

OD TEKSTU W JĘZYKU NATURALNYM DO REALISTYCZNEJ ANIMACJI

Tomasz KAPŁON

Streszczenie: Tekst zawiera propozycje rozwoju systemu oprogramowania umożliwiającego realizację cyfrowych wizualizacji przebiegu sytuacji. Sytuacją może być np. scena filmu, próba przedstawienia, przebieg zdarzenia drogowego, odtworzenie zdarzenia historycznego. Oprogramowanie będzie stanowiło swoiste laboratorium umożliwiające połączenie znanych technologii z zakresu NLP, grafiki komputerowej oraz szeregu metod i technik sztucznej inteligencji. Będzie umożliwiała automatyczne generowanie scen i animacji pozwalających użytkownikom, trenerom, aktorom, scenarzystom czy śledczym, na trening, edukację, weryfikację czy ustalenie jednoznacznego obrazu całości sytuacji oraz możliwość jej analizy.

Słowa kluczowe: przetwarzanie języka naturalnego, sztuczna inteligencja, grafika komputerowa, wizualizacja 3D

1. Wstęp

Ekspertyza procesowa, produkcja filmu pełnometrażowego czy reklamowego, analiza zdarzenia, przygotowanie wizualizacji architektonicznej w klasycznym wydaniu, czyli z zatrudnionymi aktorami, statystami, oświetleniem, wyborem i wynajęciem lokalizacji, obsługą całości, jest przedsięwzięciem drogim i czasochłonnym. W wielu wypadkach jest to proces twórczy, który wymaga prób, dokonywania zmian scenografii, kostiumów czy stworzenia lub oczekiwania na odpowiednie warunki pogodowe. Dodatkowo wymaga konsultacji i ustalania stanowisk w różnych kwestiach. Proponowane oprogramowanie umożliwi testowanie i wizualizację fragmentów lub całości zadania przed końcową realizacją. Dodatkowo umożliwi kontrolowanie zgodności czasu, miejsca i identyczności obiektów użytych w scenach - problem, który znany jest jako tak zwane wpadki filmowe. Istotą projektu jest dostarczenie użytkownikowi narzędzia do realizacji i weryfikacji projektów przestrzennych, ustalania wspólnego zdania, wspólnej wizji czy wyglądu i przebiegu odtwarzanych sytuacji. Pozwala na ustalenie, uzgodnienie bądź weryfikację różnych interpretacji, wizji danej sytuacji lub jej fragmentu. W dobie wszechobecnych urządzeń mobilnych, oprogramowanie powinno dawać możliwość pracy zdalnej i pracy interaktywnej wielowątkowej i wielozadaniowej. Przedstawiono ideę zarówno jak i architekturę takiego systemu.

2. Zadanie i rozwiązania

2.1. Istniejące rozwiązania

Pomysł systemu łączącego metody, techniki i narzędzia z zakresu przetwarzania języka naturalnego, grafiki komputerowej, sztucznej inteligencji i tworzenia wirtualnych środowisk nie jest nowy. Istnieją systemy typu *text-to-graphics*: CarSim [1], WordsEye

[2], SONAS [3], systemy wirtualnych postaci (virtual humans): Improv [4], SAM & REA [5], Gandalf [6], oraz systemy multimedialne kreujące animacje oparte o scenariusze formułowane w języku naturalnym takie jak AESOPWORLD [7], KidsRoom [8], OZ [9], czy ScriptViz [10]. Wszystkie one mają swoje zalety i wady. Dobrze radzą sobie z rozumieniem języka naturalnego czy to tekstu pisanego (*typed-in text*) czy komunikatów werbalnych. Nie wszystkie obsługują trzy kanały wejściowe: tekst pisany, mowa i obraz. Nie wszystkie też dysponują kompletem wyjść w postaci tekstu, tekstu mówionego czy sposób wykorzystują możliwości prezentacji graficznej (np. 2D lub 3D). Ponieważ dedykowane są do określonych zadań, żaden z nich nie stanowi kompletnego i kompleksowego rozwiązania problemu wizualizacji przekazu wprowadzanego w formie tekstu, dźwięku i obrazu, pozwalając na modyfikacje, a przede wszystkim nie wspomagając użytkownika w kontroli zgodności z obowiązującymi lub zdefiniowanymi prawami i regułami świata, w którym rozgrywa się sytuacja.

Idealnym rozwiązaniem byłoby połączenie zalet wszystkich systemów, uzupełnienie ich o wykorzystanie informatyki afektywnej, prezentację w czasie rzeczywistym, grafikę bardzo wysokiej rozdzielczości oraz kontrolę i weryfikację zgodności ze zdefiniowaną rzeczywistością. Proponowany system zakłada takie rozwiązanie.

2.1. Zadanie do rozwiązania

Jednoznaczne umiejscowienie w czasie i przestrzeni dla każdego obiektu i każdej akcji to kluczowy element procesu przetwarzania mającego na celu prezentację przebiegu badanej sytuacji. Aspekt miejsca i czasu są kluczowe dla właściwego działania systemu wizualizacji. Istotą przetwarzania informacji wejściowej (tekstu pisanego, mówionego czy obrazu) jest jednoznaczne wskazanie lokacji w przestrzeni (współrzędnych x , y , z) w każdym momencie (chwili czasu t) rozpatrywanej sytuacji. W celu wygenerowania animacji, czyli pokazania przebiegu sytuacji należy wygenerować kolejne sceny z odpowiednio modyfikowanymi obiektami (ludzie, przedmioty, etc.) z żadaną jakością (rozdzielczość, liczba kolorów, dźwięk, dialogi, emocje, etc.). Jednoznaczność miejsca i czasu (dla każdego z obiektów) eliminuje możliwość istnienia innego przebiegu. W przypadku wątpliwości co do faktycznego przebiegu zdarzenia takie rozwiązanie eliminuje wątpliwości, które mogą występować u obserwatorów, którzy sytuację tę wyobrażaliby sobie różnie w oparciu o ten sam opis. Częściowo problem ten został poruszony w [11]. Tak wyglądałoby rozwiązanie idealne, kiedy posiadalibyśmy kompletną wiedzę o przebiegu sytuacji.

