



Mgr inż. Grzegorz Pietruszewski



Mgr Jacek Głębocki

Poznań – Splendor Systemy Informacyjne

PROPOZYCJA WPROWADZENIA I ZASTOSOWANIA METRYKI W TEZAURUSACH

Abstract

The proposal of metric function for thesauri is presented. Each vertex of a thesaurus tree is an ordered pair where the first entry is a descriptor and the second entry is its semantic code (for several semantic codes for a descriptor, an adequate number of vertices are created). The distance between two vertices is next used to evaluate a distance between a vertex and a set of vertices and between two sets of vertices as well. Two distance functions are discussed: without a weight and with a weight assigned to each vertex. The distance evaluation in thesauri could be applied to estimate a similarity of thesauri indexed articles.

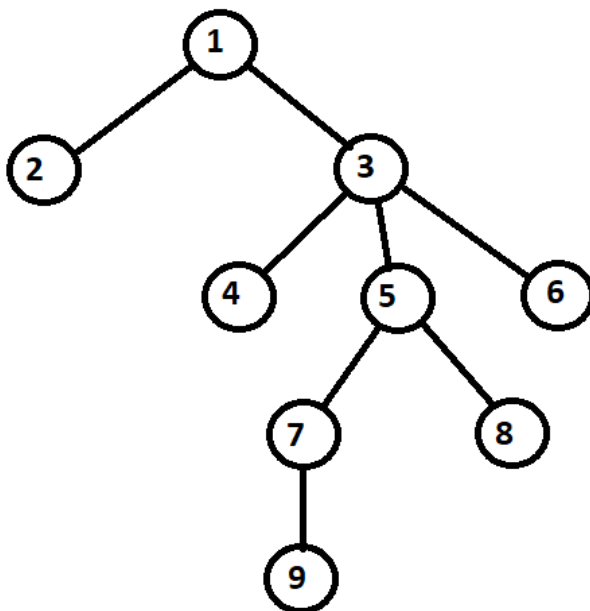
Streszczenie

Artykuł przedstawia propozycję wprowadzenia metryki w tezaursach. Jako wierzchołki drzewa tezaursu traktowane są uporządkowane pary: deskryptor oraz kod semantyczny. Wprowadzono metrykę opartą wyłącznie na zawartym w tezaursie kodzie semantycznym oraz metrykę uwzględniającą wagi wierzchołków. Omówiono odległość między wierzchołkami, odległość między wierzchołkiem a zbiorem wierzchołków oraz odległość między zbiorami wierzchołków. Możliwość oszacowania odległości między zbiorami wierzchołków daje podstawy do określania podobieństwa artykułów klasyfikowanych hasłami tezaursu.

Wstęp

Próba szacowania podobieństwa deskryptorów tezaursu wymaga wprowadzenia metryki. Wprowadzenie metryki pozwoli na posługiwanie się pojęciem odległości pomiędzy deskryptorami – można będzie dzięki temu przyjmować, że deskryptory

są tym bardziej do siebie podobne, im ich odległość jest mniejsza. Tezaurus można przedstawić w postaci drzewa lub lasu (zbioru drzew). Omówmy podstawowe pojęcia, które będą używane poniżej (patrz poz. [1] s. 16 oraz s. 23) na przykładzie drzewa przedstawionego na ilustracji nr 1:



- a) liśćmi są wierzchołki 2, 4, 9, 8, 6
- b) wierzchołkami wewnętrznymi są wierzchołki 1, 3, 5, 7
- c) synami wierzchołka 5 są wierzchołki 7 i 8
- d) ojcem wierzchołka 5 jest wierzchołek 3
- e) korzeniem jest wierzchołek 1
- f) potomkami wierzchołka 3 są wierzchołki 4, 5, 6, 7, 8, 9 oraz sam wierzchołek 3 (wierzchołek jest zarówno przodkiem, jak i potomkiem samego siebie).
- g) najbliższym wspólnym przodkiem wierzchołków 6 i 7 jest 3
- h) najbliższym wspólnym przodkiem wierzchołków 5 i 7 jest 5
- i) drogą z wierzchołka 1 do wierzchołka 9 jest ciąg wierzchołków 1, 3, 5, 7, 9
- j) drogą z wierzchołka 3 do wierzchołka 8 jest ciąg wierzchołków 3, 5, 8

Kod semantyczny określa położenie wierzchołka w drzewie. Poziom wierzchołka określamy zliczając segmenty kodu semantycznego (ilustracja nr 2, wg poz. [2]).

