

**AGH**

**AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE**

**WYDZIAŁ ELEKTROTECHNIKI, AUTOMATYKI,  
INFORMATYKI I INŻYNIERII BIOMEDYCZNEJ**

Praca dyplomowa magisterska

*Budowa internetowego kontekstowego asocjacyjnego silnika  
wyszukiwania informacji z wykorzystaniem sztucznych systemów  
skojarzeniowych.*

*Construction of internet context associative search engine using  
artificial associative systems.*

Autor:

*Daria Para*

Kierunek studiów:

*Inżynieria Biomedyczna*

Opiekun pracy:

*dr hab. Adrian Horzyk*

Kraków, 2015

*Oświadczam, świadomy(-a) odpowiedzialności karnej za poświadczanie nieprawdy, że niniejszą pracę dyplomową wykonałem(-am) osobiście i samodzielnie i nie korzystałem(-am) ze źródeł innych niż wymienione w pracy.*

*Serdeczne podziękowania dla promotora pracy  
Pana Doktora Habilitowanego Adriana Horzyka  
za poświęcony czas, życzliwość,  
cierpliwość oraz cenne uwagi.*



## Spis treści

<b>1. Wstęp</b> .....	7
1.1. Cel pracy.....	7
1.2. Biologiczne inspiracje .....	7
1.2.1. Budowa mózgu - neurony .....	8
1.2.2. Aktywacja neuronów, przewodzenie impulsów.....	9
<b>2. Teoretyczne rozważania</b> .....	11
2.1. Informacja.....	11
2.2. Skojarzenia .....	11
2.3. Wiedza .....	12
2.4. Pamięć .....	12
2.5. Inteligencja .....	13
<b>3. Dostępne narzędzia oraz teoria</b> .....	15
3.1. Mechanizmy skojarzeniowe w informatyce .....	15
3.2. Pozyskiwanie i wyszukiwanie wiedzy [8].....	16
3.3. Asocjacyjne relacje [1] .....	17
3.4. Semassele .....	17
<b>4. Matematyczne fundamenty pracy [1]</b> .....	19
4.1. Neuroasocjacyjne grafy wiedzy ANAKG .....	19
<b>5. Realizacja tematu</b> .....	23
5.1. Część sieciowa.....	23
5.1.1. Budowa .....	24
5.1.2. Działanie .....	26
5.2. Część grafu .....	26
5.2.1. Budowa .....	27
5.2.2. Działanie .....	28
5.3. Część odpowiedzi .....	31
5.3.1. Budowa i działanie .....	31

---

<b>6. Wyniki</b> .....	35
6.1. Pollyanna .....	35
6.1.1. Jednokrotne pobudzenie pytaniem.....	35
6.1.2. Dwukrotne pobudzenie pytaniem .....	36
6.1.3. Trzykrotne pobudzenie pytaniem.....	36
6.1.4. Czterokrotne pobudzenie pytaniem.....	36
6.2. Fryderyk Chopin.....	37
6.2.1. Jednokrotne i dwukrotne pobudzenie pytaniem.....	38
6.2.2. Trzykrotne i czterokrotne pobudzenie pytaniem.....	38
6.2.3. Pięciokrotne pobudzenie pytaniem .....	39
6.2.4. Piętnastokrotne pobudzenie pytaniem.....	39
6.3. Fryderyk Chopin 2.....	39
6.3.1. Jednokrotne i dwukrotne pobudzenie pytaniem.....	39
6.4. Anne of Green Gables .....	39
6.4.1. Jednokrotne pobudzenie pytaniem .....	40
6.4.2. Trzykrotne pobudzenie pytaniem.....	40
<b>7. Podsumowanie</b> .....	43
<b>A. Załącznik A - Zawartość dołączonej płyty</b> .....	47

# 1. Wstęp

W dobie internetu i powszechnego dostępu do informacji zdobycie wiedzy nie powinno być problemem. Wiele wyszukiwarek internetowych pozwala szybko znaleźć potrzebne dane. Prosty wydaje się znalezienie konkretnej informacji. Jednak kiedy zagadnienie nie jest dokładnie sprecyzowane, albo na postawione przez szukającego pytanie nie ma jasno i dokładnie określonej odpowiedzi, zamiast szybkiej odpowiedzi na użytkownika czeka mozolne przeszukiwanie zawartości poszczególnych stron, analizowanie ich treści oraz wyciąganie wniosków. Tzn. do uzyskania odpowiedzi potrzebna jest baza wiedzy (internet) oraz jednostka myśląca, kojarząca, wnioskująca (człowiek).

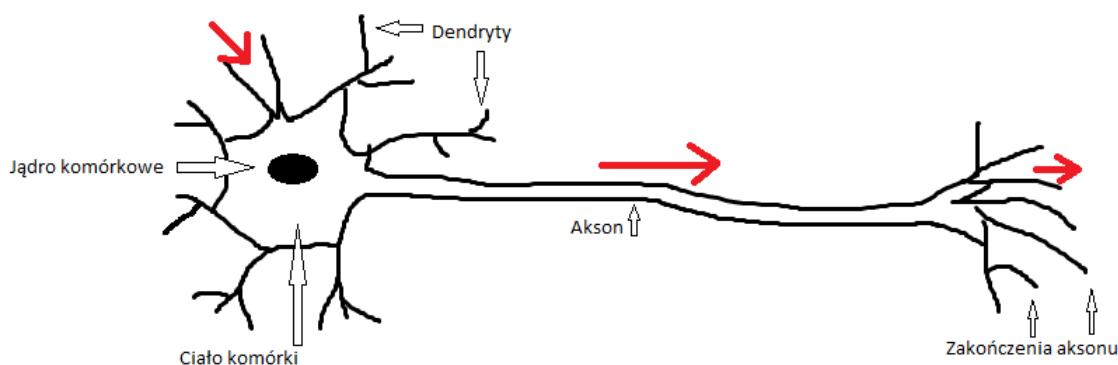
W poniższej pracy podjęto próbę stworzenia mechanizmu, który po części mógłby spełniać te wymagania.

## 1.1. Cel pracy

Zakres pracy obejmuje skonstruowanie systemu lingwistycznego zdolnego gromadzić wiedzę i hiperłącza wokół wyszukiwanych kwestii z wykorzystaniem sztucznych systemów skojarzeniowych. Celem działania tego systemu jest udzielenie pytającemu rozbudowanej i rzetelnej informacji na temat określonego zagadnienia. Najpierw system automatycznie przeszukuje zasoby internetowe oraz buduje sztuczny system skojarzeniowy, w którym formuje się wiedza wokół określonego zagadnienia opisanego przez pytającego, zawierająca skojarzenia z neuronami reprezentującymi hiperłącza do nich. Następnie pytającemu przekazywane są informacje na podstawie najmocniejszych skojarzeń, jakie wynikają z działania tego systemu. Pytający ma możliwość dopytania o szczegóły, które służą jako kontekst do wywoływania słabszych skojarzeń w ramach uformowanej wiedzy. Odpowiedzi są asocjacyjnie i kontekstowo uzupełnione hiperłączami do kwestii związanych z udzielanymi odpowiedziami na podstawie skojarzeń. [12]

## 1.2. Biologiczne inspiracje

Sieci skojarzeniowe, a wcześniej także sieci neuronowe są próbą informatycznego odzwierciedlenia działania ludzkiego mózgu. Mózg jest organem, który nie przestaje zadziwiać badaczy na całym świecie. Dzięki jego pracy jesteśmy w stanie poruszać się, odczuwać, mówić, myśleć, pamiętać. Można powiedzieć, że mózg jest organem nadzorującym pracę całego organizmu, jest jednostką centralną pod względem informacyjnym. Odpowiada za kontakt z całym ciałem oraz światem zewnętrznym.



Rysunek 1.1: Budowa neuronu z zaznaczeniem czerwonymi strzałkami kierunku przepływu impulsów

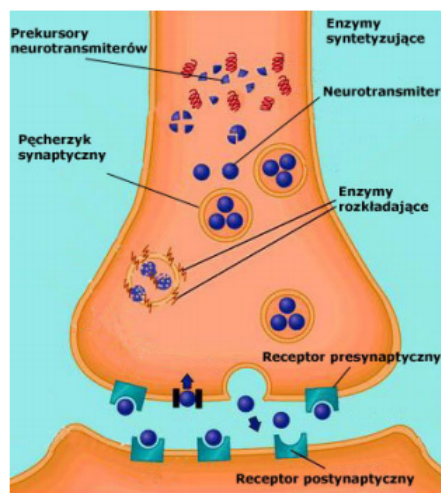
Przyjęło się uważać serce za najważniejszy element ludzkiego ciała. Jednak przeszczep serca, który teraz jest dość powszechnym zjawiskiem, nie zmienia osobowości człowieka, wspomnień (wbrew pierwotnym opiniom, z którymi stykali się prekursorzy przeszczepów serca). Z drugiej strony praktyka przeszczepu mózgu w medycynie nie istnieje. Spośród wszystkich organów ludzkiego ciała, to właśnie w mózgu „zapisane” są informacje, które tworzą osobowość człowieka – jego zdolność do myślenia, pamięć, inteligencja.

Poniżej przedstawiono pokrótce budowę ludzkiego mózgu, sposób aktywacji neuronów, metodę przenoszenia informacji, mechanizm kojarzenia oraz zjawisko pamięci. Są to biologiczne wzorce, którymi posłużono się do stworzenia sieci neuronowych.

### 1.2.1. Budowa mózgu - neurony

Mózg człowieka składa się z milionów komórek. Istniejące dwa rodzaje komórek w systemie nerwowym: to neurony (komórki nerwowe) oraz komórki glikowe. Podstawową jednostką funkcjonalną ludzkiego mózgu jest neuron. To dzięki oddziaływaniom pomiędzy neuronami człowiek może chodzić, czuć oraz myśleć. Neuron składa się z kilku części: ciała komórki z jądrem komórkowym, otaczających go wypustek - dendrytów, aksonu oraz jego zakończeń (rys. 1.1). Ciało komórki jest centrum metabolicznym neuronu. Dendryty to wypustki biegnące do ciała komórki, które zbierają impulsy dochodzące z innych komórek nerwowych. Akson jest najdłuższą wypustką neuronu, w przeciwieństwie do dendrytów przeprowadza impuls z ciała komórki do innych neuronów. Zakończenia aksonu znajdują się blisko dendrytów następnych neuronów. Obszar położenia zakończenia aksonu jednego neuronu i dendrytu drugiego nazywany jest synapsą. Komórka, która wysyła informacje nazywana jest neuronem pre-synaptycznym, a komórka odbierająca informacje to neuron postsynaptyczny. Synapsa nie jest jednak fizycznym połączeniem pomiędzy dwoma neuronami, jest to obszar zbliżenia (1.2). Przeciętny neuron tworzy około 1 000 połączeń synaptycznych z innymi neuronami. Dodatkowo połączenia synaptyczne nie są statyczne. Neurony tworzą nowe połączenia synaptyczne w reakcji na doświadczenia, przeżycia życiowe. Te dynamiczne zmiany w połączeniach neuronowych umożliwiają uczenie się.[9]





Rysunek 1.2: Schemat funkcjonowania synapsy [11]

### 1.2.2. Aktywacja neuronów, przewodzenie impulsów

W efekcie pobudzenia w komórkach pobudliwych (neurony i komórki mięśniowe) generowany jest potencjał czynnościowy. Jest to chwilowe odwrócenie elektrycznej polaryzacji błony komórkowej [3]. Potencjały czynnościowe powstają na wzniesieniu aksonowym neuronu i rozprzestrzeniają się wzdłuż błony aksonu. Mają charakter progowy, czyli zachowują się zgodnie z zasadą „wszystko albo nic”. Między początkiem bodźca, a wytworzeniem potencjału czynnościowego występuje faza latencji, czyli krótkie opóźnienie, okres utajenia. W czasie trwania potencjału czynnościowego ma miejsce okres refrakcji bezwzględnej, znaczy to, że neuron jest całkowicie niepobudliwy. Po zakończeniu fazy iglicowej komórka nerwowa może być pobudzona jedynie przez dużo silniejszy bodziec (ponadprogowy), ten stan nazywamy okresem refrakcji względnej. Występowanie okresów refrakcji gwarantuje, że potencjały czynnościowe nie będą się sumować, ograniczona jest maksymalna częstotliwość potencjału czynnościowego oraz może się on poruszać tylko w jednym kierunku. Synapsy elektryczne, o których wspomniano w podrozdziale 1.2.1 są zbudowane z grup kanałów jonowych - koneksonów. Są one połączone w obrębie złączy szczelinowych i umożliwiają występowanie sprzężenia elektrycznego między komórkami. Dzięki temu potencjały czynnościowe rozprzestrzeniają się między neuronami bez zaburzeń i z dużą prędkością [3]. Kiedy w zakończeniu nerwowym wystąpi potencjał czynnościowy ma miejsce uwolnienie neuroprzebieźnika (rys. 1.2). Po przejściu szczeliny synaptycznej neuroprzebieźnik wiąże się z receptorami postsynaptycznymi. Istnieją dwa rodzaje tych receptorów: kanały jonowe bramkowane ligandem oraz receptory metabotropowe sprzężone z układami wtórnych przebieźników. W zależności od tego, który receptor zostanie pobudzony przebieźnictwo przebiega szybciej (kanały jonowe) lub wolniej (receptory metabotropowe). Może wystąpić reakcja pobudzająca, gdy aktywowany receptor zwiększa szanse wytworzenia potencjału czynnościowego przez neuron postsynaptyczny, lub w odwrotnym przypadku reakcja hamująca.



