Rozdział 2

SPECYFIKACJA TECHNICZNA WIRTUALNEGO ASYSTENTA

"AI is not the science of building artificial people.
It's not the science of understanding human intelligence.
It's not even the science of trying to build artifacts that
can imitate human behavior well enough to fool someone
that the machine is human, as proposed in the famous Turing test...
AI is the science of making machines do tasks that humans can do or try to do."

James F. Allen, *AI Growing Up*, 1998

1. Kod źródłowy

Projektowanie i budowa wirtualnego asystenta podlega takim samym regułom inżynierii oprogramowania jak każda inna aplikacja komputerowa. Po wyspecyfikowaniu wymagań dotyczących działania i funkcjonalności wirtualnego asystenta przychodzi czas na kodowanie jego programu w jednym z języków programowania i ewentualne kompilowanie kodu (w zależności od typu użytego języka). Następnie przeprowadzane są przez betatesterów wstępne testy poprawności i oceniana jest realizacja pierwotnych założeń dotyczących działania wirtualnego asystenta, wreszcie następuje jego implementacja w określone środowisko funkcjonowania.[108]

Satysfakcjonująca realizacja implementacji wirtualnego asystenta, przejawiająca się w akceptacji użytkowników i ergonomii interfejsu, zależy w dużym stopniu od umiejętności komunikacyjnych wirtualnego asystenta, od całokształtu jego sylwetki oraz zachowań. Na uproszczonym modelu koncepcyjnym wirtual-

[108] Dobrowolski G., *Technologie agentowe w zdecentralizowanych systemach informacyjno–decyzyjnych*, Wydawnictwa AGH, Kraków 2002, s. 34.

nego asystenta, za jego zachowanie można uznać realizowanie zaprogramowanego skryptu, czyli wykonywanie w czasie rzeczywistym następujących po sobie sekwencji kodu, przejawiających się w postaci słownych wypowiedzi, mimiki twarzy, gestów połączonych z interakcjami z innymi obiektami lub programami obecnymi bezpośrednio w środowisku implementacji wirtualnego asystenta.

Najbardziej rozpowszechnioną metodą budowy wirtualnego asystenta jest napisanie zbioru skryptów (biblioteki) realizowanych działań wykonywanych przez wirtualnego asystenta w zależności od kontekstu dialogowego, w którym się znajdzie. Osobną, ale zdecydowanie najważniejszą kwestią, jest napisanie mechanizmu głównego, który niejako zarządza całą biblioteką skryptów, równocześnie rozpoznając sytuację i zlecając wykonanie odpowiedniej akcji. Stąd też bardzo ważne jest dobranie odpowiedniego środowiska programistycznego, które pozwoli najlepiej zrealizować wymagania dotyczące funkcjonowania wirtualnego asystenta. Od użytego języka programowania zależy bowiem elastyczność jego mechanizmu. Dotychczas, opierając swój wybór na osobistej, subiektywnej ocenie dotyczącej możliwości oferowanych przez dany język, twórcy wirtualnych asystentów wybierali następujące języki programowania: Visual Basic, Java Script, TVML, C++, D, Java, Python, PHP z bazą MySQL. Niezależnie od wybranego języka, w każdym przypadku staje się on swoistym interfejsem programowania aplikacji (*Application Programming Interface,* API), który pozwala na specyfikację zachowania wirtualnego asystenta na pewnym poziomie abstrakcji.[109]

Programiści mają do swojej dyspozycji od 1995 roku także język AIML (*Artificial Intelligence Mark–up Language*) — przeznaczony specjalnie do konstrukcji wirtualnych asystentów obecnych na stronach internetowych. Podstawy tego języka stworzył Richard Wallace (wspomniany wcześniej twórca A.L.I.C.E.). Zaprojektowany przez niego mechanizm konwersacji wirtualnego asystenta nie analizował wszystkich słów wpisanych przez użytkownika, lecz koncentrował się na kluczowych słowach i zdaniach, wyznaczających ogólny kierunek rozmowy. AIML ma na celu przekazywanie i przetwarzanie wiedzy w sposób ustrukturyzowany między użytkownikiem a wirtualnym asystentem w Internecie, ułatwiając jego implementację w kodzie HTML i umożliwiając wymianę danych między AIML i XML oraz jego pochodnych, jak np. XHTML. Do 2001 r. architekturę języka AIML rozwijali programiści fundacji A.L.I.C.E. AI Foundation oraz spo-

[109] André E., Rist T., *Controlling the Behaviour of Animated Presentation Agents in the Interface: Scripting versus Instructing*, AI Magazine 2001, nr 22 (4), Special Issue on Intelligent User Interfaces, AAAI Press, s. 54.

łeczności Alicebot jako niekomercyjny projekt rozwoju open source'owej technologii budowy wirtualnych asystentów.[110] AIML opiera się na obiektach składających się z kategorii i, opcjonalnie, z kontekstu. Kategorie są podstawową jednostką wiedzy w AIML. Każda kategoria ma odrębną zasadę dotyczącą rozpoznawania wzorca wypowiedzi użytkownika (*pattern*) i dopasowania do niej szablonu generującego wypowiedź wirtualnego asystenta (*template*). Wzorzec w AIML jest prosty: składa się wyłącznie ze słów, spacji oraz symboli „_" i „*", przy czym słowa mogą zawierać tylko litery i cyfry. Zasadniczą ideą tej techniki jest znalezienie w bazie wiedzy wirtualnego asystenta najlepszego i najdłuższego dopasowania szablonu wypowiedzi do danego wzorca.[111] Zasadność użycia języka AIML została potwierdzona, gdy oparta o AIML wirtualna asystentka A.L.I.C.E. trzykrotnie (w latach: 2000, 2001, 2004) zdobyła nagrodę Loebnera w corocznym międzynarodowym konkursie na chatterbota najlepiej imitującego człowieka podczas rozmowy.[112] Język AIML nadal zdobywa dużą popularność wśród twórców wirtualnych asystentów.