Przeciwnym przypadkiem jest całkowity brak informacji o czasie i lokacji poszczególnych obiektów w przestrzeni zdarzenia. Biorąc jednak pod uwagę sytuację rzeczywiste, przypadek taki, podobnie jak przypadek z kompletną informacją, nie występuje praktycznie. Wynika to z obserwacji sytuacji rzeczywistych, pewnej intuicji oraz prób znalezienia opisu sytuacji o pełnej informacji. W celu uporządkowania przyjęto kilka podstawowych definicji.

Definicja 1. *Sytuacja* to ciąg zdarzeń o określonym czasie trwania rozgrywająca się w określonej przestrzeni z udziałem obiektów wykonujących akcje, wpływających na siebie oraz przebieg zdarzeń.

Definicja 2. *Zdarzenie* to ciąg akcji wykonywanych przez obiekty w określonej przestrzeni i czasie.

Definicja 3. *Akcja* to działanie obiektu bądź obiektów wykonywane w ściśle określonej chwili czasu lub przez określony czas, w określonym miejscu.

Pomiędzy przypadkiem o pełnej informacji a całkowitym brakiem informacji o miejscu i czasie działań znajdują się wszystkie sytuacje rzeczywiste. Można je jednak podzielić na dwie grupy. Podział taki przydatny jest również w praktycznym punkcie widzenia. Pierwsza grupa, to scenariusze, w których zarówno czas jak i miejsce określone są z dużą dokładnością lub precyzyjnie. Drugą grupę stanowią opisy (powieści, zeznania, wiersze, opowiadania itp.), gdzie czas i miejsce albo określone są mniej precyzyjnie, albo w ogóle. Najczęściej zamiennie, albo brakuje pozycjonowania czasowego albo przestrzennego. Szczególną grupę stanowią wiersze, gdzie czas i miejsce albo w ogóle nie występują, albo ich wzajemny związek jest nieokreślony. Zatem, skoro celem jest precyzyjne określenie czasu i miejsca, właściwym kierunkiem przetwarzania i precyzowania jest kierunek opis, scenariusz, jednoznaczność czasu i miejsca. Stanowi to podstawę do utworzenia takiego właśnie sposobu przetwarzania informacji wejściowej w systemie. Następować będzie kolejno coraz to większe precyzowanie kolejnych elementów składających się na całościowy obraz końcowy. W szczególności będzie to dotyczyło aspektu czasu i miejsca. Elementami dodatkowymi, uzupełniającymi możliwości produkcyjne systemu są: detektor emocji (w tekście), modelowanie wyrażania emocji (afektu) obiektów ożywionych, wizualizacja scen 3D. Wygenerowana animacja musi spełniać pewne podstawowe warunki (bazując na kryteriach de Beaugrande i Dresslera dotyczących tekstu): a) kohezja rozumiana jako gramatyka obrazu, czyli kadr - ujęcie - scena - sekwencja [16], b) koherencja - spójność semantyczna, stanowiąca element tak zwanej spójności wizualnej (definicja poniżej), c) intencjonalność - obraz ma wywierać pewien efekt na odbiorcy, d) akceptowalność rozumiana jako współgrająca z oczekiwaniem odbiorcy ale również (i to nie mniej ważne) spełniająca prawa i ograniczenia świata, w którym się rozgrywa, e) informatywność i f) sytuacyjność. Jednym z podstawowych celów projektu jest utrzymanie spójności wizualnej animacji.

Definicja 4. Spójność wizualna to stan, w którym wszystkie obiekty na scenie pozostają ze sobą w takiej korelacji wizualnej, w jakiej występują w określonym świecie. Wynikiem jest zawsze animacja.

3. Przebieg analizy sytuacji

Najważniejszym elementem systemu jest bez wątpienia moduł przetwarzania, a właściwie rozumienia języka naturalnego (NLP). Podstawowym zadaniem parsera jest ekstrakcja struktury składniowej poszczególnych zdań. W przypadku języka angielskiego można wykorzystać np. probabilistyczny parser typu bottom-up Apple Pie [12]. W przypadku języka polskiego możemy skorzystać z parsera Świga 2 lub z tak zwanej Zmodyfikowanej Gramatyki Łączeń [17].

W efekcie działania parsera otrzymujemy drzewo rozkładu składniowego. Na jego podstawie możemy dokonać płytkiej analizy semantycznej zdania. Jest ona wystarczająca do tego aby określić role semantyczne pełnione w zdaniu przez konkretne słowa - tak zwane rdzenie grup semantycznych [13]. Możemy zatem ustalić np. role typu AGENT, ACTION, OBJECT, TIME i LOCATION.

Istotnymi informacjami uzyskiwanymi podczas tej analizy są np. tak zwane wyzwalacze emocji czy wyzwalacze akcji - często niezbędne do ustalenia zależności przyczynowo-skutkowej i związku pomiędzy kolejnymi akcjami oraz związków łączących kolejne działania obiektu w przypadku utraty wątku związanego z danym obiektem przez pojawianie się w tekście zdań z nim nie związanych.

Kolejnym elementem analizy jest konwersja informacji semantycznej w ciąg akcji. Zgodnie z wprowadzonym podziałem na scenariusze i opisy prześledźmy proces powstawania animacji w zależności od rodzaju wejścia.

W przypadku *scenariusza* sztuki, filmu animowanego czy działań według pewnej określonej procedury (przykładem może być opracowanie lub weryfikacja procedur wojskowych bądź policyjnych), kiedy akcje ułożone są w znanej kolejności - kolejne akcje umieszczane są w określonym ciągu i nie jest konieczna żadna dodatkowa analiza treści porządkująca je czasowo. W ramach scenariuszy pozostałe elementy, takie jak lokacje na scenie, czas trwania akcji czy emocje są ściśle określone. Jeżeli nie precyzyjnie w centymetrach czy sekundach, to na pewno za pomocą możliwie, jak na pojęcia lingwistyczne, precyzyjne stwierdzenia. W związku z tym możliwe jest generowanie animacji przebiegu sytuacji w czasie rzeczywistym. Przy ustalonych wcześniej elementach niezmiennych (scenografii), system na bieżąco interpretuje ciąg akcji, wyszukuje asocjacje w bazie animacji (zawartość bazy to krótkie animacje szkieletowe określonych akcji, które uzupełnia się przed renderowaniem o kolory, tekstury, mapy oświetlenia, wyzwalacze emocji i czasy trwania), umieszcza je odpowiednio na scenie i generuje kolejne fragmenty animacji. Całość odbywa się w sześciu fazach (po wykonaniu analizy składniowej i semantycznej).

