

**Politechnika Śląska**

**Wydział Automatyki, Elektroniki i Informatyki**

**Kierunek Informatyka**

##### Praca dyplomowa magisterska

###### Nawigacja postaci w wirtualnym świecie 3D

Autor: Radosław Bigaj

Kierujący pracą: dr Ewa Lach

Gliwice, wrzesień 2013

Spis treści

[1. Wstęp 4](#_Toc365544868)

[1.1. Geneza 4](#_Toc365544869)

[1.2. Cel pracy 4](#_Toc365544870)

[1.3. Przewodnik po pracy 4](#_Toc365544871)

[2. Sztuczna inteligencja 4](#_Toc365544872)

[2.1. Początki Sztucznej Inteligencji 5](#_Toc365544873)

[2.2. Przykłady zastosowań sztucznej inteligencji w grach 6](#_Toc365544874)

[2.3. Techniki symulacji sztucznej inteligencji 16](#_Toc365544875)

[3. Znajdowanie ścieżki 19](#_Toc365544876)

[3.1. Graf odnajdowania ścieżki 21](#_Toc365544877)

[3.2. Definicja grafu 22](#_Toc365544878)

[3.3. Grafy ważone 23](#_Toc365544879)

[3.4. Skierowane grafy ważone 25](#_Toc365544880)

[3.5. Terminologia 27](#_Toc365544881)

[3.6. Reprezentacja 27](#_Toc365544882)

[3.7. Algorytm Dijkstry 28](#_Toc365544883)

[3.7.1. Przedstawienie problemu 29](#_Toc365544884)

[3.7.2. Opis algorytmu 30](#_Toc365544885)

[3.8. Algorytm A\* 35](#_Toc365544886)

[3.8.1. Przedstawienie problemu 35](#_Toc365544887)

[3.8.2. Opis algorytmu 36](#_Toc365544888)

[4. Nawigacja postaci 40](#_Toc365544889)

[5. Koncepcja inteligentnego agenta 40](#_Toc365544890)

[5.1. Założenia projektowe gry 40](#_Toc365544891)

[5.2. Charakterystyka środowiska graficznego 41](#_Toc365544892)

[5.3. Charakterystyka silnika Unity3D 41](#_Toc365544893)

[6. Realizacja 44](#_Toc365544894)

[6.1. Konfiguracja biblioteki 45](#_Toc365544895)

[6.2. Implementacja 47](#_Toc365544896)

[6.3. Przegląd map i wygenerowanych grafów 48](#_Toc365544897)

[6.4. Efektywność algorytmów 51](#_Toc365544898)

[7. Podsumowanie i wnioski 51](#_Toc365544899)

# Wstęp

## Geneza

Proces tworzenia gier posiada własny specyficzny zbiór pojęć, własną pulę idiomów i wyzwań, którymi każdy projektant lub programista musi się zmagać. Można powiedzieć, że jest to jeden z powodów, dla których...

## Cel pracy

Celem pracy jest zapoznanie się z poszczególnymi technikami nawigacji oraz dokonanie ich analizy. W projekcie jest zawarte klika technik rozwiązywania problemu znajdowania ścieżki...

## Przewodnik po pracy

....

# Sztuczna inteligencja

Sztuczna Inteligencja jest istotnym elementem każdej gry wideo. Obszar zagadnień związanych ze Sztuczną Inteligencją istnieje, właściwie odkąd pojawiły pierwsze gry wideo (rok 1970), jednak same algorytmy pojawiły się już kilkanaście lat wcześniej. W literaturze często Sztuczną Inteligencje określa się skrótem AI (ang. *Artificial Intelligence* ) Sztuczna Inteligencja stosowana jest do [wsk]:

* Podniesienia realizmu świata gry. Stosowana głównie w grach typu cRPG (ang. *Computer Role Playing Games*). Ma za zadanie sterować zachowaniem agentów, z którymi zetknie się bohater gracza.
* Wsparcie podczas walki. Jest to najczęściej spotykana kategoria Sztucznej Inteligencji w grach komputerowych. Stosowana powszechnie w grach strategicznych oraz grach akcji. Sztuczna Inteligencja ma celu sterowanie agentami podczas walki.
* Relacjonowanie wydarzeń. Stosowane w grach sportowych. Sztuczna inteligencja pełni funkcje związane z trafnym komentowaniem zdarzeń zachodzących w świecie gry na podstawie bieżących działań gracza.

Obszar działania Sztucznej Inteligencji nie kończy się jednak tylko na symulowaniu inteligentnych zachowań, ale może również nadać agentom cechy ludzkie. Po implementacji takiej funkcjonalności w grze może sprawić, że osiągnie ona duży sukces. Efekt taki można uzyskać przenosząc do wirtualnego świata ludzkie niedoskonałości oraz tworząc sposób porozumiewania zawierający nieliniowe dialogi czy duży zasobów słów postaci.

## Początki Sztucznej Inteligencji

Pierwsza faza rozwoju gier komputerowych przykładała wagę do wyświetlanej grafiki, oczywistym tego powodem były ograniczenia czasu pracy procesora, a to grafika robiła największa wrażenie. Takie podejście skutkowało tym, że sztuczna inteligencja została zepchnięta na dalszy plan. Przykładowe implementacje zawierały sztywno zakodowane schematy zachowania oraz proste maszyny stanów. W dzisiejszych czasach większość operacji związanej z przetwarzaniem grafiki odbywa się w układach graficznych komputerów (GPU, ang. *graphic processing unit*), a wzrost jakości wyświetlanej grafiki nie przyciąga graczy, którzy wymagają czegoś więcej od gier. Dlatego producenci gier, aby spełnić żądania graczy kładą większy nacisk na rozwój Sztucznej Inteligencji [wsk].

Gra "Tennis for Two" jest jedną z pierwszych gier wideo. Jest to symulacja tenisa ziemnego, w której obraz jest wyświetlany za pomocą oscyloskopu. Została stworzona przez Williama Higinbothama w 1958 roku [wsk]. Pierwszą grą wideo stworzoną specjalne na komputer osobisty było "Space War". Gra została napisana przez S. Russel'a na minikomputer w 1962 roku. Obydwie te gry łączyło to, że wymagały dwóch graczy do rozgrywki. Dopiero w latach siedemdziesiątych zaczęto stosować, pewne proste ustalone schematy odpowiadające za poruszanie się obiektów, co można traktować jako początki Sztucznej Inteligencji.

Pierwszą grą, w której gracz posiadał przeciwników był "Pac-Man" wydany w 1979 roku. Agenci komputerowi sprawiali wrażenie inteligentnych - podczas pościgu za postacią gracza, na każdym z rozwidleń dróg agenci mieli różne szanse wyboru losowej drogi lub pogoni za graczem. W efekcie gracz miał odczucie, że komputerowi agenci współpracują ze sobą. Pac-Man zawierał implementacje prostej maszyny stanów, gdzie każdy z czterech agentów mógł gonić lub uciekać przed graczem w labiryncie. Do tej pory gry bazowały na prostych lub bardziej złożonych wzorcach, jak w klasycznych grach "Golden Axe" (1987 rok) czy "Super Mario Brothers" (1985 rok), gdzie przeciwnicy zwykle poruszali się w jednym lub dwóch kierunkach, aż do napotkania gracza[wsk].

Pierwszą grą akcji posiadającą Sztuczną Inteligencje jest "Goldeneye 007" (1997 rok). Pozwalała ona reagować na odpowiednio na ruch oraz akcję gracza. Komputerowi agenci posiadali zmysł wzroku i byli w stanie zauważyć czy pozostali agenci są martwi. Natomiast w grze "Thief: The Dark Project" (1998 rok) rozgrywka opierała się w znaczniej mierze na symulacji zmysłów wzroku i słuchu. Zostanie ona omówiona dokładniej w kolejnym rozdziale [wsk].

W latach 2001 i 2002 powstały dwie gry, które sprawiły, że gracze z niedowierzaniem patrzyli na poziom sztucznej inteligencji. Pierwsza z tych gier to "The Sims" ze studia Maxis, gdzie Sztuczna Inteligencja zajmowała się modelowaniem ludzkich emocji oraz potrzeb, przez co można powiedzieć, że gra była symulatorem życia. Drugą z gier jest "Black and White" ze studia Lionhead Studios, w tej grze komputerowy agent posiada umysł będący siecią neuronową przez, co mógł uczyć się w sztucznie stworzonym środowisku. Obecnie jednak większość gier wykorzystuje tylko podstawowe techniki związane ze Sztuczną Inteligencją. Powszechnie stosowaną techniką są maszyny stanów oraz jej pochodne.