Od roku 2000 jest również rozwijany język VHML (*Virtual Human Markup Language*), podobnie jak AIML oparty o XML. Zamierzeniem twórców języka VHML jest ułatwienie naturalnej i realistycznej interakcji wirtualnego asystenta z użytkownikiem za pośrednictwem strony internetowej lub przypisanej aplikacji. Język VHML jest głównym językiem kontrolującym i zarządzającym całością zachowań wirtualnego asystenta, natomiast poszczególne warstwy działania są modelowane przez sześć podległych komponentów:

- DMML (*Dialogue Manager Markup Language*) — modelowanie prowadzonego dialogu,
- FAML (*Facial Animation Markup Language*) — modelowanie animacji i wyrazu twarzy oraz mimiki,
- BAML (*Body Animation Markup Language*) — modelowanie animacji, postawy ciała oraz gestów,
- SML (*Speech Markup Language*) — modelowanie syntezy mowy,
- EML (*Emotion Markup Language*) — modelowanie wyrażanych emocji,
- HTML (*HyperText Markup Language*) — osadzenie w dokumencie html.

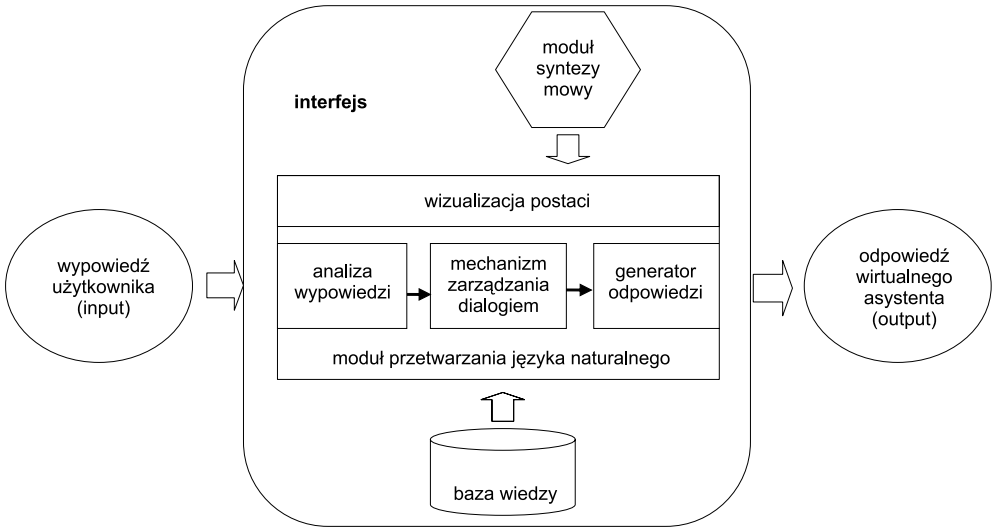
[110] Bush N., *Artificial Intelligence Markup Language (AIML) version 1.0.1*, Working Draft, A.L.I.C.E. AI Foundation 2005, www.alicebot.org/TR/2005/WD–aiml (styczeń 2011).

[111] Abu Shawar B., *Chatbots are natural web interface to information portals*, The 6th International Conference on Informatics and Systems, INFOS 2008, Cairo 2008, s. 102.

[112] Konkurs Loebnera jest oparty na teście Turinga, oceniającym konwersacyjne umiejętności chatterbotów, www.loebner.net/Prizef/loebner–prize.html (styczeń 2011).

Mając wybrane narzędzie do budowy wirtualnego asystenta w postaci języka programowania, kolejnym istotnym krokiem jest zaprojektowanie i wykonanie poszczególnych elementów składowych wirtualnego asystenta. Mogą one nieznacznie różnić się, w zależności od koncepcji twórcy wirtualnego asystenta. Są to jednak najczęściej: interfejs służący do wprowadzania i wyświetlania wypowiedzi oraz „wnętrze", czyli: baza wiedzy, moduł przetwarzania języka naturalnego wraz z mechanizmem zarządzania prowadzonym dialogiem, opcjonalny moduł syntezy mowy oraz wizualizacja postaci. Strukturę łącznego współdziałania wszystkich elementów obrazuje rys. 1, a składowe przedstawione na rysunku omówię kolejno w następnych podrozdziałach.

Rys. 1. Budowa wirtualnego asystenta.



Źródło: opracowanie własne na podstawie Lee S.-I., Cho S.-B., *An intelligent agent with structured pattern matching for a virtual representative*, Intelligent Agent Technology, Japan, Maebashi 2001, s. 3.

2. Baza wiedzy

Wiedza ma znaczenie fundamentalne dla wirtualnego asystenta. Jest „ogółem wiarygodnych informacji o rzeczywistości wraz z umiejętnością ich wykorzystywania".[113] Wiedza i informacje, które są przechowywane w bazie wiedzy wirtualnego asystenta, zawarte są przeważnie w formie zdań lub wyrażeń. Za-

[113] *Nowa Encyklopedia Powszechna*, Wydawnictwo Naukowe PWN, Warszawa 2004.

zwyczaj wprowadzane są one przez osobę czuwającą nad rozwojem bazy wiedzy, tzw. botmastera, jako wypowiedzi używane w mowie potocznej w języku danego kraju lub ewentualnie w dialekcie danego regionu.[114] Również wszelkie aktualizacje zawierające nową, dodatkową porcję wiedzy wprowadzane są każdorazowo przez człowieka. Oczywiście wirtualny asystent może posiadać mechanizm samodzielnego uczenia się lub przyswajania nowych informacji na podstawie wypowiedzi rozmówców, lecz najczęściej takie projekty rozwijane są w laboratoriach naukowych poświęconych doświadczeniom z zakresu sztucznej inteligencji.[115] Badania nad uczeniem się wirtualnych asystentów na podstawie przeprowadzonych rozmów wciąż trwają, stąd też umiejętności takiej nie mają jeszcze wirtualni asystenci implementowani w przedsiębiorstwach. Bardzo często zdarza się, że użytkownicy „testują" wiedzę i zachowanie wirtualnego asystenta podczas rozmowy, pytając go o pojęcia abstrakcyjne, często wprowadzając go specjalnie w błąd lub obrzucając go wulgarnymi wyzwiskami. Gdyby wirtualny asystent przyswajał nową wiedzę na podstawie takich niekontrolowanych rozmów, jego baza wiedzy w szybkim tempie zamieniłaby się w zwyczajny śmietnik. Dodatkowo, świeżo nauczonymi wypowiedziami wirtualny asystent odstraszyłby i obraziłby następnych swoich rozmówców, którzy są przecież potencjalnymi klientami właściciela strony. Dopiero gdy mechanizm uczenia się będzie odpowiednio nadzorowany, a proces zapamiętywania dostatecznie udoskonalony, wówczas z pewnością firmy rozwijające chatboty będą dużo bardziej skłonne do budowy wirtualnego asystenta, który potrafi się uczyć, a z kolei przedsiębiorstwa — do zatrudnienia go.