## **2. Teoretyczne rozważania**

Rozdział ten opisuje podstawowe pojęcia używane do omawiania procesów myślowych oraz będące istotą odróżniającą osobę myślącą od maszyny. Poruszając tę tematykę często używa się sformułowań takich jak: informacja, skojarzenie, wiedza, pamięć czy inteligencja. Poniższe podrozdziały przybliżą wymienione pojęcia.

### **2.1. Informacja**

Informacją można nazwać zbiór danych na określony temat. W codziennym życiu informacja jest rozumiana jako opis słowny lub wartość liczbowa. W informatyce informacja to dane wejściowe, jakie dostarczamy do programu, aby przeprowadzić na nich dalsze operacje i uzyskać szukany wynik. Innymi słowy informacja jest podstawą do dalszych działań. Według A. Horzyka informację tworzą te dane, które są ze sobą w jakiś sposób powiązane, a ich znaczenie jest znane odbiorcy. Jeżeli dane nie są ze sobą powiązane, powiązanie nie jest znane odbiorcy lub nie zmienia stanu wiedzy, działania tego odbiorcy, to dane takie nie mają dla niego wartości informacyjnej [1].

### **2.2. Skojarzenia**

W popularnych wyszukiwarkach internetowych proces znajdowania informacji przebiega poprzez użycie tzw. słów kluczowych. W artykułach przy zamieszczaniu ich na stronie internetowej należy zaznaczyć, które wyrazy posłużą jako takie właśnie słowa kluczowe. Ich wybór nie zawsze jest oczywisty. A to, które słowa zostaną oznaczone jako kluczowe, wpływa na częstość wyświetlania strony oraz jej wykorzystanie.

Jednak w naszym mózgu proces kojarzenia jest zdecydowanie bardziej skomplikowany. Według Słownika języka polskiego [14] słowo „kojarzyć” oznacza: „1. powodować połączenie, związek kogoś lub czegoś; 2. łączyć wrażenia, wyobrażenia tak, że pojawienie się jednych powoduje uświadomienie sobie innych; 3. rozumieć”.

W odniesieniu do tematu pracy należy zwrócić uwagę na pierwsze dwie definicje. Proces kojarzenia wykorzystuje połączenia neuronowe w mózgu, które utworzyły się na skutek pewnych procesów (może to być wielokrotne powtarzanie danej frazy, częste oglądanie jakiegoś obrazu z jednoczesnym słuchaniem konkretnego utworu muzycznego, bardzo silne emocje towarzyszące danemu wydarzeniu).

Sieć połączeń neuronowych w mózgu człowieka jest stale zmieniającą się, rozwijającą (pod wpływem bodźców) strukturą. Gdyby na pewnej grupie ludzi przeprowadzono test podając im różne wyrazy i oczekując pierwszego skojarzenia, które im przyjdzie na myśl (skojarzenie swobodne [5]), to dla wyrazów, które powszechnie występują najczęściej w określonym zestawieniu odpowiedzi powtarzałyby się. Jednak dla wyrazów rzadko spotykanych odpowiedzi byłyby bardzo różnorodne, ponieważ, w bardzo dużym uproszczeniu - to, jakie kto ma skojarzenia wynika z tego, jak kto żyje. Testy skojarzeniowe używane są w badaniach psychologicznych, ponieważ mogą one wiele powiedzieć o danej osobie. Oprócz skojarzeń swobodnych, bada się także skojarzenia kierowane, czyli słowa, które są związane z podanym hasłem jako zależne, synonimy, antonimy itd. [5]

### 2.3. Wiedza

Wiedzę można określić potocznie jako zbiór informacji, jakie posiada człowiek. Dokładniej mówiąc chodzi o informacje przez niego zapamiętane. Jan Strelau [5] rozróżnia informacje od wiedzy oraz proces gromadzenia danych od uczenia się. Według niego wiedza to system informacji zapisany w strukturach pamięci długotrwałej, dzięki któremu poznawczo odwzorowujemy świat. John R. Anderson zaproponował wyróżnienie dwóch rodzajów wiedzy: deklaratywnej (wiedzieć, że...) oraz proceduralnej (wiedzieć, jak...). Oba rodzaje często się uzupełniają. Trzeba posiadać wiedzę deklaratywną, aby wykształcić wiedzę proceduralną. Takie powiązanie nazywa się metawiedzą - pojęcie wprowadzone przez John'a H. Flavella.

### 2.4. Pamięć

Pamięć to utrwalone informacje „zapisane” w mózgu. To neurony z najmocniejszymi połączeniami. Pamięć to nie tylko informacje. Istnieje także pamięć smaku, węchu, pamiętać możemy także emocje. Często pod wpływem odczuć jakieś zdarzenie lub informację zapamiętujemy bardziej lub zapominaamy.

Badacze [5] podkreślają, że pamięć rozumieć można dwojako, jako wielofazowy proces lub jako właściwość danej jednostki. Podstawowym pojęciem używanym do określenia pamięci jest proces odpowiedzialny za rejestrowanie, przechowywanie oraz odtwarzanie doświadczeń. Larry Squire zaproponował podział pamięci na opisową (deklaratywną) i sposobów postępowania (niedeklaratywną) - analogiczny do podziału wiedzy Andersona). Pierwszy rodzaj dotyczy pamięci jawnej i zwerbalizowanej. Kiedy odnosi się do faktów i zależności jest określana jako pamięć semantyczna, w przypadku zdarzeń mówimy o pamięci epizodycznej. Pamięć niedeklaratywna jest utajona, niejawna, dotyczy umiejętności takich jak: pływanie, czytanie, rozróżnianie kolorów. Wymienione rodzaje pamięci mają różne mechanizmy neuropsychologiczne, można zauważyć, że w przypadku uszkodzenia określonych części mózgu wystąpi np. utrata informacji o przeszłości, otaczających osobach, ale zachowane są takie umiejętności jak czytanie. Istnieje także drugi podział rodzajów pamięci odnoszący się do czasu. Pamięć krótkotrwała nazywana jest pamięcią bezpośrednią, odnosi się do bodźców sprzed kilku sekund lub minut, ma ogra-

niczona pojemność, a jej zatrzymanie w pamięci wymaga ciągłego powtarzania. Pamięć długotrwała ma niewyczerpaną pojemność, pozwala zachować ślad wielu doznań trwale lub przez długi czas i nie wymaga powtarzania [3].

## 2.5. Inteligencja

Inteligencją można nazwać umiejętność człowieka w zakresie formowania opinii, wyciągania wniosków, tworzenia nowych rozwiązań, szukania odpowiedzi pomiędzy danymi. Według Słownika wyrazów obcych [6]: „inteligencja <łac. *intelligentia* = pojętność> 1. psych. zdolność rozumienia otaczających sytuacji i znajdowania na nie właściwych, celowych reakcji; zdolność rozumienia w ogóle, bystrość, pojętność”. Strelau [5] także formułuje definicję inteligencji jako zdolności do przystosowania się do okoliczności poprzez dostrzeżenie abstrakcyjnych powiązań, z wykorzystaniem zgromadzonych doświadczeń i skutecznej kontroli nad własnymi procesami poznawczymi. Inteligencja często utożsamiana jest z wiedzą. O człowieku odczytanym zazwyczaj mówi się, że jest inteligentny. Nie jest to do końca jednoznaczne, w końcu człowiek ten może tylko zapamiętywać przeczytane informacje. Jednak im większa wiedza, tym większa baza informacyjna, co za tym idzie, prościej jest formować nowe opinie i wyciągać poprawne wnioski. Czytanie (ze zrozumieniem) „ćwiczy” myślenie. Można w jednej historii odnaleźć klucz myślowy, schemat, który zastosowany na innych danych pozwoli odkryć nowe rozwiązania w danej dziedzinie. Aktualnie wraz z rozwojem metod obrazowania pracy mózgu metodami PET (ang. *positron emission tomography*) i MRI (ang. *magnetic resonance imaging*) szuka się biologicznych podstaw inteligencji w szybkości neuronalnej, sprawności układu nerwowego i wielkości mózgu [5]. Odnosząc się do związku inteligencji z wielkością mózgu należałoby uściślić, że chodzi nie tyle o samą wielkość mózgu, co o proporcję mózg-ciało. Jednak nawet uściślając zależność, trzeba dodać, że nie jest to związek jasno potwierdzony badaniami. Niektóre wykazują pozytywną zależność, a inne poddają ją w wątpliwość. Mózg małpki sajmiri stanowi 5% jej masy ciała, a ludzki mózg 2%. Czy to oznacza, że małpka jest inteligentniejsza od człowieka? Nieścisłość występuje także w badaniach w obrębie ludzi. W badaniu przeprowadzonym przez Posthuma i in. z 2002 roku na parach bliźniąt stwierdzono odziedziczalność rozmiarów mózgowia oraz inteligencji. W innym z kolei z 2000 roku (Schoenemann, Budinger, Sarich, Wang) z udziałem pary sióstr wykazano bardzo słabą zależność między wyższym ilorazem inteligencji i posiadaniem większego mózgu [2].



## 3. Dostępne narzędzia oraz teoria

### 3.1. Mechanizmy skojarzeniowe w informatyce

Mechanizmy skojarzeniowe nie są jeszcze bardzo rozpowszechnione w stosowaniu. Wynika to zapewne z faktu, iż zasada ich działania opiera się na funkcjonowaniu ludzkiego mózgu. Jest to, jak wiadomo, organ, który mimo wielu badań pozostaje dla naukowców zagadką w dużym stopniu. Zbadanie struktury mózgu nie dało wszystkich odpowiedzi dotyczących procesów zachodzących w nim.

Aktualne pojęcia oparte na asocjacyjności:

- Tablice asocjacyjne jest to bardzo prosty przykład asocjacji, w którym w tablicach przechowywane są dane z unikalnym kluczem. Kluczem tym nie jest jednak indeks a jedno z danych przechowywanych w tablicy. Jest to powiązanie dwóch rodzajów danych przechowywane w tablicy.
- Reguły asocjacyjne są to dowolne implikacje. Jest to przykład relacji „jeżeli A to B” z założeniem braku części wspólnej. To kolejne bardzo proste użycie reguł skojarzeniowych.
- Klasy asocjacyjne to klasy, które są ze sobą powiązane, mogą oddziaływać na siebie nawzajem.
- Pamięci asocjacyjne HAM (ang. *Hopfield's Associative Memory*) to mechanizmy bazujące na sieciach Hopfielda, umożliwiające odtworzenie wzorców na podstawie niepełnych danych. Pamięć taka może odtworzyć wzorec, dane są ze sobą połączone w obrębie jednego wzorca.
- Dwukierunkowe pamięci asocjacyjne BAM (ang. *Bidirectional Associative Memory*) to pamięci, które mogą odtworzyć wzorec na podstawie innych wzorców, które mogą być niepełne lub zaszumione, ale nie są powtarzalne. W tym przypadku możliwe jest wywołanie jednego wzorca przez inny (połączenie 1:1), ale nie ma możliwości skojarzeń czy wnioskowania.
- Wielokierunkowe pamięci asocjacyjne MAM (ang. *Multidirectional Associative Memory*) to rozszerzona wersja pamięci BAM. Zakłada możliwość wywołania jednego wzorca przez wiele innych (połączenie N:1). Nie ma możliwości określenia kolejności, czy umiejscowienia w czasie wzorców. Nie ma możliwości kojarzenia i wnioskowania.
- Sieci Hamminga stosowane jako pamięci asocjacyjne, to mechanizmy klasyfikujące dane wejściowe. Odbywa się to na zasadzie porównania wektora wejściowego z wcześniej wprowadzonymi, zapamiętanymi, wzorcowymi wektorami i znalezieniu najmniej różniącego się.