## Przykłady zastosowań sztucznej inteligencji w grach

W tym rozdziale zostanie przedstawione i opisane kilka przełomowych gier, które dzięki wykorzystaniu Sztucznej Inteligencji odniosły sukces branżowy i stały się rozpoznawalnymi markami, a co więcej niektóre z przedstawionych tytułów są do dzisiaj rozwijane. Powodem dlaczego zostały wybrane akurat te pozycje jest uzasadniony tym, że nie są to nowe produkcje i przez to dostęp do bardziej szczegółowych informacji został ułatwiony. Sztuczna inteligencja nie jest, aż tak skomplikowanym system i głowie skupia się na implementacji kilku technik, w przeciwieństwie do najnowszych produkcji. Poniżej przedstawione gry znalazły się w rankingu serwisu "AiGameDev", w kategorii najbardziej innowacyjnych gier w historii. Dzięki osiągniętemu sukcesowi zapoczątkowały całą serię kolejnych wydań i kontynuacji, jest to jeden z kilku powodów, dla których warto się im przyjrzeć. Przedstawione tutaj gry są swego rodzaju pionierami w swojej klasie. Przyszło im się zmagać z wysokimi wymaganiami przed jakimi stawiał ich silnik Sztucznej Inteligencji, co więcej udało się im te wymagania spełnić, dzięki czemu poniższe tytuły odniosły sukces. Zostanie teraz przedstawione kilka gier z wyżej opisanego rankingu, wybranych głowie dzięki dostępności materiałów opisujących implementacje Sztucznej Inteligencji.

**Thief**

"Thief: The Dark Project" jest grą skardankową, w której gracz obserwuje świat z perspektywy pierwszej osoby. Fabuła gry przedstawia historię średniowiecznego złodzieja. Często gry, w których gracz widzi świat z pierwszej osoby kojarzą się ze strzelaninami i grami akcji, w tym przypadku jest jednak inaczej. W tej grze głównym celem gracza się skradanie się, unikanie pułapek oraz wykonywanie serii kradzieży. Głównym elementem tego typu rozgrywki jest system czujników, poprzez które wirtualni agenci w grze są wstanie odbierać bodźce z otoczenia w podobny sposób jak odbiera je gracz. Moduł sztucznej inteligencji opera się na dużej liczbie stanów, a nie tylko czy gracz został dostrzeżony czy też nie. Co więcej, działa on też w stosunku do obiektów innych niż gracz, np. w przypadku pozostawionych zwłok przeciwnika, które gracz powinien schować, by straż ich nie odnalazła. Znalezienie zwłok powoduje wszczęcie alarmu, co powoduje, że strażnicy będą dokładnie patrolować okolicę przez co gracz może zostać złapany. Każde takie zachowanie powinno być zrozumiałe dla gracza, dzięki czemu jest on w stanie przewidzieć zachowanie agentów komputerowych. [wsk]



Ilustracja Straż zamku w grze Thief

Nowatorskie koncepcje wykorzystane w grze:

* Moduł odpowiedzialny za sztuczną inteligencje zbudowany jest z całego systemu czujników, dzięki któremu agenci mogą realistycznie reagować na bodźce świetlne i dźwiękowe.
* Agenci znajdujący się pod kontrolą sztucznej inteligencji korzystają z specjalnych nagrań audio, w celu oznajmienia swojego obecnego stanu. Pozwala to graczowi na zorientowanie się w jakieś sytuacji się znajduje.



Ilustracja Przykład działania zmysłów w grze Thief

Charakterystyka modułu sztucznej inteligencji:

Moduł ten zbudowany jest na systemie czujników odpowiedzialnych za wizję i dźwięk - zbiera on informacje z otoczenia w świecie gry. Z pomocą tych informacji agenci w grze są w stanie podejmować decyzje. W skład tej informacji wchodzi stan świadomości sztucznej inteligencji - są to wartości dyskretne opisujące stan wiedzy modułu na temat obecnej sytuacji w grze, położenia oraz tożsamości obiektu zainteresowania. Poziom widoczności w grze opisany jest przez światło, widoczność na tle innych obiektów, ruch oraz wielkość. Zmysł wzroku jest opisany za pomocą trzech wymiarów. Bazując za poziomie świadomości, Sztuczna Inteligencja określa odpowiednie działa w procesie podejmowania decyzji. Proces taki bazuje na zasadzie działania maszyny stanów [wsk].

**The Sims**

Grę "The Sims" można przedstawić jako symulator życia człowieka. Zadaniem gracza jest zaprojektowanie domostwa oraz jego utrzymanie. W czasie gry gracz będzie musiał się zmagać z problemami życia codziennego oraz wywiązywać się z różnego rodzaju obowiązków. Moduł Sztucznej Inteligencji zajmuje się modelowaniem agentów zbliżonych do ludzi, którzy posiadają własną osobowość, potrzeby, umiejętności oraz relacji z innymi agentami. Dzięki temu agencji mogą wpływać wzajemnie na siebie, swoje zachowanie oraz zaspokajać swoje potrzeby. Gra odniosła duży sukces i obecnie na rok 2014 przewidywana się premiera czwartej części serii.



Ilustracja Zrzut ekranu z gry The Sims

Nowatorskie koncepcje wykorzystane w grze [wsk]:

* Dokonano zamodelowania wirtualnej emocjonalnej więzi między agentami, dzięki czemu jest możliwe tworzenie związków między nimi
* Każdy agent w grze ma swój zdefiniowany charakter, umiejętności, podstawowe potrzeby emocjonalne oraz fizyczne, mające wpływ na jego poczynania w grze. Emocje postaci są mierzone w zakresie (-100, 100), a następnie są mapowane do wyjściowej formy szczęścia/nastroju.
* W grze zastosowano inteligentne obiekty, co okazało się pomocne w implementacji zachowań. To obiekt definiuje, w jak agent może wejść z nim w interakcje.

Charakterystyka modułu sztucznej inteligencji:

Model osobowości agenta w grze "The Sims" jest podzielony na 4 główne kategorie: osobowość, umiejętności, potrzeby oraz relacje z innymi agentami w grze. Osobowość jest opisana wskaźnikiem określającym każdą z pięciu podkategorii: niechlujny lub schludny, nieśmiały lub towarzyski, poważny lub żartobliwy, leniwy lub aktywny, złośliwy lub miły. Osobowość ma wpływ jakie czynności będzie wykonywała postać oraz jak te czynności będą na nią wpływać. Kategorię potrzeb podzielono na dwie grupy. Umysłowe: energia, towarzystwo, zabawa, pomieszczenie oraz fizyczne: głód, komfort, pęcherz, higiena. Każdą z potrzeb definiuje się przez wartość z zakresu <-100, 100>. Do poznania obecnego nastroju agenta wszystkie potrzeby są modyfikowane przez odpowiadająca im wartość funkcji wagi.



Ilustracja Funkcje wag potrzeb agenta

Przykładowo niski poziom higieny agenta ma bardzo wysoki wpływ na ogólny nastrój, ale wysoki poziom wpływa na ogólny nastrój już nieznacznie - widzimy tendencje spadkową. Można zauważyć, że skoro zakresy wszystkich potrzeb oraz nastroju są takie same, to dokonując ich sumowania można otrzymać ogólny poziom nastroju:

gdzie:

- waga odpowiadającej potrzebie,

- wartość potrzeby.

Jedną z innowacyjnych rzeczy jaką wprowadzono do gry są inteligentne obiekty. Obiekty te zajmują się rozpowszechnianiem informacji z wszystkimi możliwymi interakcjami w jakie można wjeść z tym obiektem oraz w jaki sposób te akcje wpłyną na agentów. Informacje te są rozprowadzane do wszystkich postaci. Załóżmy, że agent posiada ustalone parametry i jego nastrój wyjściowy wynosi +18. I teraz dostaje informacje od obiektu toalety, że gdy z niej skorzysta, zaspokoi pęcherz oraz gdy ją wyczyści to poprawi atmosferę w pokoju, co daje wynikową poprawę nastroju to liczby +26. W innym przypadku, gdyby użył wanny poprawiłby komfort oraz higienę i ogólna poprawa nastroju wynosiłaby +20. W całym procesie decyzyjnym największy wpływ na nastrój ma pęcherz, więc jako następny obiekt do użycia agent wybierze toaletę. Moduł Sztucznej Inteligencji podejmuje decyzje na podstawie listy możliwych akcji, które da się wykonać na danym obiekcie. Zachowanie agentów głównie bazuje na skryptach, jakie udostępniają obiekty dotyczące możliwych interakcji. Przykładowo: skrypt odnoszący sie do lodówki opisujący cały proces przygotowania posiłku [wsk].

**Halo**

"Halo: Combat Evolved" jest grą akcji, w której gracz patrzy na wirtualny świat z perspektywy pierwszej osoby (FPS, ang. *first-persion shooter* ). Gracz wciela się w postać jednego z elitarnych żołnierzy i wyrusza na wojnę z obcą rasą. W trakcie rozgrywki postać gracza może wykorzystać wiele elementów terenu takich jak zagłębienia, skały, drzewa oraz przemieszczać się przy pomocy różnych pojazdów wodnych, lądowych i powietrznych.

Nowatorskie koncepcje wykorzystane w grze [wsk]:

* Inteligentni agenci potrafiący się kryć przez ostrzałem oraz używać rozważenie dostępnej broni,
* Wykorzystanie drzewa zachowań, które zostało bardzo dobrze przyjęte przez przemysł gier komputerowych,
* Sytuacja na polu bitwy ma wpływ na zachowanie jednostek.