Mechanizm wirtualnego asystenta musi przewidywać możliwość zapytania go o to, co już jest zgromadzone w bazie wiedzy. Dlatego też reprezentację zawartości bazy wiedzy umożliwia język reprezentacji wiedzy. Jest on definiowany przez swoją składnię (*syntax*), która określa strukturę zdania, i swoją semantykę (*semantics*), która określa prawdziwość każdego zdania w danym kontekście.[116] Wirtualny asystent, korzystając ze swojej bazy wiedzy, powinien móc reagować odpowiednio do zadawanych pytań i dostosowywać swoje wypowiedzi do kontekstu rozmowy, zmiany tematu rozmowy lub przerwy w rozmowie.

---

[114] Abu Shawar B., Atwell E., *A chatbot system as a tool to animate a corpus*, ICAME Journal 2005, nr 29, s. 6.

[115] Korczak, J., Lipiński, P., *Technology of Intelligent Agents used in Financial Data Analysis*, w: Niedzielska E., Dudycz H., Dyczkowski M. (red.), *Nowoczesne technologie informacyjne w zarządzaniu*, Prace Naukowe AE Nr 1134, Wydawnictwo Akademii Ekonomicznej, Wrocław 2006, s. 352.

[116] Russell S., Norvig P., *Artificial Intelligence: A Modern Approach*, Prentice Hall Series in Artificial Intelli-gence, Berkeley 2003, s. 232.

Baza wiedzy zawiera zawsze podstawowe dane o pracodawcy wirtualnego asystenta i oferowanych przez niego produktach lub/i usługach. Wirtualny asystent najczęściej dysponuje również zgromadzoną odpowiednio wiedzą „ogólną", potrzebną do prowadzenia zwykłych towarzyskich pogawędek (choćby o przysłowiowej pogodzie) w celu zmniejszenia komunikacyjnego dystansu między nim a rozmówcą oraz uzyskania wrażenia przyjaznego i chętnego do współpracy konsultanta.[117] Wirtualny asystent, będący równorzędnym człowiekowi partnerem w rozmowie, bierze pełnoprawny udział w konwersacyjnych interakcjach, dlatego istotne jest włączenie do bazy wiedzy również niezbędnej warstwy psychologicznej, zawierającej osobowość i wyrażane emocje. Czyni to z wirtualnego asystenta bardziej wiarygodnego partnera w relacjach społecznych. Odrębną kwestią pozostaje decyzja, czy **każda** ludzka emocja powinna znaleźć swe odzwierciedlenie w bazie wiedzy. Sytuacje, w których wirtualny asystent wyrażałby swe negatywne emocje, raczej nie powinny mieć miejsca podczas rozmowy, gdyż jego rolą jest pomóc rozmówcy, a nie kłócić się z nim lub obrażać go.[118]

Początkowo twórcy pierwszych wirtualnych asystentów próbowali przewidzieć wszelkie możliwe pytania, które może zadać użytkownik w trakcie rozmowy, a następnie starali się napisać wszelkie możliwe odpowiedzi, których adekwatnie mógłby użyć wirtualny asystent. I faktycznie, sporo domyślnych, łatwych do przewidzenia wypowiedzi można umieścić w bazie wiedzy asystenta: powitania, pożegnania, pozdrowienia, zagajenia o zainteresowania lub ulubione potrawy.[119] Nie sposób jednak przewidzieć dokładny scenariusz każdej rozmowy, gdyż mogą zostać poruszone nietypowe wątki lub może nastąpić nagła zmiana tematu rozmowy. Stąd też wynikła konieczność zastąpienia początkowej metody budowy bazy wiedzy wirtualnego asystenta (tj. wpisywanie jak największej liczby potencjalnych pytań i odpowiedzi) przez bardziej zaawansowany mechanizm wnioskowania, czyli moduł przetwarzania języka naturalnego.

[117] Bickmore T., Cassell J., *How about this weather? Social Dialog with Embodied Conversational Agents*, Proceedings of the American Association for Artificial Intelligence (AAAI) Symposium on Narrative Intelligence, Cape Cod, Massachusetts 2000, s. 8.

[118] Becker C., Kopp S., Wachsmuth I., *Why Emotions Should be Integrated into Conversational Agents*, w: Nishida T. (red.), *Engineering Approaches to Conversational Informatics*, John Wiley & Sons 2007, s. 5, 34, 35.

[119] Stanisławski P., *Mówi, choć nie myśli*, Przekrój 2005, nr 47, Edipresse Polska S.A., Warszawa, www.przekroj.pl/cywilizacja_nauka_artykul,1082.html (styczeń 2011).

3. Moduł przetwarzania języka naturalnego

Opisana w poprzednim podrozdziale baza wiedzy jest powiązana z modułem przetwarzania języka naturalnego, będącym pewnego rodzaju „tłumaczem", za pośrednictwem którego użytkownik komunikuje się z wirtualnym asystentem. Moduł przetwarzania języka naturalnego analizuje wypowiedź użytkownika pod kątem morfologii, składni i semantyki, przekłada ją na formalny język zapytań (*query language*) skierowany do swojej bazy wiedzy, następnie korzysta z informacji, zdań i wyrażeń przechowywanych w bazie wiedzy, by na ich podstawie wygenerować zrozumiałą dla użytkownika odpowiedź poprzez „usta" wirtualnego asystenta.[120] Podstawowe trudności związane z funkcjonowaniem modułu przetwarzania języka naturalnego, z którymi borykali się twórcy pierwszych programów konwersacyjnych, dotyczyły „technicznej" natury dialogu: identyfikacji słów kluczowych w danej wypowiedzi. Największym wyzwaniem była bowiem (i jest nadal) konstrukcja mechanizmu, który potrafi choćby w minimalnym stopniu wykryć kontekst dialogu i zareagować wypowiedzią wirtualnego asystenta, a w przypadku braku rozpoznania słów kluczowych — wygenerować inną wypowiedź dostosowaną do tematu rozmowy.[121]