Prezentacja struktury hierarchicznej

C11 Choroby oka
C11.093 Astenopia
C11.187 Choroby spojówek
C11.187.169 Nowotwory spojówek
C11.187.183 Zapalenie spojówek
C11.187.183.200 Zapalenie spojówek alergiczne
C11.187.183.220 Zapalenie spojówek bakteryjne
C11.187.183.220.250 Zapalenie spojówek wtórne
C11.187.183.220.538 Zapalenie oczu noworodków
C11.187.183.220.889 Jądra
C11.187.183.240 Zapalenie spojówek wirusowe
C11.187.183.240.216 Zapalenie spojówek ostre krwotoczne
C11.187.183.394 Zapalenie rogówki i spojówki
C11.187.183.394.520 Zapalenie rogówki i spojówki zakaźne
C11.187.183.394.550 Zapalenie rogówki i spojówki suche
C11.187.183.749 Choroba Reitera
C11.187.781 Skrzydlik
C11.187.810 Kseroftalmia
C11.204 Choroby rogówki
C11.204.236 Dystrofie rogówki dziedziczne
C11.204.236.438 Zwrodnienie rogówki Fuchsa
C11.204.267 Obrzęk rogówki
C11.204.290 Neowaskularyzacja rogówki
C11.204.299 Zmętnienie rogówki

(0)

Przyjmijmy następującą terminologię oraz założenia i sformułujmy wynikające z nich spostrzeżenia:

(0a) będziemy zajmować się tylko deskryptorami i kodami semantycznymi (pozostałe elementy rekordu tezauryś, jak askryptory, określniki itd. nie mają w poniższych rozważaniach znaczenia)

(0b) wierzchołek w drzewie tezauryś definiujemy jako parę uporządkowaną, w której poprzednik jest deskryptorem, a następnik jest kodem semantycznym (para uporządkowana – patrz [3] s.16) i dalej będziemy mówili „termin” mając na myśli poprzednik tak zdefiniowanej pary oraz „kod” mając na myśli jej następnik

(0c) w tezauryś polihierarchicznym deskryptor może mieć wiele kodów i w takich przypadkach wystąpi jako termin w wielu wierzchołkach różniących się kodem

(0d) jeżeli wierzchołek wewnętrzny ma tylko jednego syna, to wbudowujemy w strukturę drzewa dodatkowy wierzchołek będący drugim synem tego wierzchołka

(na ilustracji 1 należałoby dodać wierzchołkowi 7 – mającego jednego syna 9 – drugiego syna 9bis). Dodatkowy wierzchołek konstruujemy tak, żeby terminem był dowolny, niepusty ciąg znaków różny od wszystkich innych terminów, a następnikiem kod skonstruowany w ten sposób, by różnił się ostatnim segmentem (należy wbudować w strukturę drzewa podobnie jak przy dodawaniu nowego rekordu do tezaursusa – patrz [2])

(0e) w celu uproszczenia początkowych rozważań przyjmijmy, że wszystkich synów danego wierzchołka traktujemy równoważnie – zakres w jakim znaczenie wierzchołka-syna pokrywa się ze znaczeniem wierzchołka-ojca nie jest określony w tezaursusie; zatem przyjmujemy, że każdy spośród k synów danego wierzchołka wypełnia $1/k$ zakresu znaczeniowego ojca

(0f) ponadto jeżeli tezaursus określony jest jako las, jak przykładowo MeSH (patrz [4]), to dodajemy dodatkowy wierzchołek w ten sposób, żeby korzenie dotychczasowych drzew samodzielnych stały się jego synami; nowy wierzchołek jest wtedy korzeniem całego tezaursusa; kod semantyczny korzenia określamy dowolnie (ciąg niepusty np. A) i rozbudowujemy wszystkie kody stosując kod korzenia jako prefiks pozostałych kodów, np. zamiast C01.539.800.130 stosujemy A.C01.539.800.130