- Autoasocjacyjne samoorganizujące się sieci neuronowe o wielu zwycięzcach MWSOINN (ang. *Multi Winners Self-Organizing Neural Network*) to sieci, które generują rozproszoną reprezentację (ślady pamięciowe) dla wzorców wejściowych, mogą się uczyć. Możliwe jest sterowanie ilością zwycięzców, zakres pobudzonych neuronów jest szeroki. Ślady pamięciowe bliskie sobie są pobudzane, a odległe hamowane. Brak tutaj możliwości zapisania sekwencji.
- Epizodyczne pamięci asocjacyjne EAM (ang. *Episodic Associative Memory*) mają możliwość zapamiętywania sekwencji czasowych oraz wzorców. Jednak pamięci te mają bardzo małą pojemność, co uniemożliwia formowanie się wiedzy, która potrzebuje pewnego zasobu danych, aby wyszukiwać rozwiązania.

### 3.2. Pozyskiwanie i wyszukiwanie wiedzy [8]

Zbyt mała ilość wiedzy w wielu przypadkach jest źródłem problemów. Jest to teza poprawna, gdyż nie mając podstaw informacyjnych nie ma na czym budować wiedzy, teorii i działań. Jednak w dzisiejszych czasach problemem jest także odwrotne zjawisko, czyli nadmiar wiedzy. Coraz większa ilość literatury, także cyfrowej i łatwo dostępnej, zwiększające się rozmiary nośników danych oraz praktycznie wszechobecny internet sprawiają, że prawie na każdym kroku możemy mieć swobodny dostęp do informacji na jakikolwiek temat. Wydawać by się mogło, że jest to pozytywne zjawisko. Jednak pojawia się tutaj nowy problem dotyczący wyszukiwania interesującej nas wiedzy w całym zbiorze danych.

Działania, które mają na celu rozwiązanie tego problemu nazywane są pozyskiwaniem lub odkrywaniem wiedzy w zbiorach danych (ang. *KDD-Knowledge Discovery in Databases*). Inne określenia na ten proces to ekstrakcja wiedzy lub *data mining*. Odkrywanie, kształtowanie wiedzy to połączenie wielu dziedzin naukowych takich jak: uczenie maszynowe, optymalizacja, statystyka, bazy danych, sztuczna inteligencja oraz obliczenia równoległe.

Pozyskiwanie wiedzy jest procesem etapowym, często o charakterze rekurencyjnym. Poszczególne fazy to:

- analiza problemu - poznanie danej dziedziny, aby dobrze wybrać metody pozyskiwania wiedzy;
- przetwarzanie wstępne i integracja danych - zbieranie danych z różnych źródeł i ich wstępne przetwarzanie (np. wybór przydatnych danych, usuwanie danych odbiegających od normy, interpretacja błędów);
- eksploracja danych - budowa modelu, czyli wybór danych, dobór wzorców i metod ich uzyskania (często są to drzewa decyzyjne, reguły asocjacji lub klasteryzacja), dobór wartości progowych;
- przygotowanie wyników dla potrzeb analizy i oceny - przedstawienie danych w przystępnej formie, łatwej do analizy.



### 3.3. Asocjacyjne relacje [1]

Zjawisko kojarzenia i myślenia w dany sposób uwarunkowane jest powiązaniem występującymi między danymi. Jeżeli przedstawi się komuś dwa podobne do siebie obiekty, z których tylko jeden jest mu znany, to stwierdzi on, że ten drugi obiekt jest podobny do znanego. Pewien zbiór cech aktywuje w połączeniach neuronowych ścieżkę rozpoznawania dla obiektu, który jest mu znany i w ten sposób można określić podobieństwo. Inne cechy, które nie pasują do całego wzorca, pozwolą stwierdzić, że nie jest to dokładnie ten sam przedmiot, bo czymś się jednak różnią. Nasz mózg potrafi uzupełnić wcześniej wyuczone wzorce, gdy występują w formie niekompletnej. Jeżeli dany wyraz jest nam najbardziej znany w połączeniu z innym, to gdy go usłyszymy, będziemy się spodziewać, że zaraz usłyszymy to drugie słowo. Te wszystkie zjawiska wynikają z faktu istnienia asocjacyjnych relacji pomiędzy obiektami, które są odwzorowane w grafie wiedzy poprzez połączenia asocjacyjne.

Asocjacyjne relacje można podzielić następująco:

- ASIM - powiązania asocjacyjnego podobieństwa (ang. *associative similarity*); to powiązania między neuronami, które są połączone z tymi samymi receptorami lub neuronami, reprezentują podobne cechy lub mają zbliżone wartości.
- ASEQ - powiązania asocjacyjnego następstwa (ang. *associative sequence*); to powiązania pomiędzy danymi, które występują chronologicznie po sobie, w konkretnej kolejności.
- ACON - powiązania asocjacyjnego kontekstu (ang. *associative context*); to bardziej „odległe” powiązania między danymi; charakteryzują asocjacje, które nie są wywoływane bezpośrednio od razu, ale mogą aktywować się w przyszłości. To kontekst, nie bezpośrednie pobudzenie. Aktywacja jednych neuronów prowadzi do pobudzenia następnych, które z kolei uaktywniają kolejne; to rodzaj pobudzenia wtórnego.
- ADEF - powiązania asocjacyjnego definiowania (ang. *associative defining*); to bezpośrednie połączenie od receptorów do neuronów reprezentujących określone cechy tego obiektu, który pobudził receptor.
- ASUP - powiązania asocjacyjnego tłumienia (ang. *associative suppression*); to połączenie danych dążące do ich skontrastowania, usunięcia redundancji i wyostrenia różnic.

### 3.4. Semassele

Pojęcie semassele zostało utworzone przez Doktora Adriana Horzyka z Akademii Górniczo-Hutniczej w trakcie badań naukowych dla określenia specyficznej informacji przechowywanej i dynamicznie zmieniającej się w neuronach - zdefiniowane w monografii [1]. Semassele (ang. *semantic associative elements*), czyli jednostki semantyczno-skojarzeniowe, to układy danych wejściowych na tyle intensywne i powtarzalne, że zdołały się utrwalić w neuronach systemu skojarzeniowego. Są to różne kom-

binacje impulsów wejściowych, które mogą pobudzić neuron w danym momencie. Impuls wejściowy, który składa się na kombinacje pochodzi od innych neuronów albo receptorów, więc impuls wejściowy zazwyczaj reprezentuje inne obiekty, ich cechy, fragmenty, wpływ na inne procesy asocjacyjne. Kombinacje reprezentowane przez semassele mogą być traktowane jako kombinacje innych obiektów, cech, części itd. Wynika stąd, że znaczenie, budowa semasseli jest definiowana przez inne presynaptyczne neurony i receptory z nimi połączone.

Semassele są związane z neuronem, ale mogą zmieniać się w czasie. Jeżeli semassel presynaptycznego neuronu się zmienia, to semassel postsynaptycznego neuronu także ulega zmianie. Dlatego też zmiany w reprezentacji neuronalnej (semasselach) powinny być raczej wolne i niewielkie, aby możliwy był proces asocjacji.

## 4. Matematyczne fundamenty pracy [1]

Praca bazuje na modelu matematycznym neuroasocjacyjnego grafu wiedzy stworzonym przez A. Horzyka.

Fundamentem systemów skojarzeniowych jest formowanie się wiedzy. Odpowiednie przetwarzanie informacji daje możliwość „wyłuskania” nowych skojarzeń, ścieżek myślowych. Formowanie się wiedzy następuje poprzez tworzenie asocjacji między pobudzonymi neuronami, łączeniu ich plastycznymi zależnościami. Odpowiednie sekwencje, wagi oraz sąsiedztwa są kodowane na podstawie wzoru danych źródłowych. Często powtarzające się sekwencje zostają utrwalone z każdym nowym pobudzeniem. Składowe, które nie są bezpośrednio ze sobą połączone, ale często występują w tym samym kontekście także zostaną powiązane.

### 4.1. Neuroasocjacyjne grafy wiedzy ANAKG

Reprezentacja wiedzy za pomocą grafu ANAKG (*ang. active neuroassociative knowledge graph*) pozwala przechować ją w taki sposób, aby zawrzeć jak najwięcej zależności między danymi ją formującymi. Poszczególne neurony w grafie wiedzy są utrwalane poprzez kolejne aktywacje. Aktywowane mogą być także całe łańcuchy, w przypadku ponownego wystąpienia danej sekwencji lub kaskadowego pobudzenia. W swej monografii A. Horzyk wprowadził pojęcie wiedzy skojarzeniowej, która tworzy się na podstawie kombinacji danych, układu sekwencji powstałych na podstawie pewnego zbioru uczącego. Aby zamodelować wiedzę skojarzeniową niezbędny jest również system skojarzeniowy, który będzie w odpowiedni sposób przechowywał informacje (semassele) tak, aby zachowane zostały odpowiednie zależności, powiązania, asocjacje i oddziaływania.

Realizacją takiego systemu skojarzeniowego jest aktywny neuroasocjacyjny graf wiedzy ANAKG, który umożliwia łączenie ze sobą neuronów oraz wzmacnianie połączeń między nimi na asocjacyjnej ścieżce aktywacji wyznaczonej przez łańcuchy uczące, pobudzające neurony w odpowiedniej kolejności. Neurony mogą być pobudzane dwojako. Pierwszym sposobem jest odczytywanie sekwencji pobudzającej i tym samym tworzenie jeszcze nie występujących w grafie neuronów lub wzmacnianie istniejących. Drugą drogą pobudzenia neuronu jest jego aktywowanie „nie wprost”. Ma to miejsce w momencie, gdy neuron nie jest pobudzony bezpośrednio, a aktywowany jest neuron poprzedzający go w sieci grafu, taki, który się z nim łączy. Pobudzenie następuje kaskadowo, jako swoiste echo aktywacji związanego z nim

neuronu. W efekcie wczytywania kolejnych łańcuchów powstają w grafie połączenia odzwierciedlające podzbiory relacji typu ACON i ASEQ.

W modelu ANAKG wyznaczany jest współczynnik aktywności połączenia. Jest on zależny od upływu czasu pomiędzy pobudzeniem neuronu presynaptycznego (od którego wychodzi sygnał) i postsynaptycznego (odbiorczego) (wzór 4.1).

$$\tau = t_{postsyn} - t_{presyn} \quad (4.1)$$

Jeżeli pobudzony zostanie neuron presynaptyczny, a po upływie danego czasu w konsekwencji aktywowany zostanie neuron postsynaptyczny, to współczynnik aktywności ich połączenia wyliczony zostanie ze wzoru (4.2).

$$\delta_{SN, \widehat{SN}}^{act} = \delta_{SN, \widehat{SN}}^{act} + \frac{1}{t_{postsyn} - t_{presyn}} = \delta_{SN, \widehat{SN}}^{act} + \frac{1}{\tau} \quad (4.2)$$

Czas brany pod uwagę przy obliczeniach jest wartością dyskretną, co skutkuje tym, że odstęp czasu pobudzeń jest tak naprawdę interpretowany jako odstęp w kolejności, odległość krokowa pomiędzy wczytaniem dwóch wyrazów. Im większy odstęp czasu, czyli większy kontekst ACON, tym mniejszy współczynnik aktywności połączenia dwóch neuronów. Gdy połączenie jest ponownie wzbudzone, współczynniki z wszystkich poprzednich aktywacji są sumowane (wzór 4.3).

$$\delta_{SN, \widehat{SN}}^{act} = \sum_{\{\rightsquigarrow ACON_{\tau}: SN \rightsquigarrow \dots \rightsquigarrow \widehat{SN} \in AAT\}} \frac{1}{\tau} \quad (4.3)$$

Sprowadzenie upływu czasu w tworzeniu grafu ANAKG do odległości we wczytywanej sekwencji nie odzwierciedla niestety realnego upływu czasu, tak jak to dzieje się w przypadku pracy biologicznej sieci neuronowej i zjawiska zapominania informacji, które przyswoiło się jakiś czas temu. Jest to uogólnienie, które pozwala łatwiej zaimplementować ten model w programie komputerowym. Takie uogólnienie może jednak przynieść korzyści w postaci stworzenia nowego skojarzenia na podstawie uogólnionych danych, które nie byłoby dostępne w przypadku pracy nad realnymi danymi.