Metoda dopasowywania wzorca (*pattern matching*) to najprostszy i najwcześniej stosowany mechanizm modułu przetwarzania języka naturalnego. Polega on na odnalezieniu w zdaniu wpisanym przez użytkownika jednego (lub kilku) wzorców. Przyjmuje się, że wzorzec to zdanie, którego pewne części zostały zastąpione przez wieloznaczny symbol „*" (*wild card*). Każdy wzorzec zdefiniowany w bazie wiedzy posiada odpowiadający mu schemat wypowiedzi. Podstawiając pod każdy wieloznaczny symbol „*" odpowiednio wybrane słowo z grupy słów z bazy wiedzy, chatterbot dopasowuje swoją wypowiedź do rozpoznanego wzorca.[122] Dopasowywanie wzorca funkcjonuje dobrze w prostych rozmowach, nie wystarcza jednak podczas próby analizowania kontekstu wypowiedzi. Zawodzi również w przypadku zdań złożonych, wymagających pogłębionej analizy.

Kolejną metodą stosowaną w module przetwarzania języka naturalnego jest indeksowanie. W klasycznym podejściu wydobywania informacji (*Information*

[120] Baborski A. (red.), *Efektywne zarządzanie a sztuczna inteligencja*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 1994, s. 26.

[121] Weizenbaum J., *ELIZA – A Computer Program For the Study of Natural Language Communication Between Man and Machine*, Communications of the ACM 1996, nr 9 (1), Cambridge, Massachusetts, s. 40.

[122] Vrajitoru D., *Evolutionary Sentence Building for Chatterbots*, GECCO, Late Breaking Papers 2003, s. 316.

Retrieval, IR) dany jest zbiór dokumentów (tekstowych, napisanych w języku naturalnym) oraz zapytanie wyrażone przez człowieka w języku naturalnym, przy czym zadaniem systemu IR jest odnalezienie w całym zbiorze dokumentów najlepiej pasujących do zapytania. Rozszerzając metodę IR dla mechanizmu wirtualnego asystenta, należy przyjąć uproszczenie, że każdy dokument zawiera tylko jedno lub dwa zdania będące odpowiedziami asystenta. Wówczas wypowiedź wpisana przez użytkownika w trakcie dialogu jest „zapytaniem wyrażonym przez człowieka w języku naturalnym", a zadaniem wirtualnego asystenta jest udzielenie odpowiedzi, czyli odnalezienie w całym zbiorze jednego dokumentu „najlepiej pasującego do zapytania". Wyszukiwanie odpowiedniego dokumentu jest możliwe dzięki wcześniejszemu indeksowaniu dokumentów (*indexing of documents*), czyli po prostu przypisaniu każdemu dokumentowi kilku reprezentatywnych słów kluczowych, przy czym indeksowanie może odbywać się w trybie „ręcznym" lub automatycznie. Zaletą tej metody jest większa elastyczność rozmowy prowadzonej przez wirtualnego asystenta oraz bardzo przejrzysta organizacja bazy wiedzy — do tego stopnia, że może być następnie z powodzeniem wykorzystywana przez inne aplikacje. Wadą jest natomiast niemożność wygenerowania nowej odpowiedzi — takiej, która nie istnieje wcześniej w bazie wiedzy, co oczywiście ogranicza potencjalne zróżnicowanie wypowiedzi wirtualnego asystenta.[123]

Jeśli wypowiedź wpisana przez użytkownika nie pasuje do żadnego wzorca ani do zaindeksowanych słów kluczowych, wówczas dopiero przyjmuje się regułę losowego generowania odpowiedzi. Baza wiedzy zazwyczaj zawiera pewną liczbę gotowych wypowiedzi, które mogą być użyte właśnie w tej awaryjnej sytuacji. Metoda ta może być ulepszona w taki sposób, że w zależności od wpisania przez użytkownika: wypowiedzi oznajmującej, pytania (przez rozpoznanie zaimka pytającego: kto, co, gdzie, ile itp. albo znaku zapytania) lub silnego wyrażenia emocji (przez rozpoznanie emotikonów albo znaku „!"), losowana jest wypowiedź wirtualnego asystenta z odpowiedniej kategorii lub subkategorii reakcji emocjonalnej.[124]

Dalsze próby nadania większej elastyczności rozmowie z wirtualnym asystentem zaowocowały pojawieniem się bardziej zaawansowanych nowych metod, mających na celu stworzenie naturalniejszego modelu rozmowy. Najnowsze badania dotyczą między innymi wspomnianego wcześniej indeksowania z uży-

[123] Vrajitoru D., *NPCs and Chatterbots with Personality and Emotional Response*, The IEEE Symposium on Computational Intelligence and Games (CIG 2006), Reno/Lake Tahoe 2006, s. 143.

[124] Ibidem, s. 144.

ciem semantycznych sieci Bayesowskich (*Semantic Bayes Network,* SeBN), które znalazły już zastosowanie w innych dziedzinach, takich jak automatyczne tworzenie odsyłaczy, filtrowanie informacji oraz klasteryzacja i klasyfikacja dokumentów tekstowych.[125]

Obecnie tworzeni wirtualni asystenci w swoim module przetwarzania języka naturalnego wykorzystują zazwyczaj kilka opisanych powyżej metod łącznie. Należy dodać, że silnik konwersacyjny wirtualnego asystenta można odpowiednio dostosować również do zadań niepolegających na dialogu, jak: analiza zapytań w call–center, automatyczna analiza wiadomości e–mail, wydobywanie informacji z baz danych i dokumentów tekstowych.[126] Tak więc konwersujący wirtualni asystenci to nie jedyna dziedzina, w której znajdują zastosowanie badania nad przetwarzaniem języka naturalnego.