Wprowadzenie metryki

Oznaczmy przez X zbiór wszystkich wierzchołków, przez x_0 korzeń drzewa, przez x_n dowolny liść w drzewie. Wśród wierzchołków stanowiących drogę z x_0 do x_n , wierzchołek o indeksie $i+1$ jest synem wierzchołka o indeksie i . Niech p będzie funkcją dwuargumentową, której pierwszym argumentem jest dowolny wierzchołek x_i , a drugim argumentem wierzchołek x_{i+m} , gdzie $0 \leq i \leq n, m \geq 0$ oraz $i+m \leq n$. Przez k_i oznaczmy liczbę synów wierzchołka x_i (w myśl podanych wyżej założeń wykluczamy liczbę synów równą 1, a wierzchołek x_n jako liść nie ma synów). Funkcję p zdefiniujemy następująco:

$$(1) \quad \begin{aligned} p(x, x) &= 0, \text{ dla } x \in X \\ x_i &\in X, x_{i+m} \in X \\ p(x_i, x_{i+m}) &= 1 - \frac{1}{k_i \cdot k_{i+1} \cdot \dots \cdot k_{i+m-1}} \end{aligned}$$

Z definicji widać, że p jest funkcją nieujemną, a ponadto zachodzą nierówności:

$$(2) \quad p(x_i, x_{i+m}) > p(x_{i+1}, x_{i+m})$$

ponieważ

$$p(x_i, x_{i+m}) - p(x_{i+1}, x_{i+m}) = 1 - \frac{1}{k_i \cdot k_{i+1} \cdot \dots \cdot k_{i+m-1}} - 1 + \frac{1}{k_{i+1} \cdot \dots \cdot k_{i+m-1}} = \frac{k_i - 1}{k_i \cdot k_{i+1} \cdot \dots \cdot k_{i+m-1}} > 0$$

oraz

$$(3) \\ p(x_i, x_{i+m}) > p(x_i, x_{i+m-1})$$

ponieważ

$$p(x_i, x_{i+m}) - p(x_i, x_{i+m-1}) = 1 - \frac{1}{k_i \cdot k_{i+1} \cdot \dots \cdot k_{i+m-1}} - 1 + \frac{1}{k_i \cdot \dots \cdot k_{i+m-2}} = \frac{k_{i+m-1} - 1}{k_i \cdot k_{i+1} \cdot \dots \cdot k_{i+m-1}} > 0$$

ponadto

$$(4) \\ p(x_i, x_{i+j+m}) \leq p(x_i, x_{i+j}) + p(x_{i+j}, x_{i+j+m})$$

ponieważ

$$\begin{aligned} p(x_i, x_{i+j+m}) &= 1 - \frac{1}{k_i \cdot \dots \cdot k_{i+j+m-1}} = \\ &= \frac{1}{k_{i+j} \cdot \dots \cdot k_{i+j+m-1}} \cdot \left(k_{i+j} \cdot \dots \cdot k_{i+j+m-1} - \frac{1}{k_i \cdot \dots \cdot k_{i+j-1}} \right) = \\ &= \frac{1}{k_{i+j} \cdot \dots \cdot k_{i+j+m-1}} \cdot \left(k_{i+j} \cdot \dots \cdot k_{i+j+m-1} - 1 + 1 - \frac{1}{k_i \cdot \dots \cdot k_{i+j-1}} \right) = \\ &= \frac{1}{k_{i+j} \cdot \dots \cdot k_{i+j+m-1}} \cdot \left(k_{i+j} \cdot \dots \cdot k_{i+j+m-1} - 1 + p(x_i, x_{i+j}) \right) = \\ &= \frac{1}{k_{i+j} \cdot \dots \cdot k_{i+j+m-1}} \cdot \left(k_{i+j} \cdot \dots \cdot k_{i+j+m-1} \cdot \left(1 - \frac{1}{k_{i+j} \cdot \dots \cdot k_{i+j+m-1}} \right) + p(x_i, x_{i+j}) \right) = \\ &= \frac{1}{k_{i+j} \cdot \dots \cdot k_{i+j+m-1}} \cdot \left(k_{i+j} \cdot \dots \cdot k_{i+j+m-1} \cdot p(x_{i+j}, x_{i+j+m}) + p(x_i, x_{i+j}) \right) = \\ &= p(x_{i+j}, x_{i+j+m}) + \frac{p(x_i, x_{i+j})}{k_{i+j} \cdot \dots \cdot k_{i+j+m-1}} \leq p(x_i, x_{i+j}) + p(x_{i+j}, x_{i+j+m}) \end{aligned}$$