1. Analiza tekstu i konwersja informacji semantycznej w ciąg akcji.
2. Sprawdzenie wykonalności poszczególnych akcji - Obiektywny Sędzia.
3. Weryfikacja ciągu akcji.
4. Wyszukanie w bazie animacji kolejnych akcji.
5. Dostosowanie obiektów w celu uzyskania spójności wizualnej.
6. Wygenerowanie animacji.

W przypadku *opisu* procedura jest nieco bardziej skomplikowana. Faza 1. nie przypomina już prostej interpretacji tekstu zdanie po zdaniu. Ponieważ opis (wiersz, opowiadanie, zeznanie) nie jest pisane z założeniem ścisłego uporządkowania kolejnych faktów (scen). (W przypadku wierszy dodatkowo mamy do czynienia ze zjawiskiem *licentia poetica*, co skutecznie zaburza jednoznaczność czy nawet bliską jednoznaczność interpretację.) Faza 1. zatem wymaga kilkukrotnej wielokryterialnej analizy, w której można na przykład wyróżnić rozpoznawanie zaimków, uzupełnianie zdań o podmioty domyślne, wyszukiwanie związków przyczynowo-skutkowych pomiędzy akcjami, zdarzeniami i akcjami i zdarzeniami w celu odkrycia kolejności ich występowania. Kolejnym elementem analizy jest wskazanie lokacji obiektów, które częstokroć ukryte są - podobnie zresztą jak chwile czasowe - w stwierdzeniach z pojęciami rozmytymi. Dodatkową trudnością jest śledzenie i analiza wątków. Wątki w scenariuszach są mimo wszystko precyzyjnie rozdzielone. W przypadku opisu, zwłaszcza opowiadania, zarówno liczba wątków jak i ich wzajemne przeplatanie się stanowi problem powodujący konieczność śledzenia obiektów, zapamiętywania ich lokacji i czasu ich pojawiania się. Dla każdego z obiektów (najczęściej tych ożywionych) tworzy się oddzielne ścieżki ich działań, które później wplata się w linię czasową zdarzeń i całej sytuacji [11].

Faza 2. przebiega podobnie, co do zasady, jak w przypadku scenariuszy z tą jednak uwagą, że wymagania na zgodność z prawami (np. fizyki) obowiązującymi w danym świecie są tu zdecydowanie bardziej restrykcyjne. Szczególnie w przypadku odtwarzania sytuacji opisanych jedynie przez ich świadków. W tym przypadku liczba błędnych ocen czasu trwania czy wskazań lokacji poszczególnych obiektów jest wyjątkowo wysoka. Sytuacje, które dodatkowo rozgrywały się w wielu lokacjach a były ze sobą powiązane kontaktami pomiędzy obiektami (np. rozmowa telefoniczna) lub też widziane były przez

kilku świadków z których każdy widział jedynie część sytuacji, stanowią szczególnie trudne zadanie ze względu na konieczność uzupełniania brakujących fragmentów i rozpatrywania wielu wariantów sytuacji. Rozważmy akcję ODPOWIADAĆ (Jan, Anna). Najpierw zweryfikowane zostanie to, czy Jan jest człowiekiem (w ograniczeniach świata istnieje regułą mówiąca, że mówić (tu: pytać) może tylko człowiek). Następnie, jeśli Jan okaże się człowiekiem, system (Obiektywny Sędzia) sprawdzi czy Anna jest człowiekiem - w oparciu o tę samą regułę. Kolejnym krokiem będzie sprawdzenie istnienia możliwości odpowiedzi, czyli kontaktu pomiędzy Janem i Anną - na przykład ta sama lokacja, rozmowa telefoniczna w tej samej chwili.

Faza 3. to najczęściej produkcja wielu ciągów akcji, co jest bezpośrednią konsekwencją możliwych wariantów uzyskanych w fazie 2.

Fazy 4. i 5. pozostają bez zmian. Podobnie faza 6. może być wykonywana wielokrotnie - dla każdego wariantu osobno. Jednak do fazy 6. klasyfikowane będą jedynie najbardziej prawdopodobne warianty. Ocena wystąpienia poszczególnych wariantów oparta są na ocenie prawdopodobieństwa i możliwości wystąpienia kolejnych akcji i zdarzeń.

4. Architektura systemu

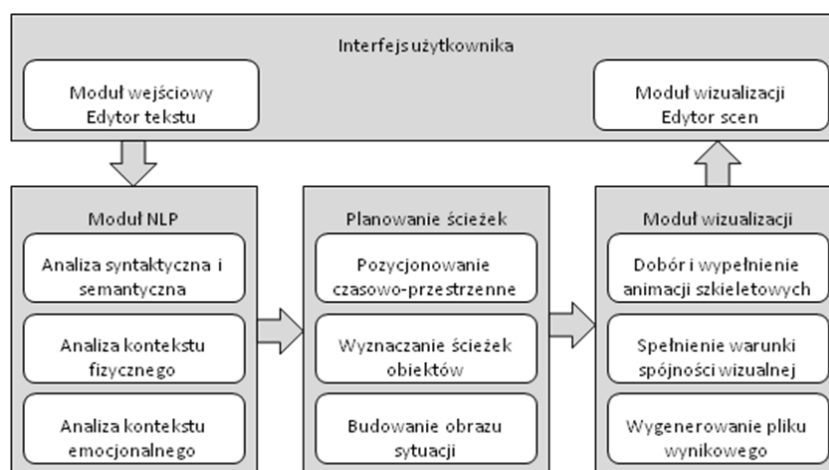
System ma docelowo posiadać przyjazny użytkownikowi interfejs umożliwiający wprowadzanie opisów i scenariuszy w języku naturalnym. Przetwarzanie powinno się odbywać w czasie rzeczywistym. Jednak jak już wspomniano spełnienie tego założenia będzie trudne ze względu na wielkość i skomplikowanie treści opisów. Najprawdopodobniej tekst będzie musiał być wielokrotnie przetwarzany, co nawet w przypadku zrównoleglenia przetwarzania nie pozwoli na jego wizualizację w czasie rzeczywistym. W efekcie generowana będzie animacja. Użytkownik będzie ją mógł następnie modyfikować przesuując obiekty, aranżując sceny. Taka możliwość będzie występowała w przypadku scenariuszy, które stanowią materiał treningowy lub koncepcyjny. W przypadku zadania polegającego na odtworzeniu sytuacji, czyli w przypadku pracy nad tekstem zamkniętym, bez możliwości modyfikacji (wersz, zeznanie) nie będzie możliwości manipulowania elementami scen.