Poza sposobem analizowania wypowiedzi użytkownika i generowania wypowiedzi wirtualnego asystenta, moduł przetwarzania języka naturalnego powinien współpracować z mechanizmem zarządzania dialogiem prowadzonym przez wirtualnego asystenta. Mechanizm ten odpowiada za utrzymywanie spójności i kontekstowości prowadzonego dialogu oraz, poprzez kontrolowanie przejmowania inicjatywy w rozmowie, zapewnia realizację zarówno celów użytkownika, jak i wirtualnego asystenta.

Podstawowymi technikami zarządzania dialogiem, wykorzystywanymi już w systemach dialogowych (*dialogue systems*), są techniki zorientowane na wykonanie zadania: **technika oparta o stany** (*state–based technique*) i **technika ramowa** (*frame–based technique*, często w literaturze określana też jako *slot–filling technique*). Technika oparta o stany reprezentuje każdy dialog jako serię stanów, czyli pytań wirtualnego asystenta i odpowiedzi użytkownika. W każdym stanie wirtualny asystent może: poprosić użytkownika o udzielenie konkretnej informacji, wygenerować własną odpowiedź lub uzyskać dostęp do zewnętrznej aplikacji. Scenariusz takiego dialogu jest z góry ustalony, w każdym ze stanów oczekuje się adekwatnej reakcji użytkownika. Technika ta sprawia, że wypowiedzi użytkownika są w miarę łatwe do przewidzenia i dialog może odbywać się sprawnie i szybko, jednakże odbywa się to kosztem elastyczności i naturalności prowadzonego dialogu. W przypadku

[125] Kim K.-M., Hong J.-H., Cho S.-B., *A semantic Bayesian network approach to retrieving information with intelligent conversational agent*, Information Processing & Management 2007, nr 43(1), s. 227.

[126] Budzikowska M., Chai J., Govindappa S., Horvath V., Kambhatla N., Nicolov N., Zadrozny W., *Conversational Sales Assistant for Online Shopping*, Demonstration at Human Language Technologies Conference (HLT'2001), San Diego, California 2001, s. 1.

prostych zadań (np. przeprowadzania użytkownika przez proces składania zamówienia w sklepie internetowym, pomocy w wypełnianiu formularza online) technika ta jest często stosowanym rozwiązaniem. Jej ograniczenia polegają natomiast na braku adekwatnych reakcji w przypadku wystąpienia nieoczekiwanych zmian w scenariuszu prowadzonego dialogu — wtedy wszelkie dodatkowe wypowiedzi użytkownika, nieprzewidziane pytania lub prośby mogą nie zostać potraktowane w oczekiwany przez rozmówcę, pomocny sposób.[127] Dlatego też w przypadku bardziej złożonych zadań lepiej sprawdza się technika ramowa, która zamiast wystąpienia serii stanów wykorzystuje pojęcie ramy. Rama, oznaczająca jedno złożone zadanie, polega na zebraniu jak największej ilości cząstkowych informacji od użytkownika, w celu wykonania tego zadania. Zaoszczędza to użytkownikowi serii krótkich, z góry zaplanowanych, nużących pytań i pozwala na mniej szablonowy dialog w porównaniu do techniki opartej na stanach. Z drugiej jednak strony, wypowiedzi użytkownika mogą być trudniejsze do przewidzenia, co wydłuża czas przeznaczony na realizację jednego ramowego zadania o ewentualne dodatkowe pytania stawiane użytkownikowi w celu doprecyzowania uzyskanych informacji.[128]

Bardziej zaawansowanymi technikami, oferującymi najwięcej możliwości w trakcie zarządzania dialogiem, są **techniki zorientowane na plan rozmowy** (*plan–based techniques*). Zamiast koncentrowania się na realizacji zadań, techniki oparte o plan polegają na identyfikacji planu użytkownika dotyczącego rozmowy i określeniu zakresu pomocy w realizacji tego planu. Jest to proces dynamiczny, w którym na bieżąco generowane są tematyczne dialogi i kontekstowe subdialogi, dzięki czemu możliwa jest wielowątkowa rozmowa, reagowanie na zmianę tematu wypowiedzi itp. Techniki te zazwyczaj dają użytkownikowi dużą swobodę i większą inicjatywę w rozmowie niż techniki zorientowane na zadania, jednakże ich zaprojektowanie i implementacja są dużo bardziej skomplikowane.[129]

Modelując konwersacyjne zachowanie wirtualnego asystenta, należy pamiętać, że podstawową jednostką interakcji wirtualnego asystenta z użytkownikiem

[127] Branting K., Lester J., Mott B., *Dialogue Management for Conversational Case–Based Reasoning*, Proceedings of the 7th European Conference on Case–Based Reasoning (ECCBR 2004), Springer–Verlag, Madrid 2004, s. 3.

[128] Spiliotopoulos D., Androutsopoulos I., Spyropoulos C.D., *Human–Robot Interaction based on spoken natural language dialogue*, European Workshop on Service and Humanoid Robots (ServiceRob–2001), Greece, Santorini 2001, s. 2.

[129] Allen J., Byron D., Dzikovska M., Ferguson G., Galescu L., Stent A., *Towards Conversational Human–Computer Interaction*, AI Magazine 2001, nr 22 (4), Special Issue on Intelligent User Interfaces, AAAI Press, s. 28.

jest pojedynczy akt dialogu (*dialogue act*). Dlatego też każdy dialog powinien osiągnąć skuteczność na każdym z trzech poziomów: interakcji, konwersacji i zawartości. Działania na poziomie interakcji dotyczą nawiązywania rozmowy (powitanie), kończenia rozmowy (pożegnanie) oraz ponownego jej podejmowania (z powracającym użytkownikiem). Działania na poziomie konwersacji dotyczą zarządzania tematem i przebiegiem rozmowy, przy czym duża doza proaktywności wirtualnego asystenta zapewnia odpowiednią spójność dialogu. Wówczas, zamiast jedynie odpowiadać na wypowiedzi użytkownika, wirtualny asystent sugeruje zmianę/powrót do głównego tematu rozmowy, na który dysponuje najobszerniejszym zasobem wiadomości w swojej bazie wiedzy. Ewentualna zmiana tematu rozmowy przez użytkownika jest tylko tymczasowa i w pełni kontrolowana przez wirtualnego asystenta. Działania na poziomie zawartości dotyczą zebrania informacji o bieżącym temacie rozmowy i umiejętności podtrzymywania rozmowy poprzez „zwykłe" ogólnikowe pogawędki (znajomość ciekawostek, opowiadanie anegdot itp.).[130]