Charakterystyka modułu sztucznej inteligencji:

Decyzje w Halo podejmowane za pomocą drzewa zachowań. Pozwala ono na kontrolowanie liczby z zachowań, w takiej architekturze wybór odpowiedniego węzła jest ustalane na podstawie jego aktywności oraz priorytetu. Proces taki można opisać w kilku krokach, najpierw węzły dokonują obliczeń na podstawie informacji zebranych ze świata gry. Wynikiem tego jest ustalenie, które węzły są aktywne, a które nie. Wybór potomka pada na aktywny węzeł posiadający najwyższy priorytet, po zakończeniu działania następny potomek w kolejności je zacznie. W trakcie działania może, jednak wystąpić taka sytuacja, że zostaje aktywowany potomek o wyższym priorytecie, który przerwie wykonywanie aktualnego procesu. Każdy węzeł w takiej architekturze jest konkretnym zachowaniem np. atakuj, wycofaj się. Co więcej istnieją też referencje do zachowań zwane impulsami. Dzięki nim możliwe są działania zależne od kontekstu.

Sprawdzanie aktywności zachowań w drzewie odbywa się bardzo często. Są jednak zdarzenia, która pojawiają się stosunkowo rzadko i warto zadbać, aby czas na ich testy nie był nie potrzebnie tracony. Zakładając istnienie impulsu "Ucieczka, gdy dowódca zginie" jest on zależy od wydarzenia "Śmierć dowódcy w grze". W procesie aktualizacji muszą zostać wykonane testy czy ten kto zginął był dowódcą oraz czy nie ma innych dowódców. W przypadku pozytywnego wyniku testów, referencja uruchamia zachowanie ucieczki. Testy tego typu są uruchamiane w bardzo krótkich odstępach czasu. Trzeba zwrócić uwagę, że drzewo zachowań musi działać płynnie, a zdarzenia opisane powyżej pojawiają się bardzo rzadko. Twórcy rozwiązali ten problemu przez aktualizację struktury drzewa w odpowiednim momencie czasu, tak aby obsłużyć takie przypadki jak śmierć generała [wsk].



Ilustracja Przykład skierowanego grafu acyklicznego zachowań

**F.E.A.R.**

„*F.E.A.R. First Encounter Assault*” jest grą akcji, w której celem gracza jest zwalczanie terroryzmu. Gracz wykonuje misje w pięcioosobowej drużynie.

Nowatorskie koncepcje wykorzystane w grze [wsk]:

* Pierwszy raz użyto systemu planowania zadań do generowanie zachowań zależnych od sytuacji.
* Agenci komputerowi sprawnie wykorzystują świat gry tak, aby zwiększyć jego realizm. Otwierają drzwi, znajdują osłony, przechodzą przez okna,
* Dokonano implementacji taktycznych technik walki - atak z flanki, przerywane serie ostrzału.

Charakterystyka modułu sztucznej inteligencji:

W grze dokonano implementacji maszyny stanów. Posiada ona tylko dwa stany przyjmujące różną liczbę parametrów - dzięki temu jest w stanie przetwarzać wiele zdarzeń. Zastosowano w grze również system planowania STRIPS (ang*. Stanford Research Institute Problem Solver*). System ten przetwarza związki pomiędzy czynnościami i informuje w jaki sposób je zrealizować.

Implementacja agenta składa się z pamięci roboczej, czujników oraz kilku modułów zarządzających nawigacją, animacją oraz bronią. Za detekcje zmian w świecie gry odpowiadają czujniki reagujące na bodźce wizualne i dźwiękowe pobierane z systemu zdarzeń. Informacje mogą być też dostarczane bezpośrednio ze świata gry np. czy gdzieś w pobliżu znajduje się miejsce, w którym można się schronić. Pozyskane dane są składowane w pamięci roboczej. Moduł zarządzający planowaniem wykorzystuje dane z pamięci roboczej do podejmowania decyzji za pośrednictwem komunikatów do konkretnych podmodułów.

**Black & White**

"Black & White" jest grą, w której gracz wciela się w rolę boga. Zamieszkuje on różne krainy wraz z cywilizacjami różniącymi się wierzeniami, mocami czy hobby. W grze oczywiście znajdują się też inni bogowie z którymi musimy konkurować, tak aby przekonać do siebie jak największą liczbę mieszkańców. Konfrontacja pomiędzy bogami polega na tym, odpowiednim wytrenowaniu stworzenia (wilka, tygrysa, niedźwiedzia itp.). Na początku gracz wybiera jedno ze stworzeń, jest ono sterowne przez Sztuczną Inteligencje. Gracz może szkolić takie stworzenie, które uczy poprzez obserwowanie zachowania gracza oraz wykorzystanie system kar i nagród.



Ilustracja Osada wraz z trenowanym stworzeniem w grze Black & White

Nowatorskie koncepcje wykorzystane w grze:

* Zastosowanie technik, takich jak drzewa decyzyjne czy sztuczne sieci neuronowe,
* Silnik korzystający z architektury BDI (Przekonanie-Pragnienie-Zamiar ang. *Belief-Desire-Intention*),
* Gra symuluje nie tylko zachowanie stworzenia, ale też życie mieszkańców pracujących tak, aby rozwijać zamieszkiwaną wioskę.

Charakterystyka modułu sztucznej inteligencji:

Drzewa decyzyjne reprezentują przekonania agenta na temat ogólnych typów obiektów, natomiast sieci neuronowe reprezentują jego pragnienia. Agent w grze może uczyć się na parę różnych sposobów, poprzez obserwowanie gracza, mieszkańców wyspy lub inny stworzeń. Może również testować różne rzeczy, a wiedza opisana w grze na temat danych obiektów skoryguje jego kolejne zachowania odnośnie tych obiektów, np. jego pragnieniem może być zaspokojenie głodu, więc spróbuje zjeść napotkany kamień, jednak z opisu w grze wynika, że nie nadaje się on do jedzenia. Tym samym zaktualizuje swoje drzewo decyzyjne dotyczące głodu i już nie będzie próbował jeść kamienia.

Kolejnym sposobem nauki jest nagradzanie lub karcenie agenta, przez gracza za wykonane czynności. Oto przykład drzewa decyzyjnego jakie zostanie zbudowane po ataku na kilka wiosek oraz odpowiedniego odniesienia się do agenta. Do utworzenia drzewa wykorzystano algorytm ID3.

## Techniki symulacji sztucznej inteligencji

Wyróżnia się wiele technik stosowanych do symulacji sztucznej inteligencji w grach komputerowych. Poniżej zostaną omówione najpopularniejsze z nich, są one powszechnie stosowane w branży gier komputerowych do gier wykorzystujących Sztuczną Inteligencję. Popularność danej techniki wynika z tego, że najprostsze rozwiązania okazują się najlepszymi, dzięki czemu techniki takie jak: automaty skończone, heurystyczne poszukiwanie drogi czy drzewa decyzyjne zyskały sobie taką popularność.

**Automaty skończone.**

Technika automatów stanów skończonych była wykorzystywana już w latach 90. jej funkcją było kontrolowanie wirtualnych agentów. Automaty stały się tak popularne i użyteczne, że są stosowane do zarządzania AI, również w najnowszych wysokobudżetowych produkcjach. Wykorzystuje się je też w komputerowych grach fabularnych (cRPG - ang. *computer Role Playing Game*) do sterowania dialogami gracza z agentami. Co więcej zarządzają obiektu w grze, przechowują stan rozgrywki (np. zwycięstwo, porażka, wykonane zadanie, postać dotarła do punktu docelowego), przetwarzają komendy gracza oraz zarządzają światem gry.

Automat stanów skończonych zbudowany jest z pewnej ściśle określonej liczby stanów znajdujących w danej puli rozwiązań. Kolejno zostają przechwycone pewne zdarzenia, które zmieniają stan maszyny. Dzięki temu istnieje możliwość podjęcia jednego lub kilku działań w zależności od stanu w jakim się aktualnie znajduje obiekt gry.

**Heurystyczne poszukiwanie drogi.**

Jednym z problemów jaki rozwiązuje Sztuczna Inteligencja jest określenie najlepszej drogi z punktu A do punktu B na terenie rozgrywki. Technika ta jest stosowana jest do rozwiązywania zagadnień bardzo skomplikowanych i złożonych takich jak poruszanie się jednostek w formacjach czy planowanie strategiczne. Rozwiązaniem jakie stosuje się dla problemów tego typu jest heurystyczny algorytm A\*. Algorytm ten podczas procesu określania drogi do celu nie szuka jej "na ślepo", tylko szacuje jej najbardziej prawdopodobny kierunek odrzucając inne mniej sensowne ścieżki.

Celem algorytm A\* jest minimalizacja obszaru poszukiwań najlepszej trasy, dzięki ustaleniu pewnego kierunku, który zawęża obszar rozważanych tras. Technika ta oblicza koszt dotarcia do punktu na mapie i dodaje do niego heurystykę określającą przewidywane koszty dotarcia do celu. Heurystyka jest liczona zwykle jako odległość od obecnego punktu do celu ignorując wszelkie przeszkody i ograniczenia umieszczone na mapie. W skrócie A\* sprawdza, po każdym wykonanym ruchu agenta wszystkie możliwe kierunki dalszej trasy i ponownie wybiera możliwe kierunki trasy o jak najniższym koszcie. W momencie, gdy rozważane położenie jest celem, algorytm kończy swoje działanie. W przeciwnym przypadku algorytm przechowuje przyległe położenie, tak aby w przyszłości móc rozważyć inne ścieżki.