Głównymi zaletami systemu będą:

- a) Wprowadzanie scenariuszy i opisów w języku naturalnym. Optymalnym rozwiązaniem jest możliwość użycia zarówno języka angielskiego ze względu na istnienie podobnych systemów akceptujących właśnie język angielski np. ScriptViz [10] - możliwa będzie porównanie efektów na tych samych danych wejściowych - jak i polskiego ze względu na istnienie znaczącego zaplecza aplikacyjnego w postaci parserów czy leksykonów (np. Słowsieć). Pozwoli to na opracowanie narzędzia przydatnego na rynku polskim. Powinno to zwiększyć efektywność pracy zarówno nad scenariuszami np. filmów, jak również, a może przede wszystkim, nad rozstrzygnięciem wątpliwości co do przebiegu niektórych zdarzeń.
- b) Realistyczne przedstawienie obiektów z ich naturalnymi ruchami i wyrażaniem emocji. Osiągnięcie wyjątkowego realizmu wirtualnych obiektów (np. aktorów) pozwoli na pokazanie emocji na poziomie bardzo bliskim rzeczywistości.
- c) Połączenie wysokiej jakości grafiki z dźwiękiem.
- d) Kontrola realizmu sytuacji. Funkcją systemu ma być istnienie tak zwanego Obiektywnego Sędziego. Ma on być aplikacją kontrolującą zgodność podejmowanych akcji i przebiegu zdarzeń z prawami obowiązującymi w określonym świecie - w przypadku świata rzeczywistego byłaby to np. kontrola nad

możliwością wykonania opisanej akcji z prawami fizyki w danym kontekście z uwzględnieniem możliwości obiektów do wykonania danej akcji w określonym miejscu i czasie.

Architektura systemu pokazana została na rysunku 1. System podzielony jest na 4 moduły: a) interfejs użytkownika, b) przetwarzania języka naturalnego, c) planowania ścieżek akcji, d) wizualizacji. Interfejs użytkownika służy do wprowadzania scenariusz, opisów lub pojedynczych scen (realizując system pierwszym krokiem było/jest sprawdzenie jego możliwości przez wygenerowanie pojedynczej sceny) oraz prezentowania wyników działania - efektów przetwarzania. Moduł przetwarzania języka naturalnego podzielony jest na trzy części. Wykonywane są tam operacje związane z fazą 1. przetwarzania między innymi analiza syntaktyczna (przede wszystkim po to, by sprawdzić poprawność tekstu) i semantyczna (w tym określanie ról semantycznych rdzeni grup w zdaniach), analiza kontekstu fizycznego, czyli wyróżnianie obiektów stanowiących tło fizyczne sytuacji oraz analiza emocji i kontekstu emocjonalnego wyrażanej treści. W module planowania ścieżek akcji realizowane są trzy podstawowe zadania: pozycjonowanie czasowe i przestrzenne oraz wyznaczanie ścieżek akcji dla kolejnych obiektów, ich wzajemna korelacja, rozpoznawanie przynależności akcji do zdarzeń i budowanie na tej podstawie obrazu sytuacji, czyli fazy 2. i 3.. Moduł wizualizacji odpowiedzialny jest za realizację faz 4, 5 i 6 procesu przetwarzania. Przede wszystkim za właściwy dobór animacji szkieletowych do konkretnych akcji, za wypełnienie szkieletów tak aby stanowiły odzwierciedlenie zapisu w tekście pod względem wizualnym (spełnienie warunku spójności wizualnej) całości oraz, co samo w sobie stanowi wyzwanie, wyrenderowanie scen, wmontowanie dźwięku czy mowy, zsynchronizowanie środków wyrazu (tak aby np. ruch warg postaci był zgodny z dźwiękiem jego słów) oraz wygenerowanie całości w postaci pliku wynikowego.



Rys. 1. Architektura systemu wizualizacji treści scenariusza/opisu

5. Podobne rozwiązania

Implementacja różnych aspektów rzeczywistości w oparciu o opisy i scenariusze podejmowana jest od czasu pojawienia się takiej możliwości - powstania współczesnych komputerów. Realizacje te obejmują przede wszystkim aspekt modelowania zachowań

postaci (wirtualnych aktorów), systemy narracyjne i aplikacje przeznaczone do produkcji filmów. Aplikacja ScriptViz [10] generuje animacje w oparciu o scenariusze jednak bez rozbudowanej strony afektywnej postaci. Nie umożliwia jednak przygotowania dłuższych form realizacyjnych o rozbudowanych scenariuszach. EML [14] jest aplikacją opartą o model emocjonalny (postaci) jednak nie umożliwia wyboru czy też kreowania postaci różnych typów emocjonalnych. Z kolei aplikacja CAMEO [15] pozwala na taką kreację, jednak jest ona dostępna tylko z poziomu użytkownika. Typ i sposób zachowania postaci nie jest budowany automatycznie w oparciu o opis postaci w tekście.

6. Wizualizacja w zarządzaniu i projektowaniu procesów produkcji

6.1. Obszary zastosowań

Obecnie ilość liczba danych i ilość informacji pozyskiwana i gromadzona w dowolnym zautomatyzowanym bądź w części zautomatyzowanym procesie produkcji jest duża i czasem zapewne będzie wzrastać chociażby ze względu na możliwość i konieczność monitorowania coraz bardziej zaawansowanych technologicznie procesów. Dane te stanowią, a przynajmniej powinny stanowić, podstawę do obiektywnej oceny każdego procesu produkcyjnego i podejmowania możliwie racjonalnych decyzji, co do sposobów działania z tymi procesami związanymi.

