeby modelu języka w module rozpoznawania mowy).



- rozwinięte zostały leksykony zawierające słownictwo dziedzinowe, gramatyka oraz reguły generujące reprezentację semantyczną. Pozwoliło to sparsować nieprzetwarzane wcześniej konstrukcje składniowe takie, jak pytania. Jak również pozwoliło rozszerzyć zakres tematyczny generowanych reprezentacji semantycznych o informacje dotyczące czasu i miejsca wykonania usługi, wykonawcy oraz komendy sterujące menedżerem dialogu
- wprowadzono wykrywanie autora czynności (roznienie między konstrukcjami „ja chcę zarezerwować” – „dla kogo zarezerwować” – „z czich usług skorzystać”)
- wsparcie dla wyboru usługi oferowanych przez system rezerwacyjny: klient wyrażając swoje potrzeby niekoniecznie korzysta ze słów, które bezpośrednio i jednoznacznie wskazują usługę z listy. Np. do wypowiedzi „Chciałbym umówić się do fryzjera” pasują wszystkie usługi dotyczące „strzyżenia męskiego”. W ramach prac wytworzony został algorytm, który na podstawie wypowiedzi klienta wskazuje usługi, które mógł on mieć na myśli. Algorytm ten przyjmuje na wejściu reprezentacje semantyczne wygenerowane przez parser. Następnie korzystając ze specjalnie w do tego celu opracowanego słownika odnajduje usługi, które pasują do informacji zawartej w dostarczonej mu reprezentacji semantycznej.
- algorytm przypisujący usługi z listy ich reprezentacjom semantycznym wytworzonym na podstawie wypowiedzi klientów został uzupełniony o analizowanie przypadków, w których klient podaje parametr nieistotny dla danej usługi, np. klient chce zrobić sobie tatuaż na ręce, podczas gdy usługi dotyczące tatuaży nie są parametryzowane przez części ciała.
- opracowany został algorytm przeprowadzający wstępna klasyfikację wypowiedzi klienta na podstawie poprzedzającej ją wypowiedzi konsultanta. Klasyfikacja ta znacznie przyspieszyła analizę wypowiedzi klientów. I tak z dialogów zostały wydobyte frazy temporalne, frazy potwierdzające, frazy pytajne, frazy dot. wykonawcy, frazy dot. wyboru usług i nazwy usług, nazwiska, emaile i numery telefonów. Na ich podstawie została przeprowadzona analiza użycia w domenie rezerwacji zaimków pytajnych oraz słów: jakiś, inny, ten; sklasyfikowane zostały pytania wydobyte z wypowiedzi klienta oraz zawarte w pytaniach frazy czasownikowe
- rozszerzono leksykony w parserze (leksykony akcji, leksykony zawierające słownictwo wykorzystywane w nazwach usług itp.), zaimplementowano semantykę pytań oraz parsowanie wypowiedzi dotyczących zmiany i anulowania rezerwacji. Na podstawie rozszerzonych leksykonów uaktualnione zostało powiązanie poszczególnych leksemów z listą usług występujących w branży beauty. Przygotowane zostały również zbior sparsowanych do postaci reprezentacji semantycznej wypowiedzi klienta i frazy wydobytych z tych wypowiedzi.
- w toku prac okazało się, że klient zwykle opisuje usługę, którą chce zarezerwować w sposób nieprecyzyjny. Np. mówi „Chciałbym się wydepilować”. Z uwagi na to, że w systemie jest ponad 250 usług dotyczących depilacji, pojawiła się konieczność zadawania przez agenta pytań doszczególowiających. Opracowana w tym celu została hierarchia usług pozwalająca określić usługę, którą klient ma na myśli za pomocą zadawania pytań zamkniętych np. „Czy interesuje Cię wosk, lycon, pasta cukrowa, laser czy fotodepilacja?”. Zaimplementowany został również dedykowany serwis rozpoznający, którą z kategorii zawartych w pytaniu klient wskazał w odpowiedzi.
- Na podstawie zbioru wydobytych z korpusu wyrażeń temporalnych opracowana została gramatyka, która przetwarza je do postaci reprezentacji semantycznej. Gramatyka ta rozpoznaje godziny, minuty, dni miesiąca, dni tygodnia, miesiące i lata. Reprezentacja semantyczna następnie poddawana jest uziemieniu (ang grounding). Jest to procedura mająca na celu wyliczenie konkretnych przedziałów czasu, o których klient mówi. Oprócz



reprezentacji semantycznej wypowiedzi klienta otrzymuje ona na wejściu informację o aktualnej dacie i czasie oraz terminie zaproponowanym przez manangera (jeśli takowy istnieje). Uziemienie pozwala wskazać konkretne terminy dla względnych określeń czasu oraz ujednoznaczyć terminy podane nieprecyzyjnie.