Oprócz tego przekształcając (1)

$$p(x_i, x_{i+m}) = 1 - \frac{1}{k_i \cdot k_{i+1} \cdot \dots \cdot k_{i+m-1}} = \frac{1}{k_i} \cdot \left(k_i - \frac{1}{k_{i+1} \cdot \dots \cdot k_{i+m-1}} \right) = \frac{1}{k_i} \cdot \left(k_i - 1 + 1 - \frac{1}{k_{i+1} \cdot \dots \cdot k_{i+m-1}} \right) =$$

$$= \frac{k_i - 1 + p(x_{i+1}, x_{i+m})}{k_i} = \frac{k_i - 1}{k_i} + \frac{p(x_{i+1}, x_{i+m})}{k_i}$$

otrzymujemy równoważną postać rekurencyjną, użyteczną w zastosowaniach programistycznych

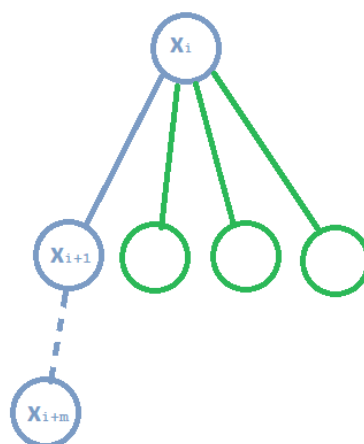
(5)

$$p(x, x) = 0, \text{ dla } x \in X$$

$$x_i \in X, x_{i+m} \in X$$

$$p(x_i, x_{i+m}) = \frac{k_i - 1}{k_i} + \frac{p(x_{i+1}, x_{i+m})}{k_i}$$

Poglądowa interpretacja geometryczna dla $k_i = 4$ może być następująca (ilustracja 3):



$$p(x_i, x_{i+m}) = \frac{3}{4} + \frac{p(x_{i+1}, x_{i+m})}{4}$$

Metryką (patrz [5] s. 191) na zbiorze X nazywamy dwuargumentową funkcję

$d : X \times X \rightarrow [0, \infty)$ spełniającą poniższe warunki:

(6)

$$x, y, z \in X$$

$$i) \quad d(x, y) = 0 \text{ wtedy i tylko wtedy, gdy } x = y$$

$$ii) \quad d(x, y) = d(y, x)$$

$$iii) \quad d(x, y) \leq d(x, z) + d(z, y)$$

Warunek iii) nazywamy nierównością trójkąta.

Zdefiniujmy metrykę d następująco:

(7)

$$d(x, y) = \max(p(w, x), p(w, y))$$

$$x, y, w \in X$$

gdzie w jest najbliższym wspólnym przodkiem x oraz y .

Tak określona funkcja d jest nieujemna i spełnia warunki i) oraz ii) z definicji.

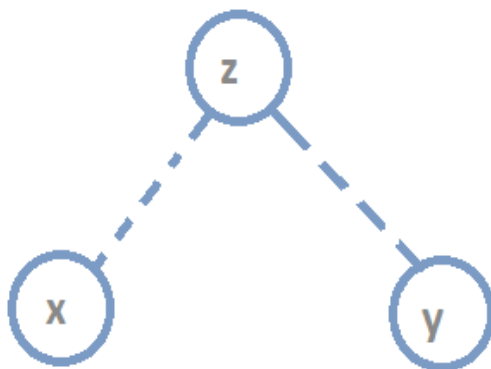
Zauważmy, że dla x i y , gdzie x jest przodkiem y zachodzi równość

(8)

$$d(x, y) = p(x, y)$$

Elementarny dowód nierówności trójkąta tj. warunku iii) polega na rozważeniu następujących przypadków:

1) z jest najbliższym wspólnym przodkiem x oraz y (ilustracja 4)



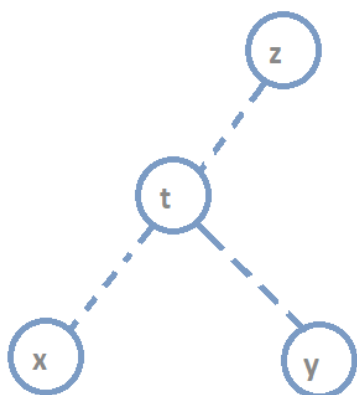
$$d(x, z) = p(z, x)$$

$$d(y, z) = p(z, y)$$

$$d(x, y) = \max(p(z, x), p(z, y))$$

czyli warunek iii) jest spełniony.