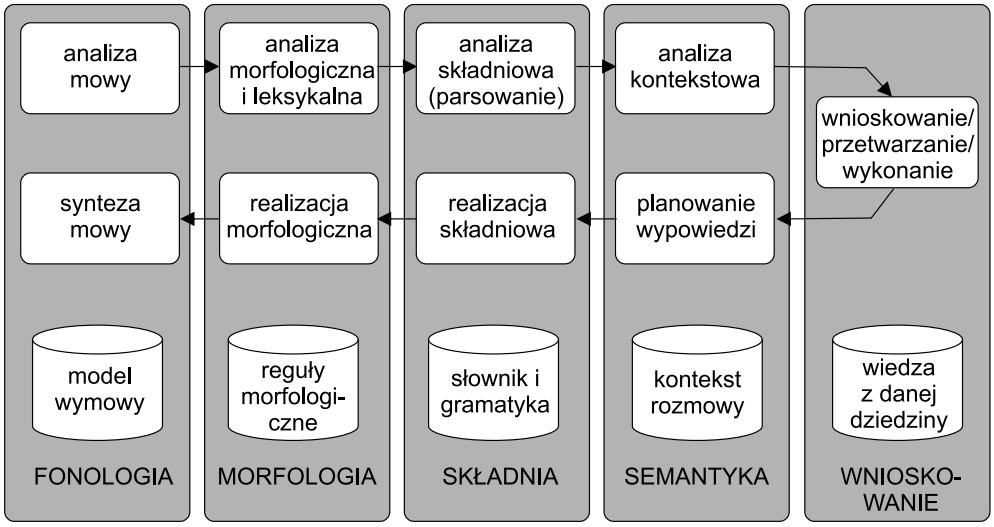
4. Synteza mowy

Konwersacja z wirtualnym asystentem odbywa się za pośrednictwem klawiatury: wirtualny asystent wyświetla swoją wypowiedź, a następnie użytkownik wpisuje swoją tekstową odpowiedź. Tymczasem naturalnym sposobem komunikacji między ludźmi stosowanym od tysięcy lat, wcześniejszym od pisma i o wiele szybszym od niego, jest... mowa. Dlatego też odkąd w latach 60. XX wieku nastąpił rozwój technologii komputerowej, naukowcy zainteresowali się możliwością generowania mowy ludzkiej na podstawie tekstu (*text to speech*) oraz automatycznego rozpoznawania słów wypowiadanych przez człowieka (*automatic speech recognition*). Początkowo prace koncentrowały się na opracowywaniu odpowiednich modeli statystycznych oraz rozwijaniu systemów zdolnych do obsługi rozległego słownictwa głosowego, tzw. korpusów fonetycznych. Próbowano także przezwyciężyć trudności związane z rozpoznawaniem mowy przez telefon, rozpoznawaniem wypowiedzi w hałaśliwym otoczeniu oraz eliminacją naturalnych zakłóceń głosu występujących podczas rozmowy. W badaniach tych czerpano nie tylko z doświadczeń prowadzonych nad technologią mowy, ale dość

[130] Kopp S., Gesellensetter L., Krämer N.C., Wachsmuth I., *A conversational agent as museum guide – design and evaluation of a real–world application*, w: Panayiotopoulos et al. (red.), *Intelligent Virtual Agents*, LNAI 3661, Springer–Verlag, Berlin 2005, s. 334, 336, 337.

nieoczekiwanie skorzystano z wyników badań w dziedzinie przetwarzania języka naturalnego i sztucznej inteligencji, które niezależnie od siebie bardzo prężnie rozwinęły się w ciągu kilku ostatnich dziesięcioleci.[131] Oczywiście przykłady zastosowania syntezatorów mowy narzucają się same: automatyczne odczytywanie tekstu osobie słabowidzącej lub niewidomej, dźwiękowe słowniki języków obcych czytające hasła, automatyczne systemy telefonicznej informacji w działach obsługi klienta. Również twórcy wirtualnych asystentów dostrzegli potencjał tkwiący w możliwościach syntezy mowy i włączają taki moduł do swoich realizacji. Gdyby wirtualnego asystenta, oprócz mechanizmu przetwarzania języka pisanego, wyposażyć zarówno w silnik rozpoznawania ludzkiej mowy, jak i jej syntezator, jego konstrukcja opierałaby się na architekturze prostego systemu dialogowego, którego schemat przedstawiony jest na rys. 2.

Rys. 2. Architektura systemu dialogowego.



Źródło: opracowanie własne na podstawie Bird S., Klein E., Loper E., *Natural Language Processing: Analyzing Text with Python and the Natural Language Toolkit*, O'Reilly Media 2009, www.nltk.org/book (styczeń 2011).

Rysunek 2 ilustruje kolejne kroki komunikacji z systemem dialogowym począwszy od analizowania głosowej wypowiedzi użytkownika (warstwa fonologiczna) i przełożenia jej na tekst, następnie zachodzą kolejne fazy analizy (warstwy: morfologiczna, leksykalna, składniowa, kontekstowa); wreszcie w wyniku

[131] McTear M.F., *Spoken Dialogue Technology: Toward the Conversational User Interface*, Springer–Verlag, London 2004, s. 35–36.

wnioskowania połączonego z wiedzą z danej dziedziny jest przekazany impuls reakcji — najpierw zostaje zaplanowana wypowiedź, następnie jej realizacja składniowa i morfologiczna, a na wyjściu ostatecznie zostaje wygenerowana wypowiedź głosowa systemu.

Niestety technologia rozpoznawania ludzkiej mowy wciąż jeszcze nie jest na takim poziomie, aby można było rozważyć prowadzenie pełnej konwersacji głosowej z wirtualnym asystentem. Stosowane obecnie metody rozpoznawania mowy oparte są najczęściej na modelach Markowa, które trudno byłoby zaimplementować w wirtualnym asystencie napisanym np. w języku AIML, opartym o mechanizm słów i zdań kluczowych.[132] Natomiast możliwość jaka istnieje obecnie i jest powszechnie realizowana, to wyposażenie wirtualnego asystenta w syntezator mowy, odczytujący ludzkim głosem swoje wypowiedzi, podczas gdy użytkownik swoje wypowiedzi musi nadal wpisywać za pomocą klawiatury.