- Opracowano reprezentację semantyczną kwantyfikatorów oraz fraz ze słowem „ten” i ze słowem „termin”. Uporządkowano reguły semantyczne.
- Opracowano też uziemienie dla wyrażeń zawierających słowa „dowolny”, „innego”, „tylko” oraz „nieco”.
- Rozwinięto groundera czasu tak by interpretował wypowiedzi zawierające preferencje klienta (frazy typu „jak najwcześniej”, „jak najpóźniejsza godzina”).
- Opracowano serwis parsujący numery telefonów oraz serwis rozpoznający imiona i nazwiska.

Kamienie milowe

Stworzono narzędzie analizy języka mówionego przy wykorzystaniu parsera ENIAM.

Rezultat badań

Zrealizowano oprogramowanie przetwarzające konstrukcje językowe na zrozumiałe dla oprogramowania parametry opisujące poszczególne fragmenty dialogu rezerwacyjnego

Nowa wiedza

- Rozszerzenie wiedzy o sposobie parsowania semantyczno-językowego w przypadku języka polskiego mówionego
- Opracowanie sposobu wyboru użytkownika z rozbudowanej bazy produktowej (wybór usługi)
- Opracowanie sposobów wyboru terminu w odniesieniu do kalendarza na poszczególnych etapach wyboru (zawężanie/rozszerzanie terminów)

Osiągnięty poziom gotowości technologicznej: poziom V

Osiągnięto założone cele i wykonano całość zaplanowanych prac badawczych.

5. Nazwa etapu: OPRACOWANIE MANAGERA DIALOGU

Opis badań i wyniki badań:

System sterowania dialogiem:

Zaprojektowano algorytm efektywnego dialogu, czyli efektywnego wypełnienia wszystkich danych niezbędnych do rezerwacji, uzyskiwanych w procesie rozmowy z użytkownikiem:

- Określenie potrzeby użytkownika (potwierdzenie chęci rezerwacji)
- Pytanie o usługę (na podstawie pełnej nazwy lub tylko jej części, atrybutów usługi, osoby wykonującej lub specyficznego słownictwa używanego przez branżę beauty) – na podstawie wstępnej wypowiedzi użytkownika określa najbardziej prawdopodobne trafienia.
- Pytanie o pracownika – opcjonalne (dla firmy pożądane jest uniknięcie wyboru pracownika przez klienta, z drugiej strony wybór pracownika dla niektórych klientów może być ważną częścią procesu rezerwacji; system posiada taką funkcjonalność, jednak



proces dialogu minimalizuje tę ewentualność sugerując płynnie przejście do dalszej części procesu rezerwacji)

Podsystem formułowania pytań

- system potwierdza wszystkie wykonywane czynności, stara się nie robić tego jednak w sposób natarczywy
- system korzysta z predefiniowanej biblioteki zapytań, wykorzystywanych w miarę postępu dialogu

Podsystem prezentowania alternatyw

- w związku z ograniczeniem ludzkiego postrzegania do 7+/-2 elementów system musi ograniczać liczbę prezentowanych alternatyw do stosunkowo małej liczby. Gdy porównamy to z przeciętnym zbiorem realnych alternatyw (to zależy od zapytania, ale sformułowanie „termin w przyszłym tygodniu” może wygenerować setki lub tysiące alternatyw), wymusiło to na systemie zastosowanie różnych strategii dostarczenia realnych możliwości wyboru – w przypadku usług celem systemu jest podróż po drzewie usług aż do dotarcia do punktu gdy liczba alternatyw umożliwia już ich zaprezentowanie; w przypadku pracowników przyjmuje strategię proponowania kilku pracowników z możliwością przedłużenia wyliczenia; w przypadku terminów system zawsze stara się przedstawić jakąś propozycję, zawsze pozostawiając jako opcję możliwość dalszego doprecyzowania terminu przez użytkownika

Podsystem analizy emocji

- prace w projekcie skupiły się nad jak optymalizacją procesu rezerwacji, by ten nie zdążył jeszcze wywołać irytacji i frustracji – co zostało uznane za lepszą strategię niż próby ratowania sytuacji w przypadku gdy te nerwy już się pojawiły

Przeprowadzono następujące prace:

- moduł do konfigurowania reguł w notacji YAML
- możliwość uzależnienia zastosowania reguł od pól znaczeniowych wykrytych przez moduł intrepretacji języka naturalnego (ENIAM)
- właściwą interpretację liczb rozpoznanych przez ENIAM
- sterowanie wyborami klienta przy pomocy liczb
- uwzględnianie kontekstu pytania do klienta w rozpoznawaniu wyboru klienta
- interfejs do serwera bazy danych rezerwacji
- interfejs do modułu ASR, przyjmujący tekst i kraty
- interfejs do modułu generacji mowy (TTS)
- interfejs do modułu ENIAM z przesyłaniem krat od ASR
- interfejs do groundera usług
- listę usług globalnych
- obsługę wielu klientów jednocześnie
- zaprojektowanie interfejsów do odbioru krat od ASR i przekazywania do ENIAM
- Zaprojektowanie interfejsu do generacji mowy (TTS, Text-to-Speech)
- Implementacja interfejsów kratowych, z możliwością dalszego korzystania z interfejsów tekstowych
- Implementacja komunikacji z TTS
- Implementacja interfejsu kratowo-tekstowego - przyjmowanie krat, korzystanie z najlepszej ścieżki tekstowej do dalszego przetwarzania



- Zaprojektowanie nawigacji w hierarchii usług
- Implementacja wyboru usługi z wykorzystaniem hierarchii
- Zaprojektowanie interfejsu do groundera kategorii
- Implementacja głosowego wyboru usług i kategorii przy wykorzystaniu groundera kategorii
- Optymalizacja struktury kategorii usług.
- Optymalizacja procesu wyboru usługi wg kategorii - wybór tylko w ramach wspólnej kategorii nadzędnej.
- Wybór usługi wg nazwy skróconej.
- Kontynuacja implementacji botflow_1.01: powrót po 3 nieudanej próbie rozpoznania.
- Zaprojektowanie interfejsu do danych od VAD.
- Zaprojektowanie interfejsu do groundera czasu.
- Implementacja analizy przedziałów czasowych przez grounder czasu.
- Opracowanie protokołu komunikacji z grounderem kategorii
- Implementacja komunikacji agenta dialogowego z grounderem kategorii
- Opracowanie protokołu komunikacji z grounderem czasu
- Implementacja komunikacji agenta dialogowego z grounderem czasu
- Opracowanie protokołu komunikacji z parserem numerów telefonu
- Implementacja komunikacji agenta dialogowego z parserem numerów telefonu
- Opracowanie protokołu komunikacji z parserem imion i nazwisk
- Implementacja komunikacji agenta dialogowego z parserem imion i nazwisk
- Implementacja strategii prezentacji terminów do wyboru
- Przygotowanie do odbioru identyfikatora firmy od GUI
- Implementacja komunikacji TCP z parserami i grounderami
- Wykrywanie niespójności w danych konfiguracji usług

Kamienie milowe

Stworzony system, który jest w stanie doprowadzić użytkownika do rezerwacji usługi, przy użyciu języka naturalnego.

Rezultat badań

Zrealizowano oprogramowanie łączące wiedzę o sposobie prowadzenia dialogu przez ludzi z zadania 1, transliterację języka z zadania 3, przetworzenie sformułowań na parametry z zadania 4 – w połączonej formie oprogramowania, które inicjuje rozmowę z człowiekiem, przeprowadza go przez proces rezerwacji i dokonuje zmian w systemie rezerwacyjnym reservis.

Nowa wiedza

- Mechanizmy uwzględniające kontekstowy sposób prowadzenia dialogu, gdzie zarówno wypowiedzi komputera, jak i człowieka mogą się odnosić do wcześniejszych informacji podanych przez obie strony
- Konieczność podziału dialogu na szereg pomniejszych podfunkcji związanych z różnymi aspektami dialogu rezerwacji (poruszanie się po drzewie wyboru, wybór elementów z listy, akceptacja sugestii komputera)
- Radzenie sobie z ograniczeniami kanału komunikacji (człowiek jest w stanie przetrzymywać w pamięci 7+-2 elementy, a elementów zwrotnych często jest więcej)
- Specyficzne użycie danych z systemu rezerwacyjnego w celu uzyskiwania odpowiedzi na wątpliwości użytkownika, które może konstruować w języku mówionym (np. system rezerwacyjny wcześniej nie przewidywał szukania możliwych terminów w przyszłym tygodniu tylko w godzinach między 16 a 19).



Osiągnięty poziom gotowości technologicznej: poziom VI

Osiągnięto założone cele i wykonano całość zaplanowanych prac badawczych.

PODSUMOWANIE:

- ✓ Zaplanowane prace badawcze zostały w pełni zrealizowane, a cel projektu został osiągnięty.

W zadaniu 1 w ramach prac rozwojowych wszystkie podsystemy sprzęgnięto i przetestowano, rozwiązując standardowo drobne problemy związane z wydajnością, desynchronizacją części formatów danych, itp. System osadzono na nowoczesnej platformie sprzętowo-systemowej wykorzystującej technologię dockerów, umożliwiających skalowalność oraz wysoką niezawodność.

Osiągnięto założone cele i wykonano całość zaplanowanych prac.

Część II: Rezultat przeprowadzonych prac badawczo-rozwojowych

W wyniku przeprowadzonych prac badawczych, na podstawie pozyskanej wiedzy, firma Selidor Sp.J. opracowała innowacyjny produkt: system rezerwacji wizyt za pomocą mowy w języku polskim.