Wizualizacja danych liczbowych jest cechą większości systemów informatycznych. Zakres informacji i wiedzy przekazywanych w ten sposób zależny jest oczywiście od zastosowanych metod analizy i możliwości sprzętowych. W każdym przypadku wizualizacja umożliwia sprawniejszy odbiór informacji i możliwość podejmowania skuteczniejszych decyzji. Wybór, pomiędzy prezentacją informacji w postaci obrazu czy tekstu jest oczywisty. Obraz, w obszarze który zajmuje niesie ze sobą więcej informacji niż tekst, a czas jego przyswojenia jest znacząco krótszy. Obraz jest medium równoległym. W odróżnieniu od tekstu, który ma naturę szeregową. Zarówno obraz jak i tekst wymagają interpretacji. Jednak zarówno adekwatność jak i aktualność informacji przekazywanej przez obraz jest większa, co przenosi się bezpośrednio na szybkość, precyzję i jakość podejmowanych na ich podstawie decyzji, zwłaszcza w sytuacjach krytycznych. Zatem podstawowym obszarem zastosowania wizualizacji jest prezentacja, uprzednio przefiltrowanych i przeanalizowanych danych - prezentacja w postaci wykresów, symulacji bądź animacji.

Innym obszarem zastosowania wizualizacji jest wsparcie projektowania i optymalizacja procesów produkcyjnych bądź technologicznych oraz związane z nimi nierozzerwalnie obszary ergonomii pracy i ekonomii produkcji. Przykładem konieczności zastosowania procesu wizualizacji jest proces projektowania logistyki wewnętrznej. Szczegółową charakterystykę projektowania systemów logistycznych można w pracy [18]. Elementem zalecanym przez autora jest wykorzystanie technologii nazywanej wizualnym modelowaniem interaktywnym VIM (*Visual Interactive Modelling*) jako narzędzia pozwalającego między innymi na wykrywanie błędów, impasów czy wąskich gardeł projektu oraz określenie ich ewentualnych skutków i wpływu na sprzęt czy pracowników, obrazowej prezentacji systemu czy zrozumienia przez użytkownika działania systemu w rzeczywistości.

Kolejnym polem działania, w którym poprzez wykorzystanie wizualizacji (komputerowych) można usprawnić pracę są szkolenia i edukacja. Szczególnie tam, gdzie najistotniejsza jest praktyka wytwarzania. Obrazowanie jest skuteczniejsze w nauczaniu niż

werbalizowanie. W przypadku niektórych zawodów szkolenia ze względu na koszt materiałów są tak duże, że albo liczba uczących się jest minimalizowana albo jakość uczenia jest niska.

Ostatnim ale nie najmniej ważnym obszarem jest projektowanie, testowanie i trenowanie procedur bezpieczeństwa, ewakuacji, kontroli tłumu czy wszystkich tych sytuacji, które zdarzają się rzadko, a których skutki są bardzo kosztowne w przypadku podejmowania niewłaściwych decyzji w trakcie ich trwania.

6.2. Wykorzystanie systemu

Projektowane rozwiązanie ma stanowić rozszerzenie i uzupełnienie możliwości prezentacji, interakcji, analizy i weryfikacji powstających projektów. Po pierwsze ma przedstawiać wizualizację z użyciem grafiki wysokiej rozdzielczości. Ten, jak się wydaje zbędny element, wiąże się jednak bezpośrednio z aspektem niezwykle istotnym a mianowicie weryfikacją zgodności działań z prawami fizyki i podniesieniem wiarygodności prezentacji. Powodem realizacji animacji w wysokiej rozdzielczości jest zmniejszenie efektu subiektywnej oceny wiarygodności prezentowanego materiału w przypadku nieodzwierciedlenia rzeczywistości w prezentacji. Systemy prezentacji przebiegu i skutków wypadków drogowych pod względem odzwierciedlenia rzeczywistości pozostawiają wiele do życzenia. Pomimo korzystania z (miejmy nadzieję) precyzyjnych modeli obiektów i przebiegu zjawisk fizycznych, ze względu na słabość grafiki wiarygodność prezentowanych zdarzeń jest niska. Zadaniem takiego systemu jest jednak odtworzenie zdarzenia ze względu na zgodność z prawami fizyki z uwzględnieniem wszelkich możliwych zjawisk mających wpływ na jego przebieg. Wizualizacja stanowi element dodatkowy, acz niezbędny.

Po drugie, system ma prezentować możliwie dokładne zachowanie się obiektów w oparciu o wiarygodny, co do zastosowanych modeli, silnik fizyki (np. PhysX, Havok czy Euphoria).

Po trzecie i być może najważniejsza z punktu widzenia przebiegu jakiegokolwiek procesu, w tym produkcji, jest kontrola i weryfikacja możliwości wykonania przez obiekty określonych działań. Projektowane przy użyciu systemów CAD ciągi technologiczne uwzględniają wszystkie bądź większość parametrów otoczenia i urządzeń niezbędnych do wykonania prezentacji. Stanowią jednak jedynie mniej lub bardziej uproszczony model środowiska i zachowania. Nie uwzględniają a przede wszystkim nie pozwalają, w sposób automatyczny, na kontrolę przebiegu procesów zachodzących w tym środowisku. W przypadku nieuwzględnienia (nieznajomość, niepewność, zaniedbanie) niektórych zmiennych nie weryfikują przebiegu działania ludzi i urządzeń oraz ich interakcji ze środowiskiem. Proponowany system będzie mógł wykonywać symulacje w oparciu o określone scenariusze - działanie poprawne, awarie, przebrojenie. To, z czego może korzystać reżyser "projektując" sceny, można (układając odpowiedni scenariusz) wykorzystać do projektowania, modernizacji czy weryfikacji procesów produkcyjnych oraz zarządzania takimi procesami. Dodatkowo system może stanowić wiarygodne narzędzie w zakresie podejmowania decyzji. Dzięki symulacjom i animacjom możliwe jest obniżenie kosztów realizacji projektów. (Praktycznie nie ma już na świecie żadnego liczącego się producenta, który nie korzystałby z projektowania z użyciem symulacji komputerowej).

Kolejną, wydaje się, że unikalną, zaletą systemu stanowić ma uwzględnianie zachowań obiektów ożywionych. W przypadku projektowania procesu produkcji być może mało istotną. Jednak w przypadku realizacji scenariuszy awarii bądź wypadków w trakcie pracy,

może dostarczać szeregu danych, informacji i wiedzy dotąd nie uwzględnianych ze względu na niemożność ich weryfikacji w warunkach rzeczywistych ze względu na oczywiste przyczyny. (Analiza wypadków *post factum* na pewno nie może być równie skuteczna, chociażby ze względu na subiektywność ocen analityków).