**Drzewa decyzyjne.**

Drzewa decyzyjne zostały zaprojektowane z myślą o rozwiązywaniu problemów decyzyjnych oraz tworzeniu planu działania. Są powszechnie stosowane w grach takich jak: szachy czy warcaby.

Drzewo decyzyjne jest reprezentowane w postaci grafu decyzji i możliwych konsekwencji. Węzły odzwierciedlają stan gry, natomiast potomkowie to stany uzyskane po przeprowadzeniu ruchu. Węzły decyzyjne i węzły konsekwencji występują naprzemiennie. Każda ścieżka w takim drzewie kończy się węzłem końcowym. Agent komputerowy dokonuje analizy drzewa rozpatrując tym samym wszystkie możliwe posunięcia dostępne w aktualnej sytuacji. Następnie wybiera te, które uznał za najbardziej korzystne.

**Logika rozmyta.**

Pojęcie logiki rozmytej ma związek z teorią prawdopodobieństwa oraz teorią zbiorów rozmytych. Można ją przedstawić jako stany pośrednie pomiędzy wartościami logicznymi (prawda, fałsz) , które określają przynależność do odpowiedniego zbioru. Daje to możliwość rozważenia odpowiedzi na pytanie "jak bardzo?", "ile?" przykładowo: "legion", "wataha", "grupa", "sporo", "kilka". Logia rozmyta wykorzystywana jest często do odwzorowania emocji w grach (np. "przyjacielski", "obojętny", "wrogi"). Za pomocą logiki jak pokazano na przykładach można modelować sferę uczuciową agentów komputerowych, co poprawia realizm gry. Kolejnym przykładem zastosowania jest sterowanie nieograniczonymi zachowaniami takimi jak: poruszające sie chmury, drżenie liści.

**Sztuczne sieci neuronowe.**

Sztuczne sieci neuronowe zostały zaprojektowane, aby działać podobnie jak sieci neuronowe w mózgu człowieka. Przetwarzają one sygnały oraz wykonują obliczenia za pomocą neuronów - są to elementy, które wykonują pewne operacje na wejściu. Za pomocą sieci neuronowej agent może się uczyć wraz z postępem gry. Natrafiając na nowy rodzaj sytuacji, dostosowuje się do niej korzystając ze zdobytego doświadczenia.

Sieci neuronowe od wielu lat zamierzano przystosować do tworzenia Sztucznej Inteligencji w grach komputerowych. W 2000 roku swoją premierę miała gra Colin McRae Rally 2.0 - gra będąca symulatorem wyścigów. Gra ta zawierała implementacje sieci neuronowej. Za dane wejściowe przyjmowała ona parametry opisujące trasę jaką miał przejechać agent przykładowo: krzywizna łuku drogi, rodzaj gruntu, parametry techniczne pojazdu, stan pogody. Zadaniem tej sieci było wygenerowanie odpowiednich danych wyjściowych bazując na parametrach wejściowych tak, aby samochód kierowany przez agenta mógł bez problemu przejechać trasę wyścigu.

**Algorytm stadny.**

W 1987 roku Craig Raynolds przedstawił artykuł, w którym opracował 3 zasady, które w połączeniu umożliwiały grupie agentów realistyczne zbiorowe zachowanie podobne do zachowań stadnych znanych ze świata zwierząt np. ławic ryb, stad ptaków. Raynolds określił te trzy zasady jako sterownie zachowaniem. Prezentują się one następująco:

* spójność - sterownie, którego celem jest zbieranie agentów znajdujących się blisko siebie w odpowiednie grupy lokalne,
* wyrównywanie - rodzaj sterowania dzięki któremu agent może dostosowywać kierunek i prędkość do innych agentów przebywających w pobliżu,
* rozdzielczość - sterownie, w które zapobiega tworzeniu się tłumu w jednym miejscu. Agenci zachowują pewną odległość względem siebie.

Reynolds opracował jeszcze czwartą zasadę, określaną mianem unikania. Jest ona stosowana, aby wirtualni agenci unikali przeszkód umieszczonych na mapie.

W każdym cyklu procesu przemieszczania się agenci każdorazowo sprawdzają środowisko, w jakim w danej chwili przebywają i to jest jedyna informacja jakiej wymaga ten algorytm. Powoduje to, że zmniejszenie wymagań związanych z pamięcią przy sterowaniu wieloma agentami oraz pozwala na szybką reakcję na ewentualną zmianę sytuacji. Podsumowują algorytm ten umożliwia nadanie grupie poruszających się agentów dynamiki ruchu jak jedno ciało oraz zdolności omijania przeszkód czy wrogich postaci.

**Dalszy rozwój sztucznej inteligencji w grach komputerowych**

Wraz z upływem czasu rola sztucznej inteligencji w grach zwiększyła się o analizowanie gry i jej dostosowanie do poziomu gracza. Powoduje ona rozwój świata wraz ze wzrostem doświadczenia gracza, które uzyskał w dotychczasowym procesie rozgrywki.

Istnieją pewne grupy badawcze, zajmujące się projektami gier, w których cały świat ma być kontrolowany przez Sztuczną Inteligencję. Gracze wchodzący do gry mają czuć się, tak jakby wchodzili do realnej rzeczywistości, gdzie każde ich działanie ma wpływ na dalszy przebieg gry.

Analizując obecny stan sztucznej inteligencji, w którym to twórcy rozwijają każdy element świata, tak aby wydawał się jak najbardziej realny, można pokusić się o stwierdzenie, że w przyszłości Sztuczna Inteligencja będzie odpowiedzialna za fizjologię i psychologię postaci gracza. Dzięki temu gracz będzie musiał bardziej zaangażować sie w grę.

# Znajdowanie ścieżki

Postacie w grach komputerowych muszą się poruszać po mapie. Czasami zdarza ten ruch jest na stałe ustawiony przez programistę. Przykładowo strażnik patrolujący drogę, porusza się ślepo po ogrodzonym terenie. Stałe trasy są łatwe do implementacji i wdrożenia, natomiast bardzo łatwo można spowodować, że obiekt zostanie przesunięty przez inny obiekt (wejdzie z nim w kolizje), co spowoduje, że wypadnie z trasy. Pozwolenie postaci na pewną dowolność w przemieszczaniu może spowodować, że jej wędrówki będą bezcelowe, co więcej postać będzie mogła łatwo utknąć. Bardziej zaawansowane postacie nie wiedzą z góry, gdzie będą musiały się przemieścić. Jednostka wykorzystywana w strategii czasu rzeczywistego może zostać przypisana do dowolnego punktu na mapie przez gracza w dowolnym momencie czasu, patrolujący strażnik w grze o tematyce skradanki może potrzebować przemieścić się do najbliższego punktu alarmowego, żeby wezwać wsparcie, a w grach platformowych może być wymagane, żeby przeciwnicy gonili gracza do przepaści używając dostępnych platform.

Dla każdej z tych postaci musi zostać obliczona odpowiednia droga przez dany poziom, żeby dostać się tam gdzie jest jej cel. Najlepszym wyjściem jest utworzenie sensownej trasy w jak najkrótszym czasie.

To właśnie jest istotą odnajdywania ścieżki (ang. Pathfinding), czasami nazywane także planowaniem ścieżki - znajduję się każdym silniku gry posiadającym moduł odpowiedzialny za sztuczną inteligencję. W przedstawionym poniżej modelu (*Ilustracja 7*) rola odnajdywania ścieżki znajduje się pomiędzy modułami odpowiedzialnymi za podejmowanie decyzji oraz poruszanie się postaci. Często odnajdywanie ścieżki jest po prostu używane do wykonania wstępnej analizy gdzie postać ma się przesunąć, aby dotrzeć do celu. Sam cel jest wyznaczany inną część modułu sztucznej inteligencji, więc można podsumować, że odnajdywanie ścieżki oblicza nam tylko jak dostać się do celu. Do uzyskania pożądanego efektu trzeba zbudować system przemieszczania w taki sposób, aby był wywoływany, kiedy jest potrzeba zaplanowania drogi. Zostanie on omówiony w kolejnych rozdziałach.

Moduł odnajdywania ścieżki może, również zostać umieszczony na siedzeniu kierowcy i zarządzać podejmowaniem decyzji, gdzie się trzeba przemieścić, a także w jaki sposób się dostać do celu. Jest to pewnego rodzaju odmiana modułu odnajdywania ścieżki zwana odnajdywaniem ścieżki otwartego celu (ang. *open goal pathfinding*), może być użyta do pracy zarówna nad ścieżką jaki i miejscem przeznaczenia.