System jest zintegrowany z oknem dodawania rezerwacji przez użytkowników końcowych (Recepja), w którym można podjąć decyzję o przeprowadzaniu całego procesu rezerwacji klasycznie, za pomocą GUI, albo tylko przy pomocy mowy po kliknięciu w przycisk mikrofonu.

Po jego naciśnięciu zostaje uruchomiony menedżer dialogu, który rozpoczyna kierowanie rozmową. Cały proces jestkończony po uzyskaniu wszystkich niezbędnych parametrów i dodaniu faktycznego terminu wizyty w harmonogramie danej firmy/danego pracownika.

Produkt umożliwia przeprowadzenie pełnoprawnej rezerwacji usług przez klienta za pomocą urządzenia wyposażonego w mikrofon (komputera stacjonarnego/ notebooka/ tabletu/ smartphone'a) i umówienie się na określony termin i określona usługę w firmie korzystającej z reservisu.

Produkt będzie oferowany jako rozszerzenie podstawowego abonamentu, dostępne za dodatkową miesięczną opłatą. Informacje o produkcie publikowane są na stronie www i cenniku firmy (załączamy zrzut ekranowy, oryginalna strona dostępna pod adresem oferta.reservis.pl/beauty/cennik)

Warszawa, 31.03.2022

SELIDOR
SP. Z O.O.
UL. KOMIANSKA 20B, 01-685 WARSZAWA
TEL. (40 22) 8331492, FAX (40 22) 8331490

abonament

Wybierz pakiet dla siebie - bez umowy, bez zobowiązań.

bb: Start

oferta na początek

25 zł netto/mies.

(30,75 brutto)

przy płatności za rok z góry;
35 zł netto przy płatności miesięcznej

2

liczba kalendarzy w abonamencie
bez możliwości rozszerzenia

100

limit rezerwacji w miesiącu



brak SMSów abonamencie
możliwość dokupienia SMSów w pakietach
(0,12 zł netto/sms)



podstawowa opieka klienta
kontakt emailowy w ciągu 24 godzin

...i wszystkie funkcje systemu, w tym

- ✓ grafiki pracy
- ✓ historia rezerwacji
- ✓ baza klientów
- ✓ rezerwacje online
- ✓ wizytówka (Twój profil www)

bb: komfort

oferta standardowa

50 zł netto/mies.

(61,5 brutto)

przy płatności za rok z góry;
65 zł netto przy płatności miesięcznej

[testuj za darmo do 30 dni >](#)

pełna funkcjonalność systemu bez SMSów
i bez zmiany liczby kalendarzy

2+

liczba kalendarzy w abonamencie
2 kalendarze z możliwością dodania następnych
(24 zł netto/każdy dodatkowy kalendarz)

∞

bez limitu rezerwacji w miesiącu

100

SMSów abonamencie
dodatkowe SMSy do kupienia w pakietach
(0,14 zł netto/sms)



rozszerzona opieka klienta
telefonicznie i online, bez czekania

...i wszystkie funkcje systemu, w tym

- ✓ grafiki pracy
- ✓ historia rezerwacji
- ✓ baza klientów
- ✓ rezerwacje online
- ✓ wizytówka (Twój profil www)

oferta obowiązuje do 10 kalendarzy w firmie, powyżej tej liczby prosimy o [kontakt](#)

zniżki dla organizacji non-profit

Jeśli robisz coś dobrego dla świata, my chcemy zrobić coś dobrego dla Ciebie. [Skontaktuj się z nami](#), by uzyskać duże zniżki, lub nawet możliwość korzystania z reservisu zupełnie bez opłat.

rezerwacja głosowa dla Twoich klientów

Za dodatkową opłatą 50 zł/mies. możesz umożliwić swoim klientom rezerwowanie wizyt w Twojej firmie za pomocą głosu ([zobacz demo](#)). [Skontaktuj się z obsługą klienta](#) by poznać szczegóły i warunki.

najczęściej zadawane pytania

▼ [czy i na jakich zasadach reservis oferuje okres próbny?](#)

▼ [co oznacza liczba kalendarzy w abonamencie](#)

^ [co oznacza limit rezerwacji w miesiącu](#)

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam feugiat, felis eget sollicitudin dictum, enim purus aliquet lacus, ac rhoncus libero eros non nibh. Curabitur rutrum pellentesque pretium. Fusce consectetur viverra velit vitae efficitur. Sed vel lorem lacinia, cursus dolor id, imperdiet tellus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Maecenas iaculis, neque sit amet varius dignissim, nibh ipsum convallis arcu, ac aliquam est eros non mi. Aliquam bibendum dolor eget enim condimentum, vel ultrices tortor commodo. Morbi blandit