W trakcie tworzenia grafu ANAKG w rezultacie kaskadowego wewnętrznego pobudzania neuronów nie związanych bezpośrednio ze sobą może dojść do utworzenia połączenia, które w danych źródłowych nie występowało. W ten sposób dochodzi do utworzenia sztucznego skojarzenia. W momencie, gdy dochodzi do wewnętrznej aktywacji sekwencji neuronów na skutek zewnętrznego pobudzenia, to te neurony oraz połączenia pomiędzy nimi tworzą asocjacyjną ścieżkę aktywacji. Ścieżki te tworzą się i zmieniają poprzez kształtowanie grafu ANAKG w oparciu o dane wejściowe, uczące. Połączenia między neuronami i ich wagi są wyliczane ze wzoru (4.4).

$$w_{SN, \widehat{SN}}^{ACON} = \frac{2 \cdot \delta_{SN, \widehat{SN}}^{act}}{\eta_{SN}^{act} + \delta_{SN, \widehat{SN}}^{act}} \quad (4.4)$$

W modelowaniu wiedzy z użyciem wyżej wymienionych zależności zawarte jest także pojęcie relaksacji. Neuron po pobudzeniu ulega relaksacji zgodnie ze wzorem (4.5).

$$\begin{aligned}
exc_t^{SN} = NRF_{\alpha,\beta}^{SQR}(exc_{t-1}^{SN}, \Theta^{SN}, x_1^t, \dots, x_K^t) &= \sum_{k=1}^K w_k * x_k^t + \\
+ \begin{cases} \alpha * exc_{t-1}^{SN} + \frac{(\alpha-1)*\beta*(exc_{t-1}^{SN})^2}{\Theta^{SN}} & 0 > exc_{t-1}^{SN} \\ \alpha * exc_{t-1}^{SN} + \frac{(1-\alpha)*\beta*(exc_{t-1}^{SN})^2}{\Theta^{SN}} & 0 \leq exc_{t-1}^{SN} < \Theta^{SN} \\ -\beta * \Theta^{SN} & exc_{t-1}^{SN} \geq \Theta^{SN} \end{cases} & \quad (4.5)
\end{aligned}$$

Powyższe wzory i zależności zostały wykorzystane do budowy grafu wiedzy skojarzeniowej ANAKG, co zostało opisane w rozdziale 5.



## 5. Realizacja tematu

W ramach niniejszej pracy napisano program, korzystając ze środowiska Visual Studio 2010 Express. Używanym językiem programowania był C#. W czasie pisania korzystano z następujących źródeł dotyczących programowania w tym języku: <https://msdn.microsoft.com/pl-pl/library/> [13], „Rusz głową! C#” [4]

Zbudowany asocjacyjny, internetowy, kontekstowy silnik wyszukiwania informacji z użyciem sztucznych systemów skojarzeniowych jest programem, który można podzielić na trzy części. Pierwsza z nich to część odpowiedzialna za przeszukiwanie stron internetowych oraz wyłuskiwanie z nich treści. Jest to rodzaj pająka internetowego zbierającego konkretną treść z internetu. Druga część to budowa grafu wiedzy ANAKG na podstawie zgromadzonej wiedzy z części pierwszej. Graf ten budowany jest na podstawie wzorów i teorii przedłożonych w rozdziale 4 tej pracy. Po przetworzeniu wcześniej zebranej wiedzy, używając wspomnianych wzorów i zależności, uzyskano grafu wiedzy, słowa połączone relacjami kontekstu ACON oraz sekwencji ASEQ. Ostatnią część stanowi wyszukiwanie odpowiedzi na zadane przez użytkownika pytanie. Wprowadzone zapytanie używane jest jako sygnał wejściowy pobudzający graf wiedzy. Na podstawie jego reakcji zwracana jest odpowiedź. W pracy przetestowano dwie wersje programu: jedna, w której do grafu przekazywane są tylko teksty zebrane ze stron oraz druga, gdzie oprócz tego w graf wplecione są hiperłącza źródłowe.

### 5.1. Część sieciowa

Jest to część odpowiedzialna za agregowanie wiedzy na zadany temat. Ze względu na zbyt skomplikowany proces tworzenia mechanizmów wykorzystywanych w wyszukiwarkach internetowych, wykorzystano gotową wyszukiwarkę Google. Mechanizm przeszukiwania wyników jest swego rodzaju pająkiem internetowym. Interfejs użytkownika (rys. 5.1) jest bardzo prostym formularzem do uzupełnienia w sieci web. Wprowadzone dane są następnie przesyłane do programu, który uruchamia przeszukiwanie sieci z użyciem zadanych parametrów. Mechanizm przeszukiwania i zbierania treści z poszczególnych stron został opisany poniżej. Wynik działania tej części programu agregowany jest w jednej zmiennej oraz jest daną wejściową dla kolejnej części programu (5.2).

Rysunek 5.1: Okno użytkownika

### 5.1.1. Budowa

Tą część reprezentuje klasa *WebForm1* (rys. 5.2) oraz *PageAnalyze*, której schemat obrazuje (rys. 5.3).

<b>WebForm1</b>
<code>+Container {tekst: string[]; link: string[]}; struct</code>
<code>#Button1_Click(sender:object,e:EventArgs): void</code>

Rysunek 5.2: Budowa klasy WebForm1

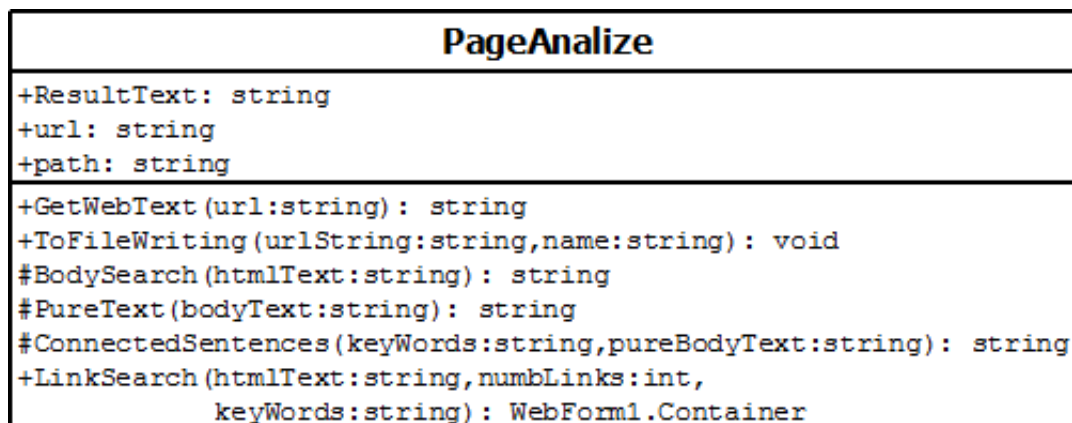
Klasa *WebForm1* obejmuje strukturę o nazwie *Container*, która składa się z tablicy zmiennych typu string o nazwie *tekst* (w niej przechowywane będą teksty wyizolowane ze stron internetowych). W drugiej metodzie ta struktura posiada również drugą tablicę zmiennych typu string o nazwie *link* (ta tablica zawierać będzie odpowiadające tekstom hiperłącza źródłowe).

Oprócz tego w tej klasie znajduje się metoda *Button1\_Click*, która jest odpowiedzialna za obsługę zdarzenia (kliknięcia przez użytkownika przycisku „Myśl” (rys. 5.1)). W przypadku nieustawienia ilości poszukiwanych linków program domyślnie wstawia tam wartość równą 3, a po zakończeniu działania części Grafu i Odpowiedzi wyświetla wyniki na stronie.

Klasa *PageAnalyze* zawiera metody pozwalające na etapowe wyluskiwanie szukanych treści ze stron internetowych.

*GetWebText* to funkcja zbierająca do zmiennej typu string całą zawartość danej strony. Jest ona wywoływana na początku, aby zapisać stronę wynikową Google. Później wykorzystywana jest do spisywa-





Rysunek 5.3: Budowa klasy PageAnalyze

nia zawartości stron z poszczególnych linków wynikowych, które zostały wybrane do dalszego przetwarzania.

*BodySearch* to pierwszy etap „okrajania” przetwarzanego tekstu. Polega on na wyszukaniu fragmentu znajdującego się pomiędzy znacznikami <body>.

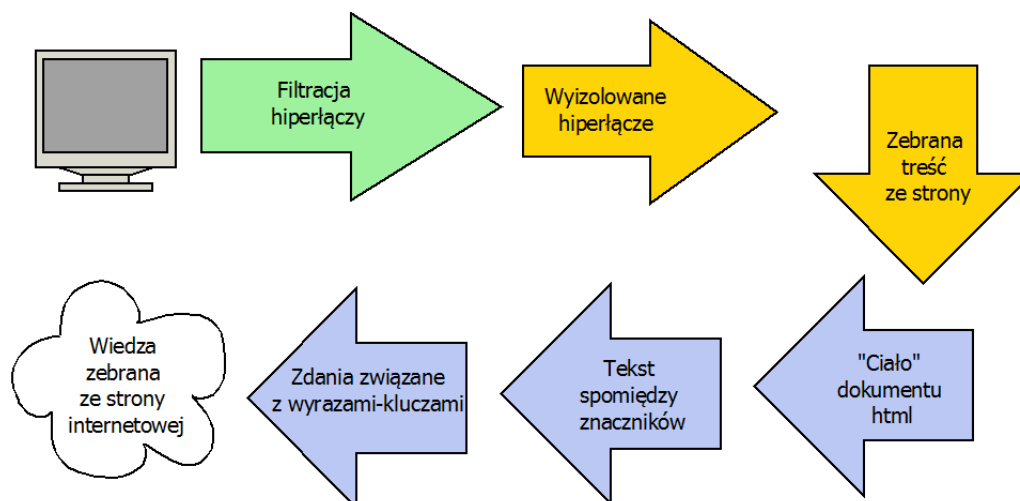
*PureText* to druga faza przetwarzania tekstu ze strony. Ta metoda przeszukuje tekst pod kątem występowania wszystkich pozostałych znaczników, a następnie wybiera treść, która znajduje się pomiędzy nimi.

*ConnectedSentences* to ostatni etap filtracji tekstu. Polega on na zebraniu tylko tych zdań, które zawierają któryś z wyrazów kluczy. Jako klucze traktowane są wyrazy z tematu wprowadzonego przez użytkownika. Dodatkowo wprowadzono ograniczenie filtrowanych treści do zdań nie dłuższych niż 200 znaków. Pozwala to wyeliminować ciągi znaków, które są fragmentem kodu, a nie treścią.

*LinkSearch* to kluczowa w tej klasie metoda. Wyszukuje fragmenty tekstu dotyczące hiperłączy, poprzez wyszukanie ciągu "http". Następnie, od tego miejsca do przewidywanego końca linka, łańcuch uważany za hiperłączy przepisywany jest do bufora pomocniczego. Kolejnym etapem jest przepuszczenie bufora przez filtry kwalifikujące hiperłączy do dalszego wykorzystania. Jako niepoprawny traktowany jest link, który zawiera słowa: schema, image, settings, google, facebook, youtube, jpg, php. Jeżeli hiperlink spełnia powyższe warunki to wywoływany jest on funkcją *GetWebText*, następnie oczyszczany przez *PureText* oraz *ConnectedSentences*. Wynikowy tekst zapisywany jest do zmiennej typu *Container*, w którym zależnie od wersji programu przechowuje się jedną lub dwie tablice. Równolegle zapisuje się do nich tekst zebrany ze strony oraz alternatywnie hiperłączy, które do niego prowadzi.

*ToFileWriting* to metoda pomocnicza, która odpowiedzialna jest za zapisywanie danych tekstowych do pliku tekstowego. Używana była w wielu miejscach programu, jako narzędzie kontroli poprawności działania kodu.

Schematyczną ścieżkę pracy tej części programu obrazuje rys. 5.4.



Rysunek 5.4: Ścieżka zbierania i przetwarzania wiedzy ze strony internetowej.

### 5.1.2. Działanie

Użytkownik wpisuje w oknie nr 1 (rys. 5.1) interesujący go temat, zagadnienie.

Następnie w oknie nr 2 (rys. 5.1) precyzuje pytanie oraz w polu nr 3 (rys. 5.1) wyznacza liczbę linków, które program ma brać pod uwagę w zbieraniu danych. Następnie po wciśnięciu przycisku "Myśl", nr 4 (rys. 5.1) zostaje uruchomiony "pająk internetowy". Polega to na tym, że wpisane w polu nr 1 zagadnienie zostaje automatycznie wyszukane w sieci za pomocą wyszukiwarki Google. Następnie analizowana jest treść strony wynikowej. Zbierane są „wartościowe” linki. Pod tym pojęciem rozumie się hiperłącza do stron wynikowych z ominięciem tych, które prowadzą do zdjęć, są reklamami Google i opcjami tej wyszukiwarki. Kiedy program zbierze wymaganą, podaną w polu nr 3 liczbę linków, każdy z nich jest wywoływany i analizowana jest treść stron wynikowych. Ograniczono tutaj zakres agregowanych zdań zebranych ze strony do takich, które zawierają słowa z zadanego tematu wyszukiwania oraz nie przekraczają 200 znaków. Zebrane zdania są wpisywane w bufor i przekazywane do dalszej analizy, opcjonalnie wraz z informacją z których stron internetowych zostały zebrane.