Ilustracja Model sztucznej inteligencji

Większość gier używa funkcjonalności znajdowania ścieżki wykorzystując algorytm A\*(ang. *A star*). Pomimo, że algorytm jest efektywny i łatwy do wdrożenia, to nie może on pracować bezpośrednio na danych zaczerpniętych z poziomu (mapy) gry. Wymaga to, aby poziom gry był reprezentowany przez odpowiednią strukturę danych - jest to zwykle graf ważony o nieujemnych wagach. Ten rozdział wprowadza pojęcie struktury danych grafu, kolejno zostanie omówiony algorytm Dijkstry, pomimo że jest on częściej stosowany w procesie podejmowania decyzji taktycznych niż w procesie odnajdowania ścieżki. Ponieważ struktura danych grafu nie jest jedynym sposobem w jaki większość gier reprezentuje swoje dane, dlatego warto się przyjrzeć się kwestii zmiany geometrii mapy na dane, które mogą zostać przetworzone przez moduł odpowiedzialny na znajdowanie ścieżki. Warto też wspomnieć o wielu dziesiątkach przydatnych wariacji podstawowego algorytmu A\*.

Po raz pierwszy w powyższym akapicie wspomniano o efektywności. Będzie ona rozumiana jako efektywność czasowa algorytmu.

## Graf odnajdowania ścieżki

Algorytm A\* oraz Dijkstra nie może pracować bezpośrednio na geometrii, z której zbudowana jest mapa. Jednak każda mapa może zostać poddana pewnemu procesowi w wyniku, którego otrzymamy jej uproszczoną wersje w postaci grafu. Jeśli proces upraszczania geometrii do grafu wykona się prawidłowo, to plan ścieżki zwrócony przez moduł odpowiedzialny za odnajdowanie ścieżki może zostać użyty oraz przetłumaczony ponownie na warunki gry. Z drugiej strony sam proces upraszczania może pozbawić pewnych informacji, które mogą się okazać znaczące. Złe uproszczenie może oznaczać, że ostateczna trasa nie jest za dobra.

Algorytmy stosowane do odnajdywania ścieżki używają zwykle struktur zwanych skierowanymi grafami o nieujemnych wagach. Dla uproszczenie opisu całego grafu odnajdywania ścieżki jest on wyrażony przez proste struktury grafowe.

## Definicja grafu

Graf jest matematyczną strukturą często reprezentowaną przez schemat graficzny. Graf składa się z dwóch rodzajów elementów. Są to węzły często rysowane jako punkty lub koła w schemacie grafu oraz krawędzie będące połączeniami węzłów przedstawiane w postaci linii. przedstawia strukturę grafu.

Formalnie graf składa się ze zbioru węzłów i zestawu połączeń, w którym połączenie jest po prostu nieuporządkowaną parą węzłów.

Każdy węzeł stanowi zwykle pewien region poziomu gry, taki jak pokój, piwnica czy schody lub mały region miejsca na zewnątrz. Połączenia pokazują, które miejsca są połączone. Jeśli pokój sąsiaduje ze schodami, to węzeł reprezentujący pokój będzie miał połączenie z węzłem reprezentującym schody. W ten sposób cały poziom gry jest podzielony na obszary, które są ze sobą połączone. Aby dostać się z jednego miejsca na danym poziomie do drugiego możemy korzystać z połączeń. Jeśli jest możliwość przejścia bezpośrednio z węzła startowego do celu, to problem jest trywialny. W przeciwnym razie możemy użyć połączeń do podróży przez węzły pośrednie znajdujące się na ścieżce.

Droga przez graf składa się z zera lub więcej połączeń. Jeśli początek i koniec są takie same, to nie ma połączenia w ścieżce. Jeśli węzły są połączone, wówczas tylko jedno połączenie jest potrzebne.



Ilustracja Graf

## Grafy ważone

Ważony graf składa się z węzłów i połączeń podobnie jest zwykły graf. Dodatkowo posiada wartość liczbową dla każdego połączenia węzłów. W matematycznej teorii grafów jest to nazywane wagą, natomiast w zastosowaniach stosowanych w grach powierzchnie jest określany mianem kosztu (chociaż graf nadal jest nazywany grafem ważonym, a nie grafem kosztu). Na *ilustracji 9* przedstawiającej graf z każdym połączeniem jest związana wartość kosztu.

Koszty w module zarządzającym odnajdowaniem ścieżki są zwykle reprezentowane czasem lub odległością. Jeśli węzeł reprezentujący platformę jest położony w dużej odległości od węzła reprezentującego następną platformę, to koszt takiego połączenia będzie duży. Podobnie będzie wyglądało to w przypadku przemieszczania się pomiędzy dwoma pokojami, które są pokryte pułapkami - taka podróż będzie trwała długo przez co koszt będzie duży. Koszty w grafie mogą reprezentować więcej niż tylko czas i odległość. Istnieje duża liczba aplikacji z odnajdywaniem ścieżki, w których koszt stanowi kombinacje czasu, odległości i innych współczynników.



Ilustracja Graf posiadający koszty

Na całej trasie przez graf, od węzła początkowego do węzła docelowego, możemy obliczyć całkowity koszt ścieżki. Jest to po prostu suma kosztów każdego połączenia na trasie. Na *ilustracji 10* zaprezentowano przykładowy graf ważony, na którym zostanie policzony koszt przykładowej trasy. Zakładając trasę z punktu A do punktu C przez węzeł B. Koszt jest obliczany następująco: z A do B wynosi 4, następnie z B do C ma wartość 5 to całkowity koszt drogi będzie wynosił 9.



Ilustracja Graf z rozważaną drogą A-B-C

**Reprezentacja punktów w regionie**

Można od razu zauważyć, że jeśli dwa regiony są ze sobą połączone (np. pokój i schody), to odległość pomiędzy nimi będzie wynosić zero. Jeśli gracz stoi w drzwiach, a następnie przemieszcza się do schodów natychmiastowo. Nasuwa się więc pytanie czy zatem wszystkie połączenie mają koszt równy zero? Istnieje tendencja do pomiaru połączenia ze względu na odległości czy czas przez punkt reprezentatywny znajdujący sie w każdym regionie. Zatem dla przykładu punkt reprezentatywny dla pokoju będzie się w jego środku, a dla schodów w ich centrum. Jeśli pokój jest duży a schody są długie to jest całkiem prawdopodobne że odległość między punktami reprezentatywnymi będzie duża, co za tym idzie koszt również będzie duży. Często można spotkać diagramy grafów odnajdywania ścieżek. Przykładem jest Ilustracja 11, gdzie punkt reprezentatywny jest przyporządkowany do każdego regionu.



Ilustracja Węzły grafu określające regiony

**Ograniczenia odnośnie wag w grafie.**

Posiadanie przez krawędź ujemnego kosztu, może wydawać się całkowicie nieuzasadnione. Nie można mieć ujemnego dystansu między węzłami oraz nie można mieć ujemnego czasu, aby dostać się do danego miejsca. Mimo to matematyczna teoria grafów dopuszcza ujemne wagi i mają one bezpośrednie zastosowanie, w niektórych praktycznych problemach. Jednak problemy te znajdują się całkowicie poza zasięgiem świata gier komputerowych i nie będę omawiane na łamach tej pracy. Pisanie algorytmów, które mogą pracować z ujemnymi wagami jest zazwyczaj bardziej złożone, niż dla tych, które mają sztywne wymogi stosowania nieujemnych wag. W szczególności, Dijkstra i algorytmy A\* powinny być stosowane wyłącznie z nieujemnymi wagami. Istnieje taka konstrukcja grafu posiadającego ujemne wagi, że algorytm zajmujący się procesem odnajdywania ścieżki zwróci rozsądny wynik. Jednak w większości przypadków, Dijkstra i A\* wejdzie w pętlę nieskończoną. I nie jest to błąd algorytmów. Z punktu widzenia matematycznego nie ma czegoś takiego jak najkrótsza ścieżka przez wiele grafów w ujemnymi wagami - takie rozwiązanie po prostu nie istnieje. Używając terminu "koszt" w tej pracy brana jest pod uwaga wyłącznie nieujemna waga. Kosz jest zawsze liczbą dodatnią. Twórcy gier wspólnie przyznają, że nigdy nie stosowali ujemnych wag ani algorytmów do nich przystosowanych w procesie tworzenia gier komputerowych.

## Skierowane grafy ważone

W wielu sytuacjach ważony graf wystarczy do reprezentacji poziomu gry i często zdarzają się implementacje grafu w takiej formie. Można jednak pójść o krok dalej. Główne algorytmy służące do odnajdywania ścieżki obsługują bardziej złożone formy grafów, takie jak graf skierowany (Ilustracja 12). Jest on często używany przez programistów gier komputerowych. Możliwe jest, aby poruszać się pomiędzy węzłem A i węzłem B (przykładowo pokój i schody), to możliwe jest też, aby przejść z węzła B do węzła A. Połączenia są dostępne w obie strony, a koszt przejścia jest taki sam w obu kierunkach. Skierowany graf zakłada, że połączenia są dostępne tylko w jednym kierunku. Jeśli postać gracza może dotrzeć z węzła A do węzła B i odwrotnie to będzie to reprezentowane na grafie jako dwa połączenia: jedno z A do B i drugie z B do A. Jest to przydatne w wielu sytuacjach. Po pierwsze, nie jest zawsze tak, że możliwość przejścia z punktu A do B oznacza że w drugą stronę jest to osiągalne. Jeśli węzeł oznacza podłogę na piętrze, a B reprezentują podłogę magazynu znajdującego się pod pokojem, to postać może łatwo spaść z A do B, ale nie będzie w stanie wrócić ponownie.