Polskim producentem syntezatora mowy, który zdobył międzynarodowe uznanie, nagrody i jest najczęściej integrowanym syntezatorem w polskich wirtualnych asystentach jest firma IVONA Software sp. z o.o. Wyprodukowała ona syntezator mowy ludzkiej IVONA, która — w wersji męskiej lub żeńskiej — potrafi przekształcić dowolny tekst na naturalny ludzki głos.[133] Syntezator IVONA uzyskał dwukrotnie (w 2006 i 2007 roku) jedną z najwyższych ocen jakości mowy w międzynarodowym konkursie Blizzard Challenge, będącym częścią projektu Carnegie Mellon University poświęconemu badaniu syntezy mowy ludzkiej. Według swoich twórców, IVONA zawdzięcza swoją naturalność i wysoką jakość dźwięku m.in. zastosowanym technikom USLTM (*Unit Selection algorithm with Limited Time–scale Modifications*).[134]

5. Wizualizacja wirtualnego asystenta

Jedną z podstawowych form interakcji międzyludzkich jest rozmowa, której zazwyczaj towarzyszy „mowa ciała": gestykulacja i mimika twarzy. W procesie porozumiewania się sygnały niewerbalne odgrywają równie ważną rolę, co prze-

[132] Atwell E., *Web chatbots: the next generation of speech systems?*, European CEO Journal, November–December 2005, s. 143.

[133] Można przetestować wirtualne głosy i wypróbować działanie syntezatora mowy IVONA na stronie http://say.expressivo.com (styczeń 2011).

[134] Kaszczuk M., Osowski L., *The IVO software Blizzard 2007 entry: improving Ivona speech synthesis system*, The Blizzard Challenge 2007, 6th ISCA Workshop on Speech Synthesis, Bonn 2007, s. 4.

kaz słowny. Zachowanie werbalne i niewerbalne to komponenty silnie ze sobą powiązane, realizujące równolegle wspólny komunikacyjny cel. Warstwa niewerbalna rozmowy jest niezbędna do tego, żeby w pełni poprawnie zinterpretować intencje rozmówcy w kontekście treści jego wypowiedzi.[135] Gesty, będące swego rodzaju komentarzem do wypowiadanych słów, pełniące funkcję ilustrującą, nazywane są w klasyfikacji Ekmana i Friesena właśnie ilustratorami.[136] Uzupełniają one wypowiedź słowną, chroniąc równocześnie przed niepotrzebnym rozbudowywaniem wypowiedzi. Z kolei mimika twarzy wyraża emocje i postawy. Każda z siedmiu podstawowych emocji (radość, zdziwienie, strach, smutek, gniew, obrzydzenie, pogarda[137]) jest reprezentowana przez odpowiadający jej wyraz twarzy. Rozmowa z kimś, kto nadmiernie kontroluje swoją mimikę twarzy i nie ujawnia w najmniejszym stopniu swoich emocji może być dla człowieka trudna lub wręcz frustrująca. Gesty i mimika twarzy dotyczą aktualnie wypowiadanych przez człowieka treści, dlatego takie uzupełnienie komunikatu słownego zastosowano analogicznie w wirtualnych asystentach za pomocą ich wizualizacji.

Początkowo wizualizację wirtualnego asystenta stanowiła statyczna fotografia lub grafika przedstawiająca kobietę, mężczyznę albo abstrakcyjną rysunkową postać. W dalszym rozwoju wirtualnych asystentów, w zależności od treści wypowiedzi wirtualnego asystenta, zmieniała się wyświetlana fotografia lub grafika, choć obrazy te nadal były statyczne. W miarę rozwoju technologii implementacji grafiki, wizualizacje zaczęły być trójwymiarowymi rysunkowymi animacjami przedstawiającymi zazwyczaj ludzką twarz albo całą postać. Wirtualni asystenci zaczęli wyglądać jak żywe osoby, gdyż ich wargi poruszały się w trakcie wypowiadanych przez siebie kwestii, ich powieki mrugały co jakiś czas, a oczy potrafiły „patrzeć", czasem nawet w kierunku przesuwanej na stronie myszki komputerowej. Wreszcie, obecnie stosowane są już trójwymiarowe sekwencje wideo przedstawiające żywego człowieka — konsultanta, który mówi, śmieje się, gestykuluje, a nawet chodzi po swoim wirtualnym pomieszczeniu, które widać za nim w tle.

Antropomorfizacja interfejsu spowodowała traktowanie wirtualnego asystenta przez użytkownika jako podmiotu społecznego. Ludzie podczas rozmowy z wirtualnym asystentem stosują się do społecznych reguł uprzejmości oraz

[135] Jarmołowicz E., *Komunikacja niewerbalna: rola gestów ilustrujących w komunikacji*, Investigationes Linguisticae, t. X, Poznań 2003, s.1–2.

[136] Ekman P., Friesen W.V., *The repertoire of nonverbal behavior: Categories, origins, usage, and coding*, Journal of the International Association for Semiotic Studies, nr 1, Semiotica, 1969, s. 55.

[137] Ekman P., Friesen W.V., *Hand movements*, Journal of Communication 1972, nr 22, s. 356.

ulegają stereotypom dotyczącym wyglądu lub płci, uważając jednych wirtualnych asystentów za bardziej kompetentnych lub bardziej atrakcyjnych niż innych.[138] Obraz „ciała" wirtualnego asystenta to coś więcej niż tylko estetycznie wyglądająca fotografia lub animacja. Dobrze opracowana wizualizacja wirtualnego asystenta dostarcza dodatkowy kanał komunikacji niewerbalnej w formie wizualnego uzupełnienia wypowiadanych kwestii, urozmaicającego interakcję z użytkownikiem. Spojrzenie, gestykulacja i przyjmowanie różnorodnych póz odgrywają szczególną rolę we właściwej realizacji wielu konwersacyjnych zachowań, choćby takich jak rozpoczynanie i kończenie rozmowy, czekanie na swoją kolej w wypowiedzi, wtrącenie komentarza, poprawienie się w przypadku błędu. Wszystkie te zachowania umożliwiają w czasie rzeczywistym wielopłaszczyznową wymianę informacji między wirtualnym asystentem a użytkownikiem.[139]

Integracja wizualizacji wirtualnego asystenta z jego możliwościami konwersacyjnymi może być reprezentowana, podobnie jak konwersacje między ludźmi, za pomocą różnych zachowań komunikacyjnych. Dla większej przejrzystości przykłady funkcji konwersacyjnych i odpowiadających im realizacji w zachowaniu wirtualnego asystenta zostały zebrane w poniższej tabeli.