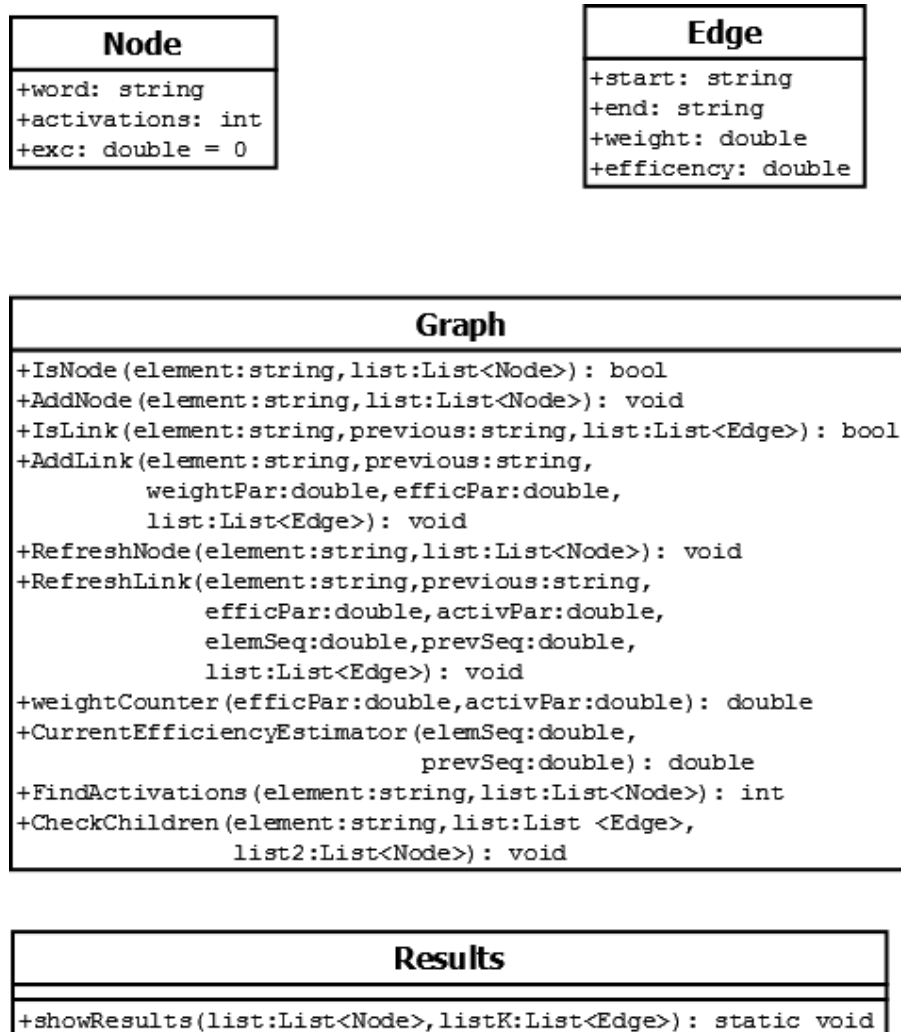
## 5.2. Część grafu

Jest to druga część programu, która odpowiada za zbudowanie grafu asocjacyjnej wiedzy na podstawie zagregowanych danych ze stron internetowych dostarczonych przez pierwszą część. Od napisania tej części zaczęto pracę nad programem, jako że jest ona kluczowa. Bazą, na której oparto działanie kodu odpowiedzialnego za budowę grafu wiedzy skojarzeniowej ANAKG jest model stworzony przez doktora Adriana Horzyka [1]. Zebrany tekst (mechanizm opisano w 5.1) jest dzielony na zdania, a zdania na poszczególne wyrazy, które są umieszczane w odpowiednich miejscach grafu, tworzone są powiązania między nimi i wyliczane odpowiednie parametry.

### 5.2.1. Budowa

W tej części utworzono kilka klas (rys. 5.5), które zawierają metody pozwalające na utworzenie grafu ANAKG.

Klasy te zostały skonstruowane na podstawie analizy wzorów matematycznych podanych w rozdziale 4.



Rysunek 5.5: Klasy utworzone na potrzeby budowania grafu wiedzy asocjacyjnej ANAKG

Klasa *Node* reprezentuje pojedynczy węzeł sieci, czyli wyraz. Do jego opisu potrzebna jest wartość tekstowa typu *string* oraz wartość numeryczna (liczba całkowita), która przechowuje informację o ilości aktywacji tego węzła. Zmienna tej klasy przechowuje również informację o wartości pobudzenia neuronu jako *exc* typu *double*.

Klasa *Edge* jest reprezentacją połączenia asocjacyjnego, czyli krawędzi w grafie łączącej dwa węzły. W tym przypadku wartościami przechowywanymi są: wyraz/węzeł od którego biegnie połączenie, wyraz/węzeł który jest z nim połączony, waga połączenia (wzór 4.4) oraz jego efektywność (wzór 4.3).

Ostatnią, najbardziej rozbudowaną klasą jest klasa *Graph*. Składa się ona z kilku metod.

*IsNode* sprawdza czy dany węzeł już istnieje w grafie.

*AddNode* dodaje nowy węzeł do sieci oraz ustawia początkową ilość aktywacji na 1.

*IsLink* sprawdza czy istnieje już połączenie pomiędzy podanymi dwoma wyrazami.

*AddLink* tworzy nową krawędź oraz ustawia jej parametry: wagę i efektywność.

*WeightCounter* oblicza wagę połączenia po podaniu ilości aktywacji neuronu presynaptycznego oraz współczynnika efektywności połączenia.

*CurrentEfficiencyEstimator* podaje obliczony współczynnik efektywności połączenia na podstawie odstępu czasowego neuronów (co jest tak naprawdę odległością w sekwencji).

*FindActivations* zwraca liczbę aktywacji szukanego neuronu.

*RefreshNode* uaktualni parametry przy ponownym aktywowaniu danego węzła.

*RefreshLink* aktualizuje parametry wcześniej utworzonej krawędzi przy ponownym jej wywołaniu.

*CheckChildren* jest to metoda sprawdzająca czy istnieją w grafie dalsze połączenia, biegnące od aktywowanego neuronu postsynaptycznego. Jeżeli tak, to zmieniana jest waga takich połączeń.

### 5.2.2. Działanie

Pierwszym krokiem w tej części programu jest przetworzenie tekstu ciągłego zebranego w części 5.1 do postaci, w której łatwo będzie go dalej konwertować na graf wiedzy ANAKG. Początkowo jest on rozbijany na zdania przechowywane w tablicy. Następnie każdy element tej tablicy jest rozdzielany na pojedyncze słowa, które ostatecznie przechowywane są w formie listy. W drugiej metodzie na koniec każdej listy dodawany jest link źródłowy. Każda ukończona lista wkładana jest na stos, który na zakończenie przetwarzania tekstu zebranego ze stron internetowych jest odwracany tak, aby kolejność zdań była zachowana. Ostatecznie tekst pierwotnie ciągły jest zapisany jako stos (cały tekst) list (zdania) wartości typu string (wyrazy).

Mechanizm budowy grafu ANAKG opiera się na wzorach przytoczonych w rozdziale 4. Można go jednak przybliżyć w formie schematu blokowego (rys. 5.6), który przedstawia „ścieżkę myślową” realizowaną w trakcie budowy grafu. W przypadku pierwszych dwóch wyrazów z pierwszego zdania są one dodawane jako nowe, bez uprzedniego sprawdzenia czy już istnieją w grafie. Tworzone jest także nowe połączenie między nimi. Każdy kolejny wyraz z zagregowanego tekstu ze stron internetowych zostaje poddany analizie. W zależności od tego czy dany wyraz występuje już w grafie czy nie, odpowiednio zwiększa się jego ilość aktywacji lub tworzy nowy neuron ze współczynnikiem równym 1. Następnie analizuje się połączenia wychodzące od tego neuronu do neuronu postsynaptycznego. Jeżeli połączenie istnieje, odświeża się jego parametry, jeżeli nie istnieje, tworzy się je. Ostatnim krokiem jest odświeżenie wszystkich istniejących powiązań i kaskadowa aktualizacja parametrów po wewnętrznej aktywacji.

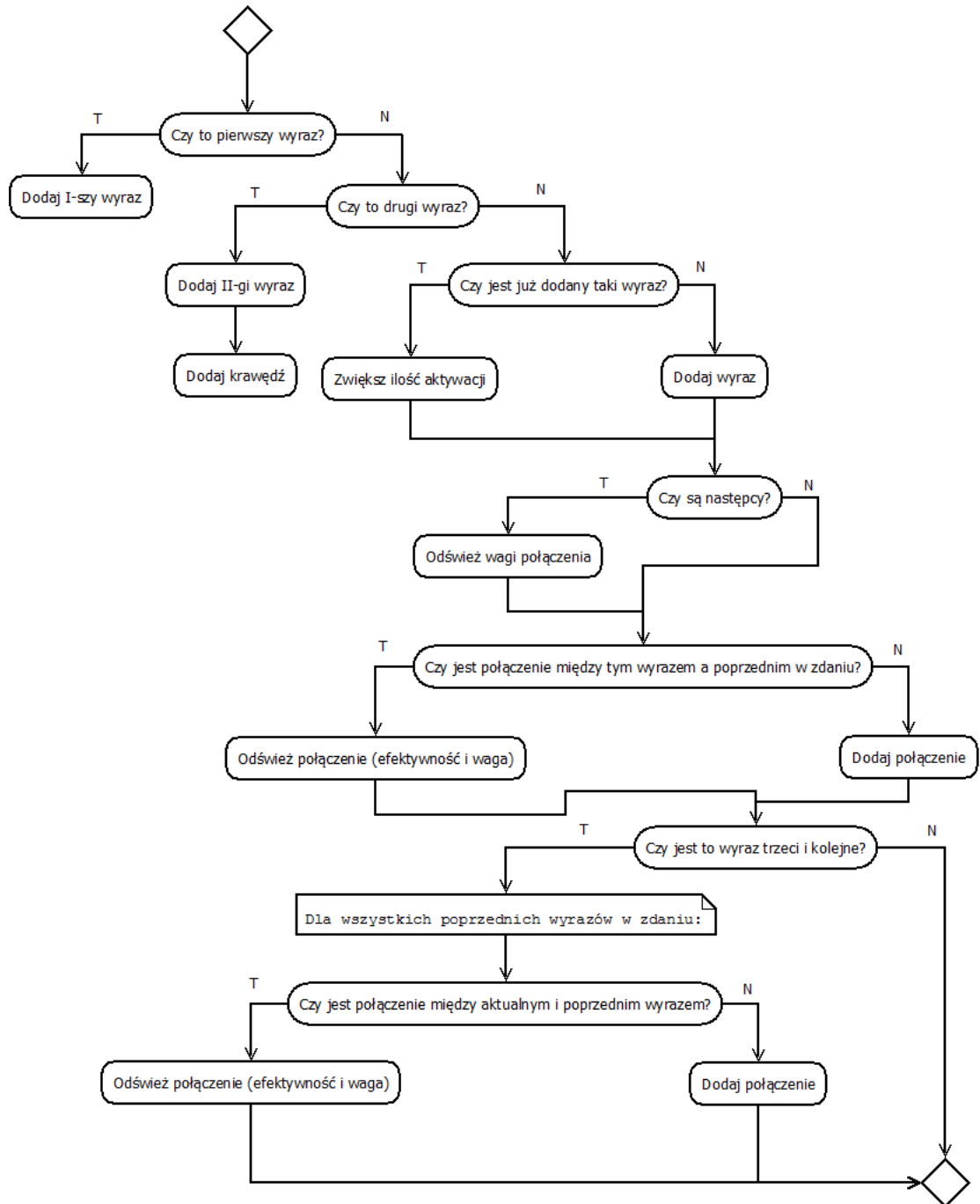
Działanie zgodne z założeniami sprawdzono na przykładzie przeanalizowanym w [1] oraz [10].