2) z jest wspólnym przodkiem x oraz y , ale nie najbliższym (którym jest wierzchołek t) – ilustracja 5



$$d(x, y) = \max(p(t, x), p(t, y))$$

$$d(x, z) = p(z, x)$$

$$d(y, z) = p(z, y)$$

czyli warunek iii) jest spełniony, ponieważ z (2) zachodzi:

$$p(z, x) > p(t, x)$$

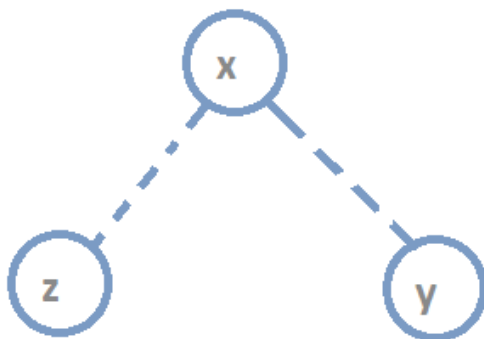
oraz

$$p(z, y) > p(t, y)$$

stąd

$$\max(p(t, x), p(t, y)) < p(t, x) + p(t, y)$$

3) x jest najbliższym wspólnym przodkiem z oraz y (ilustracja 6)



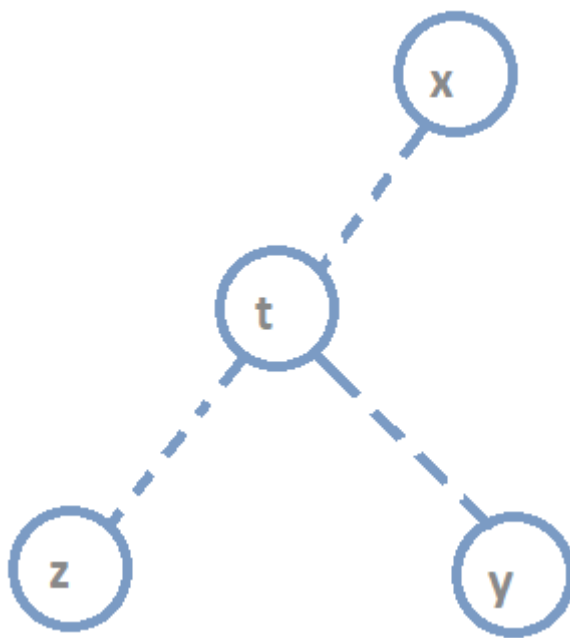
$$d(x, y) = p(x, y)$$

$$d(x, z) = p(x, z)$$

$$d(y, z) = \max(p(x, y), p(x, z))$$

Powyższe równości wskazują, że warunek iii) jest spełniony.

4) x jest wspólnym przodkiem x oraz z , ale nie najbliższym (którym jest wierzchołek t) – ilustracja 7.



$$d(x, y) = p(x, y)$$

$$d(x, z) = p(x, z)$$

$$d(y, z) = \max(p(t, y), p(t, z))$$

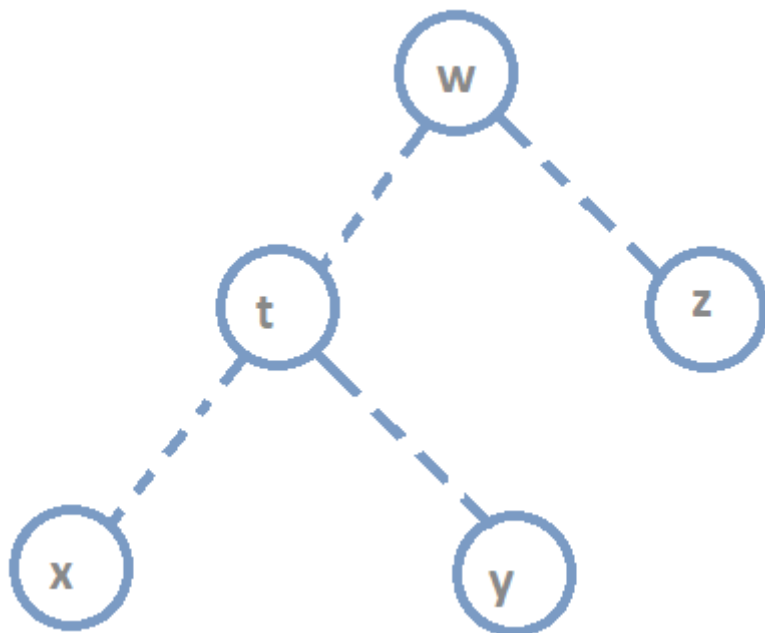
Rozważmy przypadek, w którym z jest synem t , natomiast y dowolnym potomkiem t (zakładając oczywiście, że y i z nie są identyczne z t).