Tabela 9. Przykłady funkcji konwersacyjnych wirtualnego asystenta i ich realizacji.

Funkcje konwersacyjne	Zachowanie wirtualnego asystenta
Rozpoczynanie i kończenie rozmowy	
– reagowanie	– krótkie spojrzenie
– zachęcenie do rozmowy	– dłuższe spojrzenie, uśmiech
– powitanie	– patrzenie, skinięcie głową, podniesienie brwi, pomachanie ręką, uśmiech
– przerwa w rozmowie	– rozglądanie się dookoła, zajęcie się np. przeglądaniem gazety lub pracą na laptopie
– pożegnanie	– patrzenie, skinięcie głową, pomachanie ręką
Kolejność wypowiedzi	
– kolej użytkownika	– patrzenie, podniesienie brwi
– czekanie na swoją kolej	– gesty rąk
– kolej wirtualnego asystenta	– spojrzenie, rozpoczęcie mówienia

Źródło: opracowanie własne na podstawie Bickmore T., Cassell J., *Social Dialogue with Embodied Conversational Agents*, w: van Kuppevelt J., Dybkjaer L., Bernsen N. (red.), *Natural, Intelligent and Effective Interaction with Multimodal Dialogue Systems*, Kluwer Academic, New York 2004, s. 3.

[138] Cassell. J., *More than Just Another Pretty Face: Embodied Conversational Interface Agents*, Communications of the ACM 2000, nr 43 (4), s. 2.

[139] Cassell J., *Embodied Conversational Agents: Representation and Intelligence in User Interface*, AI Magazine 2001, nr 22 (3), AAAI Press, s. 78–79.

6. Wymagania techniczne implementacji

Wszystkie moduły wymienione we wcześniejszych podrozdziałach, zintegrowane ze sobą, stanowią całość narzędzia będącego wirtualnym asystentem, którego można zaimplementować w przedsiębiorstwie. Jest on złożony z komponentów działających w dwóch otoczeniach informatycznych:

- otoczenie narzędziowe, które zlokalizowane jest na przeznaczonym do tego serwerze,
- otoczenie wykonawcze, czyli interfejs użytkownika umieszczony w serwisie WWW.

Otoczenie narzędziowe obejmuje wszelkie narzędzia dotyczące przetwarzania języka naturalnego i zasobów tekstowych. Można uznać je za „mózg" wirtualnego asystenta. Bardziej szczegółowe wymagania techniczne dotyczące serwera obsługującego implementację różnią się w zależności od dostawcy rozwiązania, czyli firmy produkującej wirtualnych asystentów. Najczęściej firma zakupująca wirtualnego asystenta ma również możliwość wyboru, czy wirtualny asystent ma być umieszczony na wewnętrznym serwerze firmy zakupującej wirtualnego asystenta, czy też ma być wykupiony hosting na serwerze zewnętrznym, należący do firmy produkującej wirtualnych asystentów.

Do programu wirtualnego asystenta funkcjonującego na serwerze dołączony jest panel administratora. Jego wygląd, funkcjonalności, zbierane dane i możliwości analityczne zależą wyłącznie od koncepcji dostarczanej przez daną firmę produkującą wirtualnych asystentów. Panel taki służy do wprowadzania zmian w konfiguracjach wirtualnego asystenta, umożliwia modyfikację i aktualizację bazy wiedzy. Ponadto jest to również narzędzie umożliwiające uzyskanie wglądu w informacje udzielone wirtualnemu asystentowi podczas rozmów. Panel oferuje zintegrowany system monitorujący aktywność asystenta: liczba rozmów, czas trwania rozmów, najczęściej zadawane pytania, poruszane tematy rozmowy itp. — wszystkie te dane są rejestrowane, jak również wszelkiego rodzaju dane wprowadzane przez użytkowników, np. adresy email, miejsce zamieszkania, preferencje produktowe, opinie dotyczące firmy itp. Dane te zazwyczaj można wyeksportować w dostępnym formacie w celu dalszych analiz albo można generować raporty zawierające zestawienia, diagramy i wykresy związane z tekstem rozmów.

Otoczenie wykonawcze, czyli po prostu interfejs wirtualnego asystenta, jest umieszczone w serwisie WWW, najczęściej bezpośrednio na stronie głównej lub w postaci okienka pop–up wyskakującego po kliknięciu np. w odpowiedni ba-

ner, przycisk „pomoc" lub zakładkę „porozmawiaj z asystentem". Poza wyświetlanym tekstem wypowiedzi, coraz powszechniej wirtualny asystent odczytuje je głosowo dzięki syntezatorowi mowy, wykonuje też odpowiednie akcje dotyczące chociażby otworzenia nowego okna przeglądarki z informacjami, których szuka użytkownik. Prowadzone rozmowy z wirtualnym asystentem mogą być częściowo personalizowane, jeśli ustawienia przeglądarki użytkownika mają ustawioną akceptację plików cookies. Użytkownikowi końcowemu do porozmawiania z wirtualnym asystentem wystarczy dowolna przeglądarka internetowa i klawiatura oraz oczywiście połączenie z Internetem. W przypadku, gdy wirtualny asystent oparty jest na technologii Adobe Flash, może się zdarzyć, że przeglądarka zażąda pobrania odpowiedniego plug–inu, jeżeli nie był on wcześniej zainstalowany. Nie stanowi to obecnie żadnego problemu, gdyż ściągnięcie i instalacja takiego dodatku zajmuje zazwyczaj kilkanaście sekund.