W tym celu wczytano z pliku tekstowego następujące zdania:

*Ala ma kota.*

*Jan Nowak ma psa.*

*Ewa Nowak ma liście.*



Rysunek 5.6: Schemat blokowy przedstawiający ścieżkę budowy grafu wiedzy ANAKG

*Ewa lubi liście.*

*Jan ma kota.*

Zdania te są tak skonstruowane, żeby odpowiadały wczytywanym sekwencjom S1, S2, S3, S4, S5 z przykładu. Występuje tutaj następująca zależność: *Ala -> A1 ma -> A2 kota -> A3 Jan -> A4 Nowak -> A5 psa -> A6 Ewa -> A7 liście -> A8 lubi -> A9*

W wyniku działania części programu odpowiedzialnej za budowę grafu otrzymano następującą odpowiedź (tab. 5.1), (tab. 5.2):

Liczba węzłów : 9, liczba krawędzi : 17

Tablica 5.1: Wartości współczynnika  $\eta$  dla poszczególnych węzłów

Poszczególne węzły:	
nazwa węzła	$\eta$
Ala	1
ma	4
kota	2
Jan	2
Nowak	2
psa	1
Ewa	2
liście	2
lubi	1

Tablica 5.2: Współczynniki krawędzi

Poszczególne krawędzie:			
neuron presynaptyczny	neuron postsynaptyczny	waga	delta
Ala	ma	1	1
ma	kota	0,67	2
Ala	kota	0,67	0,5
Jan	Nowak	0,67	1
Nowak	ma	1	2
Jan	ma	0,86	1,5
ma	psa	0,4	1
Nowak	psa	0,4	0,5
Jan	psa	0,29	0,3333
Ewa	Nowak	0,67	1
Ewa	ma	0,4	0,5
ma	liście	0,4	1
Nowak	liście	0,4	0,5
Ewa	liście	0,59	0,8333
Ewa	lubi	0,67	1
lubi	liście	1	1
Jan	kota	0,4	0,5

Powyższe wyniki są zgodne z tymi przedstawionymi w przykładzie (rys. 5.7)

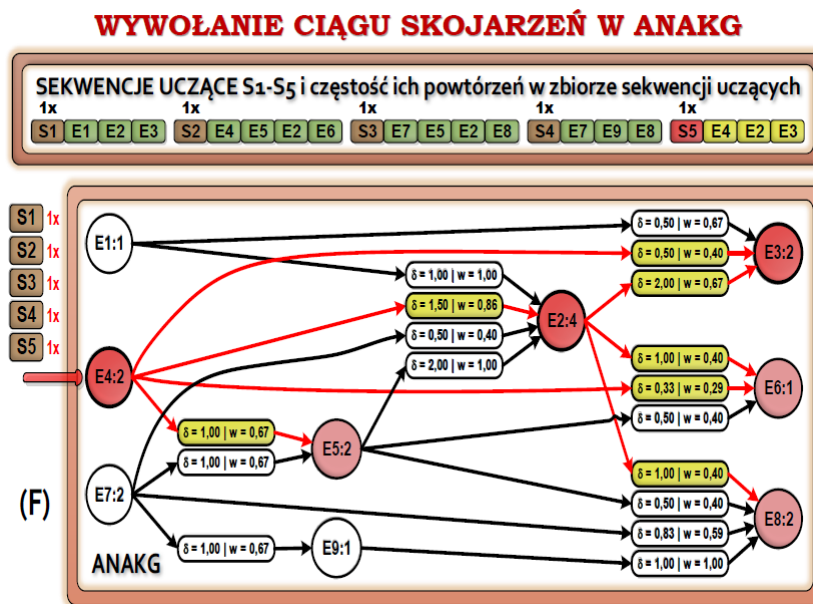
### 5.3. Część odpowiedzi

Jest to ostatnia część programu. W tym miejscu następuje wyszukanie w grafie odpowiedzi na zadane przez użytkownika konkretne pytanie wpisane w polu nr 2 (rys. 5.1).

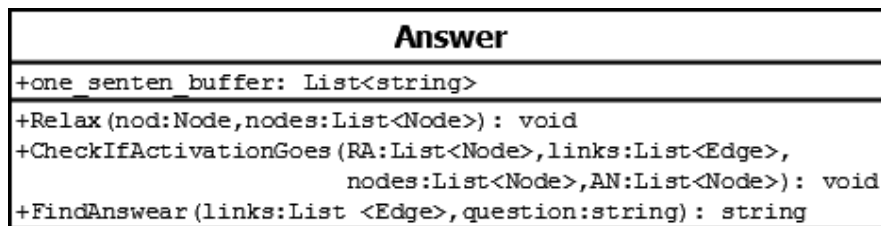
#### 5.3.1. Budowa i działanie

W tej części zbudowano klasę *Answer* (rys. 5.8), która zawiera metody: *Relax*, *CheckIfActivationGoes* oraz *FindAnswer*. Metoda *Relax* jest zastosowaniem wzoru na relaksację (4.5) dla neuronu przekazanego jako parametr.

Pierwszym krokiem w funkcji *FindAnswer* jest rozdzielenie pytania zadanego przez użytkownika na poszczególne wyrazy. Następnie realizowane jest kilkakrotne pobudzanie grafu pytaniem. Schemat jednokrotnego pobudzania sieci pytaniem zamieszczono na rys. 5.9. Pojedynczego pobudzanie grafu pytaniem polega na jednokrotnym pobudzeniu sieci poszczególnymi wyrazami pytania w kolejnych chwilach



Rysunek 5.7: Przykład wywołania ciągu skojarzeń - graf typu ANAKG ([10])

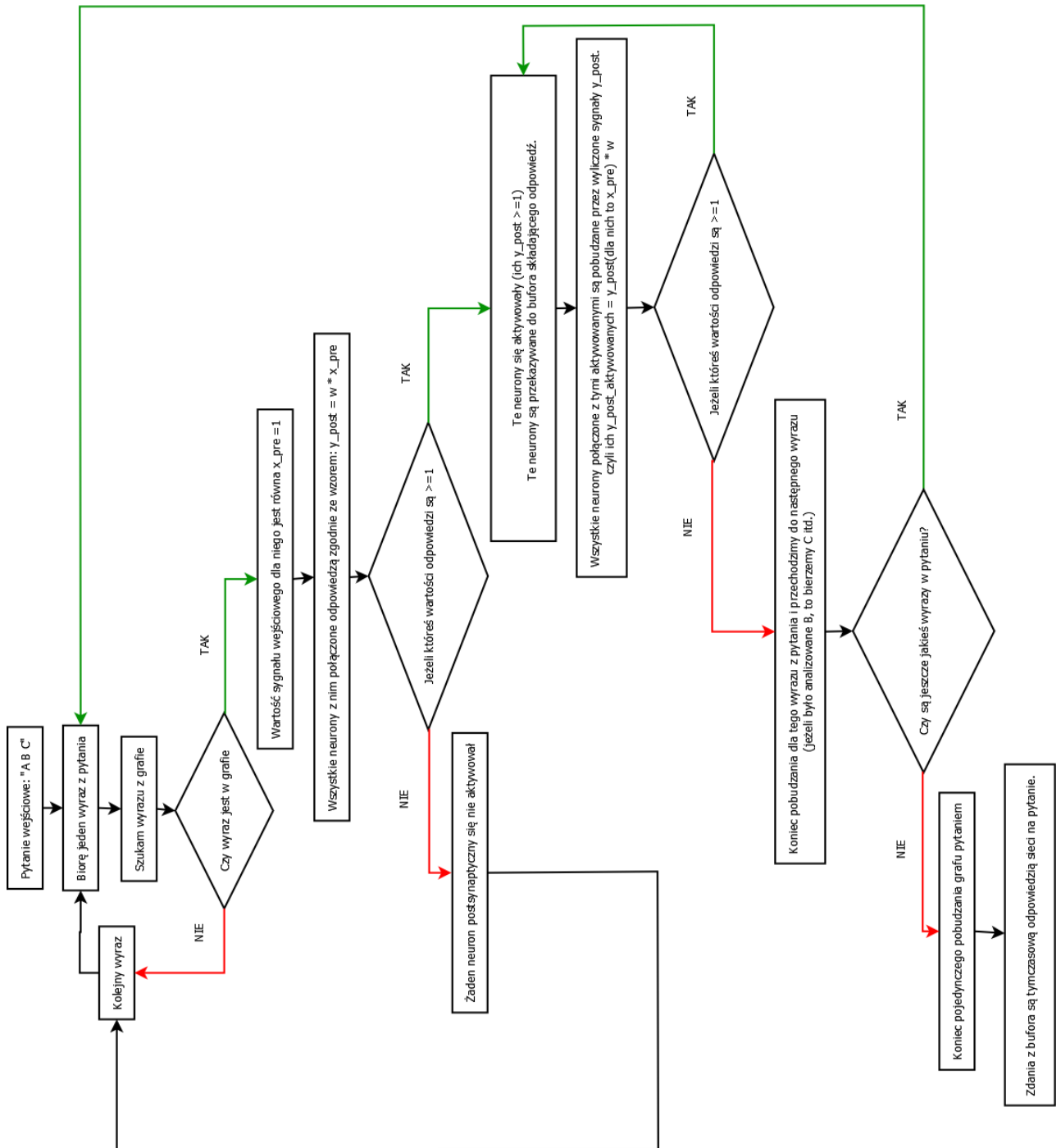


Rysunek 5.8: Klasa Answer

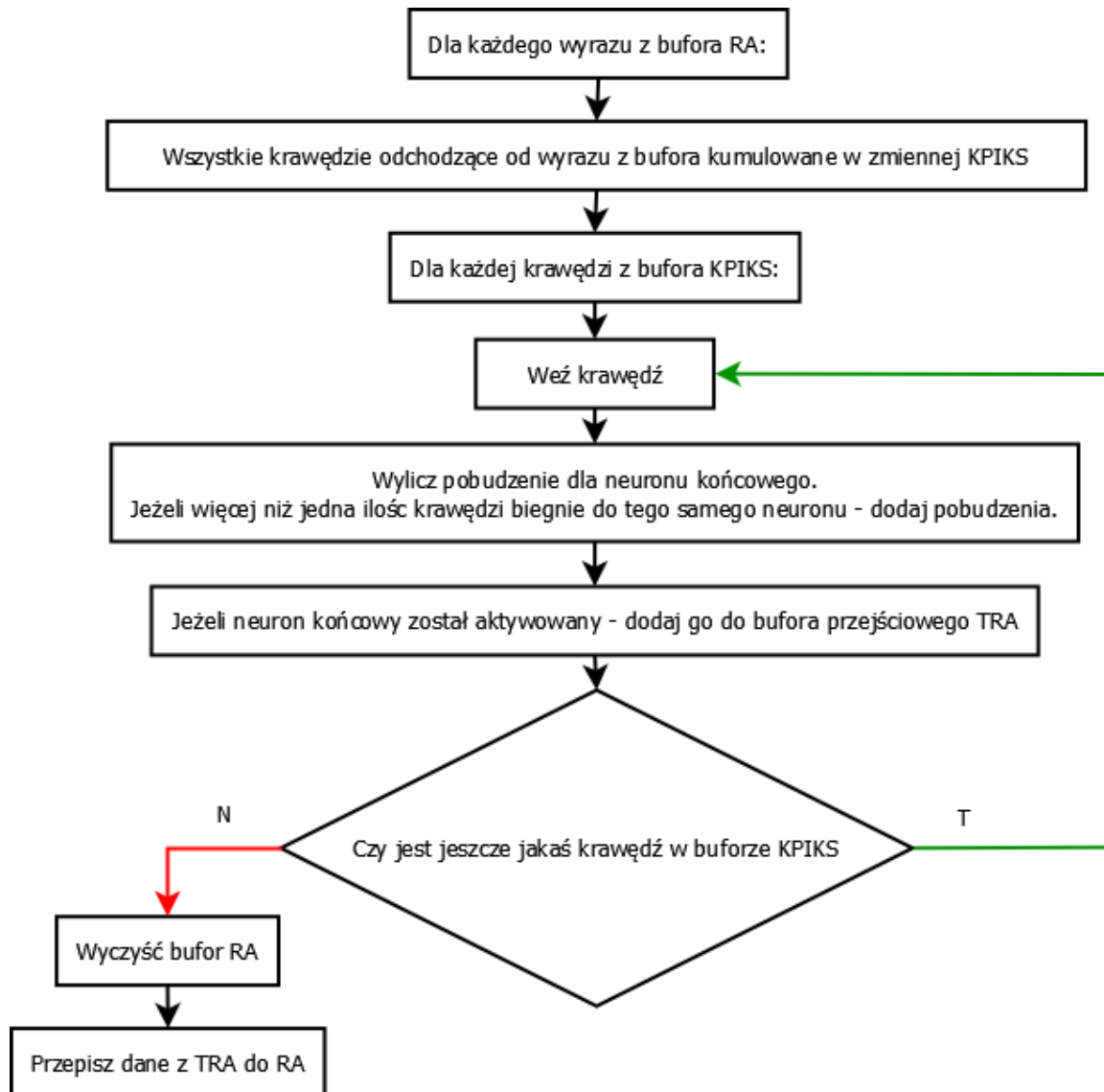
czasowych. Jeżeli pobudzony neuron pytania został aktywowany, to neurony z nim połączone zostaną pobudzone w kolejnej chwili czasowej, razem z następnym wyrazem z pytania. Aktywowane neurony tworzą odpowiedź sieci.