Ilustracja Graf skierowany ważony

Po drugie posiadając dwa połączenia w różnych kierunkach, oznacza to, że mogą istnieć dwa różne koszty. Za przykład można ponownie podać pokój na piętrze oraz magazyn znajdujący się pod nim, z tym, że do naszego świata dodajemy drabinę. Zastanawiając się o kosztach w kategoriach czasu skok z pokoju do magazynu nie zajmuje w ogóle czasu, ale może upłynąć kilka sekund zanim postać gracza zdąży się wspiąć z powrotem na górę po drabinie. Ponieważ koszty są związane z każdym połączeniem mogą być po prostu reprezentowane: połączenie A(piętro) z B(magazyn) ma mały koszt, a połączenie z B do A ma większy koszt. Matematycznie graf skierowanie jest identyczny jak graf nieskierowany z wyjątkiem pary węzłów, które stanowią połączenie jest teraz uporządkowane. Podczas gdy połączenie [węzeł A, węzeł B, koszt] w nieskierowanym grafie jest identyczne do połączenia [węzeł B, węzeł A, koszt] (tak długo jak koszty są identyczne ) w grafie skierowanych są one po prostu innymi połączeniami.

## Terminologia

Terminologia dotycząca gafów jest zmienna. W literaturze matematycznej często można natknąć się na określenie wierzchołek niż węzeł oraz krawędź niż połączenie (oczywiście zamiast pojęcia wagi stosowany jest koszt jak już było wspominane). Wielu programistów, którzy aktywnie angażują się w badania nad Sztuczną Inteligencją stosuje terminologie naruszającą matematyczną literaturę. Może to okazać się mylące w kontekście procesu tworzenia gier, ponieważ pojęcie wierzchołka ma zupełnie inne znaczenie. Niestety nie ma spójnej terminologii dotyczącej grafów w kontekście znajdowania ścieżki. Można spotkać w różnego typu artykułach oraz na konferencjach poświęconych programowaniu gier określenie "kropka" na węzeł, natomiast łuk, ścieżka, łącze i linia są określeniami dla połączenia. Na potrzeby tej pracy będzie stosowana terminologia węzłów i połączeń, ponieważ jest to powszechne, dość sensowne (w przeciwieństwo do kropek i linii) oraz jednoznaczne (łuki i wierzchołki mają znaczenie zarówno w grafice komputerowej). Ponadto, podczas omawiania skierowanych grafów o nieujemnych wagach prawie każda pozycja literaturowa dotycząca zagadnienia odnajdowania ścieżki nazywa te grafy właśnie w taki sposób .

## Reprezentacja

Reprezentacja grafu musi być przedstawiona w taki sposób, żeby algorytmy odnajdywania ścieżki takie jakie A\* czy Dijkstra mogły pracować na nich. Jak można się przekonać algorytmy muszą zdobyć informacje o połączeniach wychodzących z danego węzła. I dla każdego z tych połączeń muszą one mieć dostęp kosztu i punktu przeznaczenia.

Można dokonać reprezentacji grafu za pomocą następującego interfejsu:

interface Graph

{

/\*Zwraca tablice połączeń (interfejs Connection) wychodzących z podanego jako parametr węzła \*/

Connection [] getConnetions(Node fromNode);

}

interface Connection

{

/\* Zwraca nieujemny koszt połączenia \*/

uint getCost();

/\* Zwraca węzeł z które to połączenie pochodzi \*/

Node getFromNode();

/\* Zwraca węzeł do którego to połączenie prowadzi \*/

Node getToNode();

}

Interfejs Graph będzie zwracał tablicę obiektów połączeń dla każdego węzła, który został o to zapytany. Za pomocą tych obiektów można pobrać koszt oraz węzeł końcowy. Prosta implementacja takiej klasy będzie przechowywać połączenia dla każdego węzła oraz dzięki niej można w prosty sposób zwrócić listę tych połączeń.

Bardziej zaawansowana implementacja może liczyć koszt tylko wtedy, kiedy jest to wymagane, korzystając z obecnej informacji na temat struktury poziomu. Należy zauważyć, że nie ma określonego typu danych dla węzła w tym interfejsie, ponieważ nie ma potrzeby specyfikowania go. W wielu przypadkach wystarczy tylko nadać węzłowi unikalny numer i użyć liczb całkowitych jako typ danych. Jest to bardzo dobra implementacja, ponieważ otwiera pewne specyficzne i bardzo szybkie możliwości optymalizacyjne algorytmu A\*.

## Algorytm Dijkstry

Nazwa algorytm Dijkstry wzięła sie od holenderskiego matematyka Edsger'a Dijkstrty, który jest jego twórcą. Algorytm Dijkstry nie był pierwotnie zaprojektowany do odnajdywania ścieżki w rozumieniu gier komputerowych. Został on zaprojektowany, aby rozwiązać problem matematycznej teorii grafów, łudząco nazywany "najkrótsza droga".

Gdzie odnajdywanie ścieżki w grach ma jeden punkt początkowy i jeden punkt docelowy, natomiast algorytm znajdujący najkrótszą ścieżkę jest zaprojektowany tak, aby znaleźć najkrótsze trasy do wszystkich węzłów od punktu startowego. Rozwiązanie tego problemu zawiera w sobie rozwiązanie problemu odnajdywania ścieżki w grach komputerowych, ale

jest to duże marnotrawstwo, jeśli potem w kolejnych krokach musimy odrzucać wszystkie pozostałe ścieżki. Można by pokusić się o pewną modyfikację algorytmu, która wygeneruje interesującą z punktu widzenia programisty ścieżkę, ale wciąż jest ona tworzona nieefektywnie.

Z tego powodu bardzo praktycznie w ogóle nie korzysta się z algorytmu Dijkstry do rozwiązania problemu odnajdywania ścieżki. Jego wykorzystanie zostało raz zaimplementowane, jednak nie jako algorytm odnajdywania ścieżki, ale do analizy właściwości ogólnych danej mapy w zaawansowanym systemie odnajdywania ścieżki w pewniej symulacji wojskowej. Niemniej jednak, jest to ważny algorytm taktycznej analizy i ma zastosowanie w kilku innych obszarach Sztucznej Inteligencji w grach. Zbadana zostanie tutaj jego prostsza wersja jako ogólnego algorytmu to odnajdywania ścieżki,

## Przedstawienie problemu

Dany jest graf (skierowany o nieujemnych wagach) i dwa węzły (początkowy i końcowy) w tym grafie. Zadaniem algorytmu jest wygenerowanie ścieżki tak, aby całkowity koszt ścieżki był minimalny spośród wszystkich ścieżek od startu do celu. Może dojść do sytuacji że będzie istnieć wiele ścieżek o takim samym minimalnym koszcie. Ilustracja 13 zawiera 10 możliwych ścieżek, wszystkie z tym samym kosztem minimalnym. Gdy istnieje więcej niże jedna optymalna ścieżka, oczekuje się, że tylko jedna z nich zostanie zwrócona, bez znaczenia która. Warto wspomnieć, że oczekiwany rezultat powinien składać się z pewnego zbioru połączeń, a nie węzłów. Dwa węzły mogą być połączone przez więcej niż jedno połączenie i każde połączenie może mieć inny koszt (przykładowo koszt zeskoku z piętra do magazynu lub koszt wspięcia się po drabinie z magazynu na piętra). Dlatego trzeba dokładnie wiedzieć które połączenia są potrzebne, a lista węzłów byłaby niekompletna. Wiele gier nie robi tego rozróżnienia. Istnieje tam co najwyżej jedno połączenie pomiędzy dowolną parą węzłów. Pomimo to, jeśli są dwa połączenia pomiędzy parą węzłów, to moduł odpowiedzialny za odnajdywanie ścieżki powinien zawsze wybrać to połączenie z najniższym kosztem. W niektórych zastosowaniach, jednakże koszt zmienia się w trakcie gry lub pomiędzy różnymi postaciami oraz w przypadku śledzenia wielu połączeń okazuje się przydatne. Tak właśnie można zamknąć temat w którym algorytm ma sobie radzić z wieloma połączeniami. Na potrzeby tej pracy należy przyjąć, że ścieżkę będziemy identyfikować jako listę połączeń.



Ilustracja Przykład grafu, gdzie każda ścieżka jest optymalna

## Opis algorytmu

Nieformalnie, algorytm Dijkstry działa przez rozsiew od węzła początkowego wzdłuż swoich połączeń. Podczas rozprzestrzeniania się do, co raz bardziej odległych węzłów prowadzi rejestr kierunku, z którego pochodzi (można to sobie wyobrazić jako rysunek strzałek wykonanych kredą na podłodze, tak aby wskazać drogę powrotną). W końcu jak algorytm dotrze do celu zaczyna on podążać z powrotem po "strzałkach" do punktu startowego, generując tym samym kompletną ścieżkę. Ze względu na sposób działania przez proces rozprzestrzeniani, gwarantuje to strzałki z kredy wskazują zawsze najkrótszą drogę do punktu startowego. Bardziej szczegółowo można powiedzieć, że Dijkstra działa iteracyjnie. W każdej iteracji rozważany jest jeden węzeł grafu i podążą wychodzącymi z niego połączeniami. W pierwszej iteracji sprawdzi węzeł początkowy. W kolejnych krokach wybiera węzeł do rozważenia za pomocą algorytmu, który zostanie potem krótko omówiony. Na potrzeby nazewnictwa iteracji węzła zostanie przyjęte określenie "bieżący węzeł".