To tylko część, jak na razie, potencjalnych ale możliwych do uzyskania funkcji systemu.

7. Procedura realizacji scenariusza

Procedura działania zostanie omówiona czysto teoretycznie. Pomimo istnienia pewnych elementów systemu - analizator syntaktyczny z elementami podstawowej analizy semantycznej pozwalający generować ścieżki dla prostych scenariuszy złożonych jedynie ze zdań prostych oraz program wizualizacji ścieżek w postaci wykresów czasowych i pozycji w przestrzeni - nie jest jeszcze możliwe zaprezentowanie realizacji pełnego procesu *scenariusz - analiza - ścieżki - wizualizacja* (fragment takiego procesu pokazany został w [11]). Ze względu na bardzo szeroki zakres projektu i oczywistą konieczność jego podziału, aktualnie trwają prace nad zintegrowaniem elementów pozwalających na wprowadzenie scenariusza, analizę syntaktyczną i semantyczną w zakresie rozpoznania nazw obiektów, akcji, miejsc i czasu, pozycjonowanie czasowo-przestrzenne, wyznaczenie ścieżek, budowę obrazu sytuacji i wygenerowanie pliku z możliwością jego wizualizacji. Efektem działania ma być animacja obiektów w formie szkieletowej (z ewentualnym kolorem wyróżniającym) i minimalistycznym obrazowaniem ruchu.

Ponieważ na tym etapie zaawansowania projektu nie są prowadzone jeszcze analizy scenariuszy rzeczywistych (stąd brak w bazie odpowiedniego scenariusza odnoszącego się bezpośrednio do procesu produkcyjnego), przebieg analizy pokazany zostanie schemacie scenariusza operującego na kategoriach semantycznych wytworzony na podstawie analizy scenariusza hipotetycznej sytuacji opisującej przemieszczanie się obiektów i ich wzajemne relacje. Rozważmy następujący schemat scenariusza.

(1) AGENT (o_1) ACTION1 MANNER1 LOCATION1 LOCATION4 TIME1.

(2) AGENT (o_3) ACTION2 LOCATION3 TIME1.

(3a) AGENT (o_2) ACTION1 LOCATION4 TIME2 *and*

(3b) AGENT (o_2) ACTION4 LOCATION5 OBJECTp4 (op_4).

(4a) AGENT (o_3) ACTION5 TIME21 *and*

(4b) AGENT (o_3) ACTION6 MANNER2 LOCATION5 TIME22,

(4c) AGENT (o_1) ACTION6.

(5a) AGENT (o_1) ACTION61 LOCATION5 *and*

(5b) AGENT (o_1) ACTION7 MANNER3 TIME3.

(6a) AGENT (o_2) ACTION8 OBJECT (op_4) TIME4,

(6b) AGENT (o_2) ACTION9 TIME4 *and*

(6c) AGENT (o_2) ACTION62 LOCATION5 TIME51.

(7a) AGENT (o_3) ACTION9 MANNER3 LOCATION5 *and*

(7b) AGENT (o_2) ACTION10 TIME51 LOCATION5.

gdzie: (s): numer kolejnego zdania (zdania złożone dzielone są na zdania proste a, b, c,...; o_i - obiekty ożywione mogące wykonywać działania autonomicznie; AGENT, ACTION, MANNER, LOCATION i TIME - kategorie semantyczne funkcjonalnych grup słów (rzeczownikowej, czasownikowej, itd.); *and* - symbol łącznika zdań pojedynczych w zdaniu złożonym.

Korzystając z formalizmu obrazowania semantyki [19] otrzymujemy zbiór schematów semantycznych przedstawiony w tabeli 1. Niezgodność numeracji kategorii ACTION z numeracją akcji (act_i) wynika ze sposobu numerowania kategorii semantycznych przez analizator semantyczny i program obrazowania semantyki, który korzysta ze słowników z parami (w uproszczeniu) *słowo - symbol* (np. *iść - act₁*).

Tab. 1. Schematy semantyczne zdań scenariusza

sentence	AGENT	ACTION	MANNER	OBJECT	LOCATION	DESTINATION	TIME
(1)	o_1	act_1 (ACTION1)	1	---	p_1 (LOCATION1)	p_3 (LOCATION4)	t_1 (TIME1)
(2)	o_3	act_5 (ACTION2)	---	---	p_2 (LOCATION3)	p_2 (LOCATION3)	t_1 (TIME1)
(3a)	o_2	act_1 (ACTION1)	---	---	p_7^5	p_3 (LOCATION4)	t_2 (TIME2)
(3b)	o_2	act_2 (ACTION4)	---	op_4	p_7^6	p_7^6	t_3
(4a)	o_3	act_6 (ACTION5)	---	---	p_7^7	p_7^7	t_3 (TIME21)
(4b)	o_3	act_7 (ACTION6)	2	---	p_7^7	p_4 (LOCATION5)	t_4 (TIME22)
(4c)	o_1	act_7 (ACTION6)	---	---	p_7^8	p_7^9	t_4
(5a)	o_1	act_{10} (ACTION61)	---	---	p_4 (LOCATION5)	p_4 (LOCATION5)	t_5 (TIME3)
(5b)	o_1	act_8 (ACTION7)	3	---	p_4 (LOCATION5)	p_4 (LOCATION5)	t_5
(6a)	o_2	act_9 (ACTION8)	---	op_4	p_7^{10}	p_7^{10}	t_6 (TIME4)
(6b)	o_2	act_4 (ACTION9)	---	---	p_7^{10}	p_7^{10}	t_6 (TIME4)
(6c)	o_2	act_{11} (ACTION2)	---	---	p_4 (LOCATION5)	p_4 (LOCATION5)	t_8 (TIME51)
(7a)	o_3	act_8 (ACTION9)	3	---	p_4 (LOCATION5)	p_4 (LOCATION5)	t_7
(7b)	o_2	act_{12} (ACTION10)	---	---	p_4 (LOCATION5)	p_4 (LOCATION5)	t_8 (TIME51)