Funkcja *CheckIfActivationGoes* pobudza neurony połączone z węzłami przekazanymi w buforze RA -*Recently Activated* (takimi, które zostały aktywowane w poprzedniej chwili czasowej). Następnie sprawdza czy któryś z tych neuronów został aktywowany (czy ich pobudzenie przekroczyło zadany próg). Jeżeli tak, to zostają one umieszczone w wyczyszczonym buforze RA, a neurony z nimi połączone zostaną pobudzone w kolejnej chwili czasowej (przy następnym wywołaniu tej funkcji). Schemat działania funkcji przedstawiono na rys. 5.10





Rysunek 5.9: Schemat przedstawiający jednokrotne pobudzenie sieci pytaniem użytkownika



Rysunek 5.10: Schemat przedstawiający działanie funkcji *CheckIfActivationGoes*

## 6. Wyniki

### 6.1. Pollyanna

Przykładem, na którym zobrazowano działanie programu jest „Pollyanna”. Dane wprowadzone do programu przedstawiono na rys. 6.1

Treść wczytana ze stron internetowych powtarzała się w każdym z czterech linków, ponieważ wszystkie pochodziły z jednego źródła:

*http : //pl.wikipedia.org/wiki/Pollyanna*

*https : //pl.wikipedia.org/wiki/ < b > Pollyanna < /b >< /cite >< divclass =*

*http : //pl.wikipedia.org/wiki/Pollyanna%23Fabu.C5.82a\_powie.C5.9Bci*

*http : //pl.wikipedia.org/wiki/Pollyanna%23Wybrane\_e\_kranizacje*

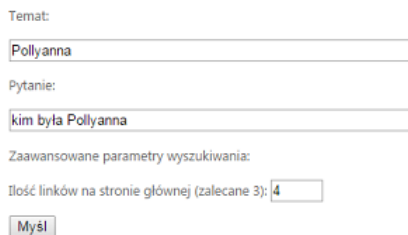
Graf wiedzy ANAKG budowano na tekście który był zarysem fabuły książki.

Przeprowadzono próby pobudzania sieci dla różnych ilości powtórzeń pytania. W poniższych podrozdziałach umieszczone zostały otrzymane wyniki.

#### 6.1.1. Jednokrotne pobudzanie pytaniem

była Pollyanna

Przy jednokrotnym pobudzeniu aktywacji uległy tylko zewnętrznie pobudzane neurony wyrazów zapytania.



Temat:  
Pollyanna

Pytanie:  
kim była Pollyanna

Zaawansowane parametry wyszukiwania:  
Ilość linków na stronie głównej (zalecane 3): 4

Myśl

Rysunek 6.1: Dane wprowadzone przez użytkownika - przykład Pollyanna

### 6.1.2. Dwukrotne pobudzenie pytaniem

była Pollyanna

Przy dwukrotnym pobudzeniu, nie widać różnicy w wynikach. Jest to spowodowane zjawiskiem refrakcji neuronów zapytania, których stan pobudzenia na skutek aktywacji ich w pierwszej pętli jest ujemny i ponowne pobudzenie sygnałem nie jest w stanie ich aktywować, ponieważ proces relaksacji tych neuronów nie przebiega na tyle szybko, aby powróciły one do stanu początkowego (stan pobudzenia = 0).

### 6.1.3. Trzykrotne pobudzenie pytaniem

była Pollyanna była bardzo udziałem Pollyanna

Za trzecim pobudzeniem całym pytaniem dochodzi do ponownej aktywacji neuronów zapytania, i zwiększenia stanu pobudzenia neuronów z nimi połączonych. Pobudzenia z pierwszej i trzeciej pętli się sumują (odpowiednio zmniejszone poprzez relaksację), w wyniku czego dodatkowe neurony ulegają aktywacji.

### 6.1.4. Czterokrotne pobudzenie pytaniem

Fragment odpowiedzi:

była Pollyanna była bardzo udziałem Pollyanna smutna;  
 jej *https* : *//pl.wikipedia.org/wiki/Pollyanna* panny Polly *https* :  
*//pl.wikipedia.org/wiki/Pollyanna* *https* : *//pl.wikipedia.org/wiki/Pollyanna* Polly  
 jej smutna; ojciec jednak panny jej ojciec jednak powiedział, że  
 powinna się ona cieszyć, gdyż nie o *https* : *//pl.wikipedia.org/wiki/Pollyanna*  
 i o *https* : *//pl.wikipedia.org/wiki/Pollyanna* ojciec jednak powiedział, że  
 powinna się ona cieszyć, gdyż nie musi nich chodzić panny Polly Polly  
*https* : *//pl.wikipedia.org/wiki/Pollyanna* i *https* : *//pl.wikipedia.org/wiki/Pollyanna*  
 przychodziła jej obiad i Anna *https* : *//pl.wikipedia.org/wiki/Pollyanna*  
 przychodziła wraz z Nancy jej obiad przestrzegać i o *https* :  
*//pl.wikipedia.org/wiki/Pollyanna* ojciec jednak powiedział, że powinna  
 się ona cieszyć, gdyż nie musi nich chodzić ciotki Nancy panny Polly  
 obiad włosy jej ojciec jednak powiedział, że powinna się ona cieszyć,  
 gdyż nie musi o nich chodzić *https* : *//pl.wikipedia.org/wiki/Pollyanna* ...

W tym przypadku aktywacja sieci dotarła do węzłów zawierających hiperłącze. Cała odpowiedź zawiera 2 626 wyrazów. Należy wziąć jednak pod uwagę fakt, iż jest to odpowiedź rozwijająca się. Jeżeli jeden neuron był kilka razy aktywowany, to tyle razy wpisywany był do bufora odpowiedzi. W powyższym fragmencie można zauważyć stopniowe rozbudowywanie odpowiedzi.

Wynik dla wpisywania do bufora niepowtarzalnych neuronów:

była Pollyanna bardzo udziałem smutna; jej *https* :  
 //pl.wikipedia.org/wiki/Pollyanna panny Polly ojciec jednak powiedział, że  
 powinna się ona cieszyć, gdyż nie o i musi nich chodzić przychodziła  
 obiad Anna wraz z Nancy przestrzegać ciotki włosy w dobrych manier  
 zabrania wspomniania ojcu dziewczynki, pastorze Wittherze jakąś  
 dziwną niechęć; podejrzewa, jest to spowodowane pewną piękną  
 fryzurę wpina we różę tłumaczy jej, wcale chora uczy ją bawić radość  
 przyprowadza do domu - bez patrzeć na swoje piegi żywi on historią  
 miłosną, która ta, po mu służba (której wcześniej miał) doktor  
 Chilton, ale oświetlone za karę ma go nia, zyskuje poprawia humor  
 wyraża pałęta może zastanawia swoim wyzdrowieniu rychłym powrocie  
 cześć swoich ciotek) choćby słowem zjeść kuchni przez słońce się,  
 dziesięcioletniego chłopca dziecka lustro wzruszona, daje rzucają  
 one ścianę nic zachwycony, czy naprawdę odmówić, proponuje kolei, by  
 przynosząc została przechodzi pryzmaty niego blask tęczy tak jedynie  
 co dach zawiesić nad oknem zaadoptował Jimmy'ego jego najdziwniejsze  
 nawet odkryła Jimmy'ego Beana, dziewczynce lepszy zgody uwierzyć  
 Beana pokój sierotę najmłodsza najtrudniejszej stodoły, bowiem, gdzie  
 później sercu

W tym przypadku aktywowanych zostało 166 różnych neuronów. Tak nagły wzrost liczby aktywowanych neuronów pomiędzy trzykrotnym i czterokrotnym pobudzeniem pytaniem wiąże się prawdopodobnie ze zbliżonymi wartościowo wagami połączeń dla dalszych neuronów i ich aktywacje w tej samej chwili czasowej.

## 6.2. Fryderyk Chopin

Drugim testowanym przykładem jest „Fryderyk Chopin”. Dane wprowadzone do programu przedstawiono na rys. 6.2

Program zbierał tekst z dwóch stron internetowych:

*http : //odfoundation.eu/a/6456, koncerty – chopinowskie – codziennie – w – ukraińskim – świecie*

Temat:

Pytanie:

Zaawansowane parametry wyszukiwania:  
 Ilość linków na stronie głównej (zalecane 3):

Rysunek 6.2: Dane wprowadzone przez użytkownika - przykład Fryderyk Chopin

*https : //pl.wikipedia.org/wiki/Fryderyk\_Chopin*

Przy czym w przypadku pierwszej strony, wynikiem był pusty łańcuch. W drugim przypadku zebrano tekst biografii Fryderyka Chopina.

W tym przypadku także przeprowadzono próby pobudzania sieci dla różnych ilości powtórzeń pytania. W poniższych podrozdziałach umieszczone zostały otrzymane wyniki.

### 6.2.1. Jednokrotne i dwukrotne pobudzanie pytaniem

jest Fryderyk Chopin

Analogicznie jak w pierwszym przykładzie, przy jednokrotnym pobudzeniu aktywacji uległy tylko zewnętrznie pobudzane neurony wyrazów zapytania. Przy drugim obiegu pętli nie zostały aktywowane żadne dodatkowe neurony (neurony wyrazów z zapytania w stanie refrakcji).

### 6.2.2. Trzykrotne i czterokrotne pobudzanie pytaniem

jest Fryderyk Chopin jest Fryderyk *https : //pl.wikipedia.org/wiki/Fryderyk\_Chopin*  
 Chopin

W trzeciej pętli doszło do pobudzenia neuronu hiperłącza. Jest to o jeden obieg pętli wcześniej niż w przypadku przykładu 6.1. Przy drugim pobudzeniu te wyrazy „jest”, „Fryderyk” i „Chopin” są w stanie refrakcji i nie zostają aktywowane - nie otrzymujemy ich ponownie w odpowiedzi. Jednak za trzecim i czwartym obiegiem pętli neurony już się zrelaksowały na tyle, że pobudzenie tych wyrazów sygnałem o wartości 1, skutkuje kolejną aktywacją tych neuronów.

### 6.2.3. Pięciokrotne pobudzanie pytaniem

jest Fryderyk Chopin jest Fryderyk [https://pl.wikipedia.org/wiki/Fryderyk\\_Chopin](https://pl.wikipedia.org/wiki/Fryderyk_Chopin)  
Chopin jest Fryderyk Chopin

Pięciokrotne pobudzanie tylko dodało kolejne pobudzenia wyrazów z pytania.

### 6.2.4. Piętnastokrotne pobudzanie pytaniem

jest Fryderyk Chopin jest Fryderyk [https://pl.wikipedia.org/wiki/Fryderyk\\_Chopin](https://pl.wikipedia.org/wiki/Fryderyk_Chopin)  
Chopin jest Fryderyk Chopin jest Fryderyk Chopin [https://pl.wikipedia.org/wiki/Fryderyk\\_Chopin](https://pl.wikipedia.org/wiki/Fryderyk_Chopin) jest Fryderyk Chopin jest Fryderyk  
Chopin [https://pl.wikipedia.org/wiki/Fryderyk\\_Chopin](https://pl.wikipedia.org/wiki/Fryderyk_Chopin) jest Fryderyk Chopin  
jest Fryderyk Chopin [https://pl.wikipedia.org/wiki/Fryderyk\\_Chopin](https://pl.wikipedia.org/wiki/Fryderyk_Chopin)

Piętnastokrotne pobudzanie jest powtarzaniem wcześniej zebranych łańcuchów złożonych z wyrazów zapytania oraz neuronu hiperłącza. Na tym etapie można stwierdzić, że inne połączenia są tak słabe, że ich proces relaksacji przebiega szybciej niż proces kumulowania pobudzeń z kolejnych obiegów pętli.