**Przetwarzanie bieżącego węzła**

Podczas iteracji, rozważa się każde wychodzące połączenie z bieżącego węzła. Dla każdego połączeni znajduje się węzeł końcowy i zapisuje całkowity koszt ścieżki znaleziony do tej pory wraz z połączeniami, do których dotarł algorytm.

W pierwszej iteracji, gdzie węzeł startowy jest bieżącym węzłem, kosztem dotychczasowym dla każdego połączenia w węźle końcowym jest po prostu koszt połączenia. Ilustracja 14 pokazuje sytuacje pierwszej iteracji. Każdy węzeł połączony z węzłem startowym ma koszt dotychczasowy równy kosztowi połączenia, które doprowadziło go tam - w tym przypadku to po prostu zapis wagi z tego połączenia.

Dla kolejnych iteracji dotychczasowy koszt węzła końcowego każdego połączenia jest sumą kosztów połączeń i kosztów dotychczasowych bieżącego węzła (to znaczy węzeł, z którego połączenie przychodzi.) Ilustracja 15 przedstawia kolejną iteracja tego samego grafu. Tutaj koszt dotychczasowy w węźle E jest sumą kosztu dotychczasowego z węzła B i kosztem połączenia IV z B do E.

W implementacji algorytmu, nie ma różnicy między pierwszą i kolejną iteracją. Ustalając koszt dotychczasowych wartości węzła początkowego jako 0 można skorzystać z jednego kawałka kodu dla wszystkich iteracji algorytmu.



Ilustracja Dijkastra - pierwsza iteracja



Ilustracja Dijkstra - kolejna iteracja

**Lista węzłów**

Algorytm śledzi ścieżki wszystkich węzłów, które miały miejsce w dwóch listach zwanych: lista otwarta i lista zamknięta. W otwartej liście rejestruje się wszystkie węzły dostępne, które nie miały jeszcze swojej własnej iteracji. Węzły przetworzone są zapisywane w liście zamkniętej. Na początku lista otwarta zawiera tylko węzeł początkowy (z zerowym dotychczasowym kosztem), a lista zamknięta jest pusta. Każdy węzeł może być należeć do jednej z trzech kategorii: może on należeć do zamkniętej listy - został on przetworzony w procesie swojej własnej iteracji, może być na liście otwartej - będąc odwiedzonym przez inny węzeł, ale jeszcze nie przetworzony w swojej iteracji, może również nie znajdować się na żadnej z wymienionych wyżej list. Czasami mówi się, że węzeł jest zamknięty, otwarty lub nieodwiedzony. W każdej iteracji, algorytm wybiera węzeł z listy otwartej, który ma najmniejszy dotychczasowy koszt. Następnie zostaje przetworzony w normalny sposób. Przetworzony węzeł jest usuwany z otwartej listy i dodany do listy zamkniętej. Jest jeszcze jedna komplikacja. Kiedy algorytm podąża za połączeniem od bieżącego węzła, zakłada się, że ostatecznie znajdzie się w węźle nieodwiedzonym. Może się zdarzyć taka sytuacja, że zamiast tego algorytm kończy w węźle, który jest otwarty lub zamknięty, aby rozwiązać taki przypadek trzeba postąpić nieco inaczej.

**Obliczanie dotychczasowych kosztów dla węzłów otwartych i zamkniętych**

W przypadku, gdy algorytm dochodzi do otwartego lub zamkniętego węzła podczas iteracji, to rozpatrywany węzeł będzie już posiadał dotychczasowych koszt i zapis połączenia, które doprowadziło do niego. Teraz wystarczy wstawić wartość która nadpisze wynik poprzedniej iteracji algorytmu. Zamiast tego, można sprawdzić czy trasa, która obecnie jest rozważana jest lepsza od poprzednio znalezionej. Obliczenie dotychczasowego kosztu przebiega normalnie - jeśli wartość kosztu jest większa niż ta aktualnie zarejestrowana, to nie aktualizujemy węzła w ogóle i nie zmieniamy jego listy. Jeśli nowy koszt jest mniejszy niż obecny dotychczasowy koszt węzeł, to wtedy dokonywana jest aktualizacja węzła nową, lepszą wartość, kolejno jest zapisywana w połączeniu. Następnie węzeł powinien zostać umieszczony na liście otwartej. W przypadku jeśli był on poprzednio na liście zamkniętej, to zostaje z niej usunięty. Podsumowując Dijkstra nie znajdzie lepszej trasy do zamkniętego węzła, więc może sprawdzić czy węzeł jest zamknięty najpierw i kontynuować dalsze testy dotychczasowych kosztów.

Fachowa implementacja algorytmu Dijkstry właśnie tak działa. W kolejnych rozdziałach można się jednak przekonać, że przypadku algorytmu A\* jest jednak inaczej, jakkolwiek w obydwu przypadkach muszą zostać sprawdzone najszybsze ścieżki. Ilustracja 16 przedstawia aktualizacje otwartego węzła w grafie. Nowa ścieżka przez węzeł C jest szybsza i zapis dla węzła D jest odpowiednio zaktualizowany.



Ilustracja Aktualizacja otwartego węzła

**Przerywanie algorytmu**

Podstawowa wersja algorytmu przerywa działanie, kiedy otwarta lista jest pusta, a oznacza to, że każdy węzeł został rozważony oraz wszystkie węzły znajdują się na liście zamkniętej.

W przypadku odnajdywania ścieżki obiektem zainteresowania jest tylko węzeł docelowy, więc dzięki temu można przerwać algorytm wcześniej. Dokonuje się przerwania w momencie, kiedy węzeł docelowy jest najmniejszym węzłem na liście otwartej.

Warto zwrócić uwagę na warunek przerwania. Algorytm nie jest przerywany w momencie napotkania węzła docelowego.

Ilustracja 17. Węzeł D jest węzłem docelowym i zostaje odnaleziony podczas przetwarzania węzła B. W tym momencie algorytm zostaje przerwany i zwracana jest ścieżka A-B-D, która nie jest najkrótszą ścieżką. Aby mieć pewność, że znaleziona ścieżka jest najkrótszą, algorytm musi działać dalej, dopóki cel nie będzie miał najmniejszego dotychczasowego kosztu. W tym i tylko w tym punkcie obliczeń, każdy inny punkt(znajdujący się na liście otwartej lub nieprzetworzony) będzie posiadał dłuższą drogę.

W praktyce ta zasada jest często łamana. Algorytm jest modyfikowany tak, że jeśli zostaje znaleziona ścieżka do celu to zakłada się, że ona jest najkrótsza, nawet jeśli istnieje inna, krótsza - zwykle jest ona o krótsza o niewielką wartość liczbową.

Z tego powodu większość programistów implementując algorytm Dijkstry przerywa algorytm odnajdywania ścieżki dopóki nie natknie się on na węzeł końcowy, zamiast pozwolić algorytmowi działać dalej, dopóki węzeł docelowy zostanie wybrany z otwartej listy.

**Odtwarzanie ścieżki**

Ostatnim krokiem algorytmu jest odtworzenie ścieżki. Dokonuje się tego zaczynając od punktu końcowego i podąża się połączeniami, które doprowadziły ścieżkę do tego miejsca. Można powiedzieć, że dokonuje się powrotu i szukanym punktem jest punkt początkowy. Kontynuuje się proces zapamiętując połączenie dopóki punkt startowy nie znajdzie się na powrotnej ścieżce. Ostatnim krokiem jest odwrócenie kolejności połączeń ścieżki i można zwrócić gotowe rozwiązanie.

*Ilustacja 17* przestawia prosty graf zaraz po tym jak algorytm zakończył działanie. Znaleziona lista połączeń zawiera węzły od celu do początku - jej odwrócenie pozwoli uzyskać poszukiwaną drogę.



Ilustracja Graf po zakończeniu obliczeń

## Algorytm A\*

Odnajdywanie ścieżki w grach jest synonimem algorytmu A\*. A\* jest prosty w implementacji, bardzo efektywny oraz posiada bardzo duże możliwości optymalizacji. Każdy moduł zajmujący się odnajdywaniem ścieżki na przestrzeni 10 lat wyżywał pewnej wariacji algorytmu A\* jako jego kluczowego algorytmu. Może również zostać wykorzystany do planowania złożonych akcji dla postaci.

W odróżnieniu do algorytmu Dijkstry, A\* został zaprojektowany dla odnajdywania ścieżki z punktu do punktu i nie jest stosowany do rozwiązywania problemu najkrótszej ścieżki w teorii grafów. A\* może zostać śmiało rozszerzony o bardziej zaawansowane przypadki.