W skład schematu wchodzi wszystkie grupy słów, których kategorie zostały rozpoznane i nadane we wszystkich analizowanych zdaniach. W tabeli 2 pokazany został zbiór $DP_j(t_i)$ utworzony na podstawie tabeli 1. ($DP_j(t_i)$ to punkty zmiany aktywności (piątki), określające obiekt, akcję, lokację i czas trwania danej akcji wykonywanej przez wskazany obiekt). Oznaczenia $t_{?i}$ oraz $p_{?v}$ dotyczą nieokreślonych chwil lub też ich określenie nie wiąże się bezpośrednio (a jedynie kontekstowo) ze zidentyfikowanymi chwilami i lokacjami. Punkty $DP_j(t_i)$ tworzą (śląd aktywności) ścieżkę $TR(o_j)$ zmian pozycji obiektu w czasie i przestrzeni. (Wykorzystywane oznaczenia pochodzą z [11]. Tam też

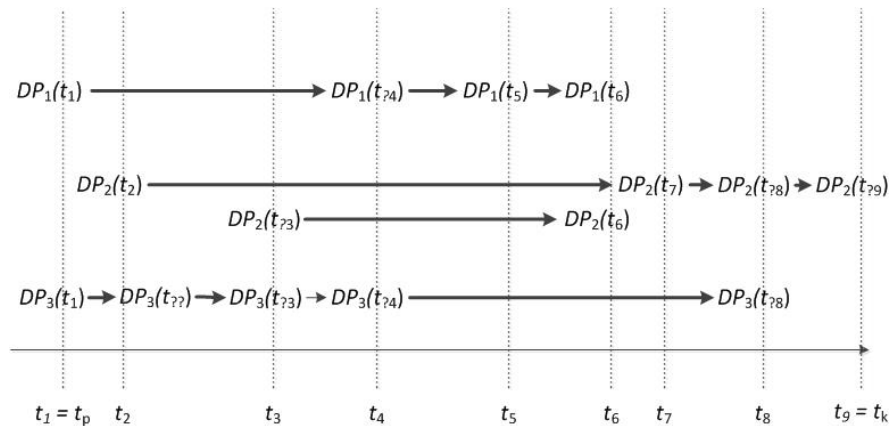
znaleźć można wprowadzenie (definicje i rozważania teoretyczne) do analizy sytuacji od strony rekonstrukcji zdarzeń opisywanych przez obserwatorów).

Tab. 1 Punkty zmiany aktywności $DP_j(t_i)$ i ścieżki czasowo-przestrzenne obiektów $TR(o_j)$

$TR(o_1)$	$TR(o_2)$	$TR(o_3)$
$DP_1(t_1) = \langle o_1, act_1, p_1, t_1, t_{24} \rangle$		$DP_3(t_1) = \langle o_3, act_5, p_2, t_1, t_{77} \rangle$
	$DP_2(t_2) = \langle o_2, act_1, p_3, t_2, t_{73} \rangle$	
	$DP_2(t_{73}) = \langle o_2, act_2, p_{76}, t_{73}, t_6 \rangle$	$DP_3(t_{73}) = \langle o_3, act_6, p_{77}, t_{73}, t_{73} \rangle$
$DP_1(t_{24}) = \langle o_1, act_7, p_{78}, t_{24}, t_{25} \rangle$		$DP_3(t_{24}) = \langle o_3, act_7, p_4, t_{24}, t_{77} \rangle$
$DP_1(t_5) = \langle o_1, act_{10}, p_4, t_5, t_6 \rangle$		
$DP_1(t_6) = \langle o_1, act_8, p_4, t_5, t_k \rangle$	$DP_2(t_6) = \langle o_2, act_9, p_{710}, t_6, t_6 \rangle$	
	$DP_2(t_6) = \langle o_2, act_4, p_{710}, t_6, t_{78} \rangle$	
		$DP_3(t_{77}) = \langle o_3, act_8, p_4, t_{77}, t_k \rangle$
	$DP_2(t_8) = \langle o_2, act_{11}, p_4, t_{78}, t_k \rangle$	
	$DP_2(t_8) = \langle o_2, act_{12}, p_4, t_{78}, t_k \rangle$	

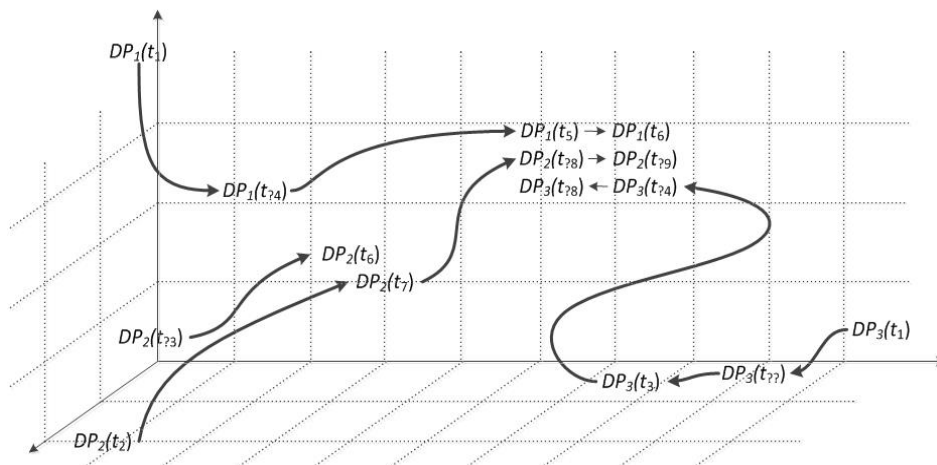
t_k - czas zakończenia sytuacji

Na podstawie tabeli 2 generowane są dwa plany: czasowy i przestrzenny. Na rysunku 2 pokazano czasowe uporządkowanie $DP_j(t_i)$ dla poszczególnych obiektów o_j . Natomiast na rysunku 3 pokazano przestrzenne uporządkowanie $DP_j(t_i)$ dla rozważanej sytuacji. Na podstawie śladów $TR(o_j)$ tworzony jest czasowo-przestrzenny obraz EI_R sytuacji.



Rys. 2. Czasowo uporządkowane punkty zmiany aktywności $DP_j(t_i)$ tworzące ślady obiektów $TR(o_j)$

Wizualizacja sytuacji jest zbiorem scen, które w sumie składają się na animację pokazującą ruchy obiektów o_1 , o_2 i o_3 oraz obiektu o_4 (pasywnego - nieożywionego, bez możliwości samodzielnego wykonywania akcji) zgodnie z pokazanymi na rysunku 3. ścieżkami.