## 6.3. Fryderyk Chopin 2

Ten przykład różni się od poprzedniego (6.2) jedynie pytaniem jakie pobudzało sieć. Temat oraz ilość linków pozostała bez zmian. Pytanie brzmiało: „co skomponował Chopin”

### 6.3.1. Jednokrotne i dwukrotne pobudzanie pytaniem

co do skomponował [https://pl.wikipedia.org/wiki/Fryderyk\\_Chopin](https://pl.wikipedia.org/wiki/Fryderyk_Chopin) Chopin

Przy jednokrotnym pobudzeniu aktywacji uległy zewnętrznie pobudzone neurony wyrazów zapytania (co, skomponował, Chopin) oraz dodatkowy wyraz „do”. Przy drugim obiegu pętli nie zostały aktywowane żadne dodatkowe neurony (neurony wcześniej pobudzone w stanie refrakcji).

W porównaniu z pobudzaniem sieci pytaniem z przykładu 6.2 można zauważyć, że pytanie „co skomponował Chopin” pobudziło dodatkowy neuron „do”.

## 6.4. Anne of Green Gables

Ostatnim prezentowanym przykładem będzie „Anne of Green Gables”. Dane wejściowe wprowadzone przez użytkownika przedstawiono na rys. 6.3

Temat:

Pytanie:

Zaawansowane parametry wyszukiwania:

Ilość linków na stronie głównej (zalecane 3):

Rysunek 6.3: Dane wprowadzone przez użytkownika - przykład Anne of Green Gables

Tekst bazowy do utworzenia grafu pochodził z 10 linków:

[https://en.wikipedia.org/wiki/Anne\\_of\\_Green\\_Gables\(1985\\_film\)](https://en.wikipedia.org/wiki/Anne_of_Green_Gables(1985_film))

<http://www.maxima.net/tours/anne-green-gables/>

<http://reederreads.com/2010/07/22/review-anne-of-green-gables-lucy-maud-montgomery/>

<http://allamericanparents.com/event/anne-of-green-gables/>

[https://en.wikipedia.org/wiki/Anne\\_of\\_Green\\_Gables](https://en.wikipedia.org/wiki/Anne_of_Green_Gables)

[https://en.wikipedia.org/wiki/Anne\\_of\\_Green\\_Gables](https://en.wikipedia.org/wiki/Anne_of_Green_Gables)

[https://en.wikipedia.org/wiki/Anne\\_of\\_Avonlea](https://en.wikipedia.org/wiki/Anne_of_Avonlea)

[https://en.wikipedia.org/wiki/Gilbert\\_Blythe](https://en.wikipedia.org/wiki/Gilbert_Blythe)

[https://en.wikipedia.org/wiki/Anne\\_Shirley](https://en.wikipedia.org/wiki/Anne_Shirley)

[https://en.wikipedia.org/wiki/Lucy\\_Maud\\_Montgomery](https://en.wikipedia.org/wiki/Lucy_Maud_Montgomery)

Zebrano treść z każdej z wymienionych stron.

#### 6.4.1. Jednokrotne pobudzenie pytaniem

what was Anne like

Po jednokrotnym pobudzeniu sieci pytaniem użytkownika otrzymano odpowiedź, która składa się jedynie z wyrazów zawartych w pytaniu. Były one aktywowane zewnętrznie. Aktywacja nie rozprzestrzeniła się na inne neurony grafu.

#### 6.4.2. Trzykrotne pobudzenie pytaniem

what was Anne like Anne of what was like Anne [https://en.wikipedia.org/wiki/Anne\\_of\\_Green\\_Gables](https://en.wikipedia.org/wiki/Anne_of_Green_Gables)



W trzeciej pętli doszło do aktywacji neuronu „of”, który nie był zawarty w pytaniu. Aktywował się także neuron hiperłącza, wskazując na ten link, który zawierał treść najbardziej odpowiadającą pytaniu spośród 10 możliwych hiperłączy.



## 7. Podsumowanie

W niniejszej pracy zbudowano system, który na podstawie tekstu zebranego z hiperłączy źródłowych, buduje graf wiedzy skojarzeniowej ANAKG, a następnie poprzez pobudzenie sieci pytaniem użytkownika zwraca odpowiedź wraz z hiperłączem źródła.

Część programu odpowiedzialna za tworzenie grafu wiedzy ANAKG z tekstu ciągłego buduje dokładnie taki model, jaki został przedstawiony w monografii doktora Horzyka [1], co zostało udowodnione przez porównanie przykładowych danych w rozdziale 5.2.2. Implementowany model grafu ANAKG działa poprawnie pod względem relaksacji w przypadku pobudzania wspólnych ścieżek. Wtedy waga połączenia jest zwiększana, a połączenie wyrazów ponownie pobudzonych z wyrazami, które nie wystąpiły ponownie jest osłabiane. Jednak w takim grafie, dużą wagę ma także ścieżka, która wystąpiła jeden raz. Jeżeli wszystkie wyrazy danego zdania pojawiły się w całym tekście tylko raz, połączenia między nimi mają wysoką wagę.

Przetestowane metody pobudzania grafu zakładają aktywację neuronów, które podlegają refrakcji. Neuron nie może być aktywowany dwa razy pod rząd, a pobudzenie następujące w krótkim czasie po aktywacji musi być mocniejsze, aby został przekroczony próg aktywacji neuronu. Pokazano na przykładach jak przebiega rozwijanie aktywacji sieci skojarzeniowej poprzez etapowe pobudzanie neuronów.

Włączenie hiperłączy jako węzła po każdym ze zdań miało znaczny wpływ na odpowiedź wynikową. Jeżeli dany tekst składał się z wielu zdań i po każdym dodany został link, to w efekcie, każde z tych zdań będzie połączone z jednym neuronem (tym hiperłączem). Zwiększa to znacząco wagi połączeń z tym neuronem. Może skutkować prawie bezpośrednim przechodzeniem do neuronu hiperłącza w odpowiedzi, co wykazano na przykładzie „Fryderyk Chopin”. Użytkownik dostaje wskazówkę, gdzie ma szukać dalej, ale nie dostaje konkretnej odpowiedzi na swoje pytanie od razu, musi tak jak w klasycznym wyszukiwaniu sam przejrzeć tekst z hiperłącza. Jednak drugi przykład („Pollyanna”) wykazał, że obecność bezpośredniej odpowiedzi w metodzie wyszukiwania z hiperłączami w grafie może zwrócić, oprócz źródła, także bezpośrednią odpowiedź. Prawdopodobnie wynika to z konstrukcji tekstu bazowego. Jeżeli słowa występujące w nim są różnorodne i nie ma powtarzalnych sekwencji, które uzyskiwałyby wyższą wagę, to dominującą asocjacją jest sekwencja zawierająca neuron z pytania oraz hiperłącze.

Kwestią, która miała wpływ na działanie programu, jest „sztywność” wyrazów i ich rozpoznawanie. Chodzi tutaj o poziom skomplikowania języka polskiego, w którym obecna jest odmiana wyrazów. Dla programu, słowa „pustynia”, „pustyni”, „pustynie” są reprezentowane przez trzy różne neurony. Umysł człowieka jest w stanie stwierdzić, że te trzy wyrazy odnoszą się do tego samego, ale program je roz-

różnia. W momencie zadania w zapytaniu tylko jednej z tych opcji w wynikach odrzucone zostaną dwie pozostałe.

Na podstawie wyników stwierdza się również, iż niezwykle istotnym jest zadanie właściwego pytania. Często zdarza się, że osoba pytana nie rozumie jakiego zwrotu w zapytaniu i nie jest w stanie poprawnie odpowiedzieć. Podobna sytuacja występuje w przypadku grafu. Jeżeli w pytaniu zawarte są wyrazy, których nie ma w sieci (czyli graf ich „nie zna”), to są one pomijane w procesie pobudzania sieci i wyszukiwania odpowiedzi. Odpowiednio zadane pytanie może o wiele bardziej pobudzić sieć, co będzie skutkowało pełniejszą odpowiedzią. To zjawisko zostało przedstawione na przykładzie „Fryderyk Chopin”, w którym to ten sam graf wiedzy ANAKG był pobudzany dwoma pytaniami. Wyniki były mniej lub bardziej rozbudowane, w zależności od stopnia „dopasowania” pytania do wiedzy zawartej w grafie.

Największą komplikacją napotkaną w pracy była analiza i zbieranie treści ze stron internetowych. Wynika to ze zróżnicowanej budowy stron, co stwarzało trudności ze sformułowaniem założeń do przeszukiwania. Dodatkowo w wynikach często pojawiały się strony, które zawierały listę do kolejnych stron, a nie treść, którą można by poddać analizie, co skutkowało pustym tekstem zebrany z strony. Był to problem, ponieważ w założeniu program przeszukuje tylko pierwszą warstwę wyników, a nie wchodzi w linki zawarte na podstronach.

Dalsze prace na rozwijaniem i udoskonalaniem programu mogłyby dotyczyć innych metod przechowywania hiperłączy w grafie oraz wprowadzenia dodatkowych danych, reguł opisujących odmianę wyrazów w języku polskim. Należałoby także zwrócić uwagę na dokładniejsze wyszukiwanie treści ze stron internetowych. Innym rozwiązaniem mogłoby być zastąpienie tego kroku wczytywaniem gotowych tekstów (bazy wiedzy), np. w postaci całych książek na interesujący użytkownika temat. Wyeliminowałoby to problem „zazumienia” danych w sieci.

Obecnie nic nie jest w stanie zastąpić ludzkiego mózgu pod względem kompleksowości myślenia, wnioskowania, kojarzenia, ponieważ mózg człowieka to nie tylko suche fakty, nie tylko przetworzone w określony sposób informacje, ale także przeżycia, a to trudno jest zaimplementować. Jednak prowadzone badania i ciągle rozwijające się modele są kolejnymi krokami do coraz lepszego poznania działania ludzkiego mózgu.

## Bibliografia

- [1] A. Horzyk. *Sztuczne systemy skojarzeniowe i asocjacyjna sztuczna inteligencja*. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2013.
- [2] J.W. Kalat. *Biologiczne podstawy psychologii*. Wydawnictwo Naukowe PWN, Warszawa, 2011.
- [3] A. Longstaff. *Neurobiologia. Krótkie wykłady*. Wydawnictwo Naukowe PWN, Warszawa, 2009.
- [4] A. Stellman, J. Greene. *Rusz głową! C#*. Wydawnictwo Helion, Gliwice, 2010.
- [5] J. Strelau. *Psychologia, Podręcznik akademicki, t.2, Psychologia ogólna*. Gdańskie Wydawnictwo Psychologiczne GWP, Gdańsk, 2000.
- [6] J. Tokarski. *Słownik wyrazów obcych*. Wydawnictwo Naukowe PWN, Warszawa, 1980.
- [7] A. Horzyk. *How does generalization and creativity come into being in neural associative systems and how does it form human-like knowledge?*. Neurocomputing 144 (2014) 238-257.
- [8] B. Jędrzec. *Pozyskiwanie wiedzy z dużych zbiorów danych z zastosowaniem adaptacyjnych procedur generowania zapytań*. Rozprawa doktorska, AGH, Kraków, 2008.
- [9] BSCS organisation. *The Brain: Understanding Neurobiology Through the Study of Addiction*. BSCS and Videodiscovery, Inc., Colorado Springs, 2000.
- [10] A. Horzyk. *Sztuczne systemy skojarzeniowe i asocjacyjna sztuczna inteligencja*. prezentacja habilitacyjna
- [11] R. Tadeusiewicz. *Prezentacja z sieci neuronowych pt. Jak są zbudowane biologiczne sieci neuronowe*. [http : //www.uci.agh.edu.pl/uczelnia/tad/Sieci\\_neuronowe/02 - Jak\\_s\\_a\\_zbudowane\\_biologiczne\\_sieci\\_neuronowe.pdf](http://www.uci.agh.edu.pl/uczelnia/tad/Sieci_neuronowe/02 - Jak_s_a_zbudowane_biologiczne_sieci_neuronowe.pdf).
- [12] A. Horzyk. *Strona internetowa*. <http://home.agh.edu.pl/horzyk>.
- [13] Baza wiedzy MSDN [https : //msdn.microsoft.com/pl - pl/library/](https://msdn.microsoft.com/pl-pl/library/).
- [14] *Internetowy Słownik języka polskiego*. Wydawnictwo Naukowe PWN. [http : //sjp.pwn.pl](http://sjp.pwn.pl).



## **A. Załącznik A - Zawartość dołączonej płyty**

Na płycie dołączonej do pracy magisterskiej znajduje się:

- Wersja elektroniczna pracy magisterskiej pt. „Budowa internetowego kontekstowego asocjacyjnego silnika wyszukiwania informacji z wykorzystaniem sztucznych systemów skojarzeniowych.”
- Folder o nazwie PROGRAM zawierający trzy testowane w pracy wersje programu.
- Wyciąg z indeksu - plik w formacie XLS.