## Przedstawienie problemu

Problem jest identyczny do tego przedstawionego w algorytmie Dijkstry.

Dany jest graf (skierowany o nieujemnych wagach) i dwa węzły w tym grafie (startowy i docelowy). Zadaniem algorytmu jest wygenerowanie takiej ścieżki, że całkowity koszt jej jest minimalny spośród wszystkich dostępnych ścieżek od startu do celu. Wyliczona ścieżka jest tak naprawdę listą połączeń od węzła początkowego do końca.

## Opis algorytmu

Nieformalnie, algorytm działa na takiej samej zasadzie jak Dijkstra. Podobnie występuje tam otwarta lista w najniższymi dotychczasowymi kosztami, ponadto zostaje tam wybrany węzeł, który jest najbardziej prawdopodobny do dotarcia do celu. Użyte sformułowanie "najbardziej prawdopodobny" mówi, że jest to kontrolowane przez heurystykę. Jeśli heurystyka została dobrana trafnie to algorytm będzie efektywny, w przeciwnym wypadku efektywność będzie gorsza od Dijkstry.

Algorytm A\* działa iteracyjnie. W każdej iteracji jest przetwarzany jeden węzeł grafu i podąża dalej wychodzącymi połączeniami. Węzeł (ponownie nazywany bieżącym węzłem) jest wybierany używając algorytmu selekcji podobnego do tego zastosowanego w Dijkstrze, ale z znaczącą różnicą heurystyki.

**Przetwarzanie bieżącego węzła**

Podczas iteracji A\* rozważa każde wychodzące połączenie z bieżącego węzła. Dla każdego połączenia znajduje węzeł końcowy i przypisuje całkowity koszt ścieżki tymczasowej i połączenia, z którego przyszedł - tak jak miało to miejsce w poprzednim algorytmie .

Dodatkowo algorytm przypisuje jeszcze jedną wartość: estymacje całkowitego kosztu dla ścieżki od węzła startowego przez obecny węzeł do celu (ta wartość zostanie teraz nazwana: estymowany koszt całkowity). Ta estymacja jest sumą dwóch wartości: kosztu dotychczasowego i wartości określającej jak daleko węzeł znajduje się od celu. Estymacja jest generowana przez oddzielną część kodu, która nie jest częścią algorytmu.

Estymacja jest powszechnie nazywana "wartością heurystyki" węzła, nie może być to wartość ujemna (wynika to z tego, że koszty w grafie są nieujemne, więc bezsensownym było by, aby estymacja miała wartość ujemną). Proces generowania wartości heurystyki jest punktem kluczowym w procesie implementacji algorytmu A\* - zostanie on omówiony później.

*Ilustracja 18* prezentuje obliczone wartości dla paru węzłów w grafie. Węzły są opisane, również wartościami heurystyki oraz są przestawione dwie wartości (dotychczasowy koszt, estymowany koszt całkowity), które informują, że dane węzły zostały przetworzone przez algorytm.



Ilustracja A\* - działanie

**Lista węzłów**

Jak w poprzednim przypadku algorytm trzyma w liście otwartej węzły, które zostały odwiedzone, ale nie przetworzone, a w liście zamkniętej te które zostały już przetworzone. Węzły są przenoszone do listy otwartej jeśli były węzłami końcowymi na liście dotychczasowo rozważanych połączeń. W przypadku listy zamkniętej - umieszczane są tam węzły, które zostały przetworzone w swojej iteracji.

W odróżnieniu od poprzedniego algorytmu węzeł z najmniejszym estymowanym kosztem całkowitym jest wybierany do każdej iteracji. Jest to prawie zawsze inny węzeł niż ten posiadający najmniejszy koszt dotychczasowy.

Ta różnica powoduje, że algorytm będzie przeszukiwał bardziej obiecujące węzły. Jeśli węzeł ma małą wartość estymowanego kosztu całkowitego, to musi on mieć też relatywnie niski koszt dotychczasowy i relatywnie małą estymacje dystansu do przebycia do celu. Jeśli estymacje są dokładne to węzły, które są bliższe celu są brane pod uwagę najpierw, nakierowując poszukiwania na najbardziej korzystny obszar.

**Obliczanie kosztu dotychczasowego dla otwartej i zamkniętej listy**

Jak wspomniano poprzednio podczas przebiegu procesu algorytmu może on dotrzeć do węzła oznaczonego jako zamknięty lub otwarty i trzeba dokonać aktualizacji zapisanych wartości.

Algorytm analogicznie dokonuje obliczenia wartość kosztu dotychczasowego i jeśli nowa wartość jest mniejsza od tej istniejącej w węźle, to dokonywana jest aktualizacja węzła. Warto zwrócić uwagę na fakt, że dokonywane jest porównanie na wartości kosztu dotychczasowego, a nie na estymowanym koszcie całkowitym.

W porównaniu do Dijkstry, A\* może znaleźć lepsze drogi do węzłów, które już znajdują się na liście zamkniętej. Jeśli poprzednia estymacja była bardzo optymistyczna, to węzeł możesz zostać przetworzony z przekonaniem, że był to najlepszy wybór, jednak faktycznie nie był.

To powoduje pewien problem. Jeśli niepewny węzeł zostanie przetworzony i umieszczony na liście zamkniętej, to znaczy że wszystkie jego połączenia zostały sprawdzone. Może się zdarzyć, że zbiór wszystkich węzłów ma koszt dotychczasowy obliczony na podstawie kosztu jednego z niepewnych węzłów. W takim przypadku aktualizacja tylko tego węzła nie wystarczy. Należy dokonać aktualizacji wszystkich połączeń, przez propagacje nowej wartości. W przypadku węzła na liście otwartej nie jest to konieczne - jak wiadomo połączenia węzłów na liście otwartej nie zostały jeszcze przetworzone. Istnieje metoda, która pozwoli na ponownie przeliczenie i propagacje nowej wartości. Można to osiągnąć poprzez usunięcie węzła z listy zamkniętej i umieszczenie go na liście otwartej. Algorytm będzie dalej kontynuował swój proces przetwarzając usunięty węzeł, umieszczając go ponownie na liście zamkniętej. Każdy węzeł, którego wartość kosztu jest związana z ponownie rozpatrywanym węzłem, zostanie przetworzony jeszcze raz.

*Ilustracja 19* przestawia aktualizacje grafu - jest to analogiczna sytuacja do poprzednio przedstawionego grafu, lecz dwie iteracje później. Przedstawia on sytuacje, w której aktualizowany jest zamknięty węzeł. Nowa trasa do węzła E przez węzeł C jest szybsza, więc dane węzła E odpowiednio aktualizowane i zostaje on umieszczony na otwartej liście. W następnej iteracji wartość węzła G zostaje zmieniona. Tak, więc węzły znajdujące się na zamkniętej liście mają zmienioną wartość kosztu i zostają z niej usunięte oraz przeniesione do listy otwartej. Otwarte węzły, które mają zmienione wartość zostają na otwartej liście.

**Przerywanie algorytmu**

W wielu implementacjach A\* przerywa swoje działanie, kiedy węzeł docelowy jest najmniejszym węzłem na liście otwartej.

Często jednak zdarza sie, że węzeł posiadający najmniejszy estymowany koszt całkowity może zostać zaktualizowany. Nie da się zagwarantować, że węzeł, który jest pierwszym na otwartej liście jest tym, który posiada najkrótszą ścieżkę do tego miejsca. Dlatego przerwanie algorytmu w tym momencie nie daje gwarancji, że została znaleziona najkrótsza ścieżka.

Jest to oczywiste, że proces algorytmu mógłby trwać troszkę dłużej, żeby wygenerować optymalny wynik. Można to osiągnąć dzięki przerwaniu wykonania algorytmu w momencie, w którym węzeł na otwartej liście z najmniejszym dotychczasowym kosztem ma ten koszt większy niż koszt znalezionej ścieżki do celu. Wtedy i tylko wtedy jest zagwarantowane, że żadna przyszła ścieżka nie będzie krótsza.

Implementacje A\* biorą pod uwagę fakt, że mogą one zwrócić teoretycznie nieoptymalny rezultat. Istnieje jednak funkcja heurystyczna, która kontroluje algorytm. Zależnie od wyboru takiej funkcji jest zagwarantowane, że zostaną zwrócone optymalne rezultaty lub można świadomie dopuścić nieoptymalne wyniki, co daje szybszy czas wykonania.

Z racji tego, że proces A\* często jest bliski znalezieniu optymalnego wyniku, duża liczba implementacji przerywa algorytm, kiedy cel jest pierwszym odwiedzonym węzłem bez czekania, żeby był pierwszy na otwartej liście. Wzrost wydajności nie jest tak duży jak w przypadku zrobienia identycznej rzeczy w Dijkstrze.

**Odtwarzanie ścieżki**

Proces odtwarzania ścieżki zaczyna się od miejsca docelowego i gromadzi połączenia cofając się do punktu startowego. Połączenia muszą zostać odwrócone, żeby utworzyć poprawną ścieżkę.



Ilustracja Aktualizacja zamkniętego węzła