$$d(y, z) \geq p(t, y)$$

zatem

$$d(x, y) = p(x, y) \leq p(x, t) + p(t, y) \leq p(x, t) + d(z, y) \leq d(x, z) + d(z, y)$$

5) najbliższym wspólnym przodkiem dla x oraz y jest t , natomiast najbliższym wspólnym przodkiem dla t oraz z jest w (ilustracja 8)



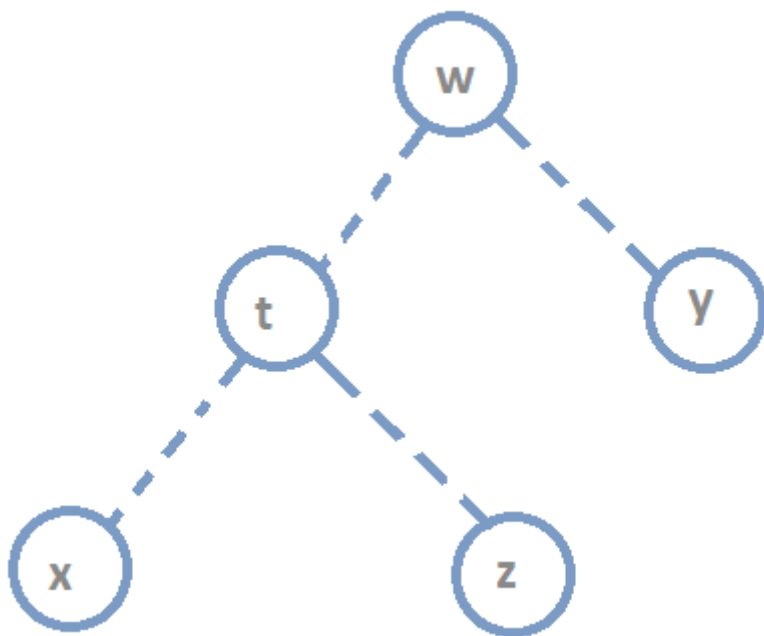
$$d(x, z) > p(t, x)$$

$$d(y, z) > p(t, y)$$

$$d(x, y) = \max(p(t, y), p(t, x))$$

więc warunek iii) jest spełniony

6) najbliższym wspólnym przodkiem dla x oraz z jest t , natomiast najbliższym wspólnym przodkiem dla t oraz y jest w (ilustracja 9)



$$d(x, z) \geq p(t, x)$$

$$d(z, y) \geq p(w, t)$$

zatem:

$$d(x, z) + d(z, y) \geq p(t, x) + p(w, t) \geq p(w, x) = d(x, y)$$

W związku z powyższym d jest metryką na X i może być stosowana do obliczania odległości w teaurusie, którego elementami są w myśl założeń **(0)** pary uporządkowane złożone z terminu i kodu semantycznego. Metryka d jest metryką ograniczoną przez jedynkę. Wadą przyjętej definicji jest to, że minimalną odległością pomiędzy różnymi wierzchołkami jest $\frac{1}{2}$, co wynika z założenia **(0e)** o równoważnym traktowaniu synów wierzchołka.

Wprowadzenie metryki z wagą wierzchołków

Jeżeli w teaurusie byłyby zawarte informacje dodatkowe pozwalające na nierównoważne traktowanie synów danego wierzchołka poprzez przypisanie każdemu z nich wagi, to można by odrzucić założenie **(0e)** i określić funkcję p_w analogicznie do funkcji p zdefiniowanej w **(1)**, ale z uwzględnieniem wagi wierzchołków:

(9)

$$p_w(x, x) = 0, \text{ dla } x \in X$$

$$x_i \in X, x_{i+m} \in X$$

$$x_i \text{ jest ojcem } x_{i+1}$$

$$w_i \text{ jest wagą wierzchołka } x_{i+1} \text{ (czyli syna wierzchołka } x_i)$$

$$s_i \text{ jest sumą wag synów wierzchołka } x_i$$

$$p_w(x_i, x_{i+m}) = 1 - \frac{(s_i - w_i) \cdot \dots \cdot (s_i - w_{i+m-1})}{s_i \cdot s_{i+1} \cdot \dots \cdot s_{i+m-1}}$$

Można pokazać, że dla (9) podobnie jak dla (1) zachodzą nierówności analogiczne do (2), (3), (4) tj.