Rys. 3. Przestrzenny obraz sytuacji EI_R

8. Podsumowanie

W pracy przedstawiono propozycję rozwoju systemu umożliwiającego generowanie wysokiej jakości animacji w oparciu o opis (scenariusz) sformułowany w języku naturalnym (wstępnie założono, że będzie to język polski i angielski). System z założenia ma być platformą integracyjną i testową dla różnych metod i technik z zakresu a) przetwarzania języka naturalnego, w szczególności analizy semantycznej, kontekstowej i pragmatycznej związanej z rozpoznawaniem i klasyfikacją zachowań i emocji, automatycznym rozpoznawaniem i śledzeniem wątków w tekście (być może z zastosowaniem elementów uczenia maszynowego) oraz porządkowaniem czasowym i przestrzennym zdarzeń (akcji) i b) grafiki komputerowej, ze szczególnym naciskiem na generowanie wysoce realistycznych animacji w czasie rzeczywistym oraz możliwie bliskiemu odzwierciedleniu naturalnych dla ludzi (i zwierząt) zachowań (np. mimiki twarzy). Najważniejszym zadaniem (funkcją), na którą szczególnie zwrócona będzie uwaga wykonawców, jest zdolność systemu do automatycznej kontroli i weryfikacji zgodności opisanych sytuacji (zdarzeń, akcji) z regułami i prawami zdefiniowanego świata. Przedstawiono również krótki przykład sposobu działania systemu w wersji podstawowej pokazujący proces analizy treści scenariusza i wygenerowanie na tej podstawie planów czasowego i przestrzennego sytuacji. Proponowane rozwiązanie, ze względu na bardzo szeroki zakres badań i zadań, jest zadaniem trudnym. Jednak biorąc pod uwagę istnienie podobnych, choć w części, aplikacji, wydaje się być możliwym do realizacji.

Literatura

1. Dupuy S., Egges A., Legendre V., and Nugues P., Generating a 3D Simulation of a Car Accident From a Written Description in Natural Language: The CarSim System. In Proc. of The Workshop on Temporal and Spatial Information Processing, 1-8, ACL 2001 Conference, Toulouse, 7 July.
2. Coyne, B and Sproat R., WordsEye: An Automatic Text-to-Scene Conversion System. Computer Graphics Annual Conference, SIGGRAPH 2001 Conference Proceedings, Los Angeles, 12-17 Aug., 487-496.

3. Kelleher, J., Doris T., Hussain Q. and Nualláin S. Ó., SONAS: Multimodal, Multi-user Interaction with a Modelled Environment. In *Spatial Cognition*, S. Ó Nualláin (Ed.), 171-184, Philadelphia, U.S.A.: John Benjamins B.V., 2000.
4. Perlin K. and Goldberg A., Improv: A system for scripting interactive actors in virtual worlds. In *Proceedings of SIGGRAPH 96*, New Orleans, LA, 1996, 205-216.
5. Cassell, J., Sullivan J., Prevost S., and Churchill E. (Eds.) *Embodied Conversational Agents*. Cambridge, MA: MIT Press, 2000.
6. Thórisson, K., *Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills*. Ph.D. thesis, Media Arts and Sciences, MIT Media Lab, Massachusetts Institute of Technology, U.S.A, 1996.
7. Okada, N., Integrating Vision, Motion and Language through Mind. In *Artificial Intelligence Review*, Vol. 10, Issues 3-4, August, 1996, 209-234.
8. Bobick, A., Intille S., Davis J., Baird F., Pinhanez C., Campbell L., Ivanov Y., Schtte A. and Wilson A., The KidsRoom: A Perceptually-Based Interactive and Immersive Story Environment. In *PRESENCE: Teleoperators and Virtual Environments*, 1996, 8(4): 367-391.
9. Loyall, A. B., *Believable agents: building interactive personalities*. Ph.D. thesis, CMUCS- 97-123, Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, 1997.
10. Liu, Z. and Leung, K.: Script visualization (ScriptViz): a smart system that makes writing fun. *Soft Computing*, 2006, 34-40.
11. Kapłon T., *Strategy of event reconstruction, Information systems architecture and technology (ISAT): models of decision making in the process of management in a risky environment*, Wrocław Oficyna Wydawnicza Politechniki Wrocławskiej, 2013.
12. Sekine S., *Corpus-based Parsing and Sublanguage Studies*, Ph.D. Dissertation, Department of Computer Science, New York University, 1998.
13. Kapłon T., *Generowanie reguł wnioskowania i ich klasyfikacja, Przetwarzanie języka naturalnego*, Praca zbiorowa pod redakcją Jana Kazimierczaka, WKiŁ, Warszawa 2005.
14. De Melo, C. and Paiva, A.: *Multimodal Expression in Virtual Humans*. *Computer Animation and Virtual Worlds*, 2006, 17 (3-4), 239-348.
15. Shim, H. and Kang, B. G.: CAMEO - Camera, Audio and Motion with Emotion Orchestration for Immersive Cinematography. In: *Proc. of the 2008 International Conference on Advances in Computer Entertainment Technology. ACE '08*. ACM, New York, NY. 352, 2008, 115-118.
16. Arijon D., *Gramatyka języka filmowego*, Wydawnictwo Wojciech Marzec, Warszawa, 2008.
17. Mierzwa J. *Modified Link Grammar in syntactic-semantic analysis of natural language texts*, *Polish Journal of Environmental Studies*, 2009, vol. 18, no 3B, 11-15.
18. Szymonik A., *Projektowanie systemów logistycznych w przedsiębiorstwie produkcyjnym*, IZIP Zakopane 2015, Wydawnictwo Polskiego Towarzystwa Zarządzania Produkcją, Opole 2015, Tom I 992 - 1003.
19. Kapłon T., *Model formalny obrazowania semantyki zdań języka naturalnego do wnioskowania przez komputer (praca doktorska)*, Wydawnictwo Politechniki Wrocławskiej, Wrocław 2003.

Dr inż. Tomasz KAPŁON

Katedra Informatyki Technicznej, Politechnika Wrocławska

50-370 Wrocław, ul. Janiszewskiego 11/17

tel./fax: (0-71) 320 27 45

e-mail: tomasz.kaplon@pwr.edu.pl