(10)

$$p_w(x_i, x_{i+m}) > p_w(x_{i+1}, x_{i+m})$$

(11)

$$p_w(x_i, x_{i+m}) > p_w(x_i, x_{i+m-1})$$

(12)

$$p_w(x_i, x_{i+j+m}) \leq p_w(x_i, x_{i+j}) + p_w(x_{i+j}, x_{i+j+m})$$

co pozwala na zdefiniowanie metryki tak jak (7) poprzez zastąpienie p przez p_w .

Tak zdefiniowana metryka nadal jest metryką ograniczoną przez jedynkę, ale jej wartość dla różnych wierzchołków może być dowolnie bliska zeru, w zależności od zastosowanych wag.

Odległość wierzchołka od zbioru wierzchołków

Metrykę d można wykorzystać do określenia odległości wierzchołka od zbioru wierzchołków. Znajduje tu zastosowanie definicja odległości punktu od zbioru (patrz [5] s. 199), która w szczególnym przypadku zbiorów skończonych brzmi następująco:

(13)

A niepusty podzbiór X , x ustalony wierzchołek X , d metryka

$$D(x, A) = \min_{a \in A} d(x, a)$$

D jest zatem najmniejszą z odległości od x do poszczególnych elementów zbioru A

Odległość pomiędzy zbiorami wierzchołków

Do określenia odległości pomiędzy zbiorami wierzchołków można wykorzystać odległość Hausdorffa. Definicja odległości Hausdorffa (odstępu Hausdorffa, patrz [5] s. 269) w zastosowaniu do zbiorów skończonych brzmi następująco:

(14)

A, B niepuste podzbiory X , $x \in A, y \in B$, d metryka

$$D_H(A, B) = \max(\max_{x \in A} D(x, B), \max_{y \in B} D(y, A))$$

Przykłady

Poniżej (ilustracja 10, wg poz. 4) przedstawiająca jedną z gałęzi MeSH. Z uwagi na to, że wszystkie wierzchołki należą do jednej gałęzi warunek (Of) jest nieistotny, zatem nie będziemy stosować dodatkowego prefiksu przed kodami semantycznymi.

Ocular Physiological Phenomena [G14]

Accommodation, Ocular [G14.010]

Adaptation, Ocular [G14.020]

Afterimage [G14.040]

Blinking [G14.152]

Dominance, Ocular [G14.264]

Evoked Potentials, Visual [G14.330]

Eye Color [G14.340]

Eye Movements [G14.350]

Convergence, Ocular [G14.350.217]

Fixation, Ocular [G14.350.253]

Nystagmus, Physiologic [G14.350.378]

Pursuit, Smooth [G14.350.453]

Saccades [G14.350.500]

Figural Aftereffect [G14.360]

Flicker Fusion [G14.370]

Glare [G14.400]

Intraocular Pressure [G14.440]

Lacrimonal Elimination [G14.540]

Ocular Absorption [G14.590]

Optic Flow [G14.700]

Refraction, Ocular [G14.760]

Vision Disparity [G14.930]

Vision, Ocular [G14.935]

Visual Acuity [G14.940]

Contrast Sensitivity [G14.940.500]

Emmetropia [G14.940.750]

Visual Fields [G14.950]

Oznaczmy:

$z = (, \text{Nystagmus, Physiologic}, G14.350.378)$

$x = (\text{Fixation, Ocular}, G14.350.253)$

$y = (\text{Emmetropia}, G14.940.750)$

$r = (\text{Saccades}, G14.350.500)$

$v = (\text{Ocular Physiological Phenomena}, G14)$

$w = (\text{Eye Movements}, G14.350)$

$A = \{z, x\}$

$B = \{y, r\}$

Zastosujemy metrykę wg definicji (7).

Odległość wierzchołków w oraz y

Najbliższym wspólnym przodkiem w oraz y jest v .

$$d(w, y) = \max(p(v, w), p(v, y))$$

$$p(v, w) = 1 - \frac{1}{20} = \frac{19}{20} = \frac{38}{40}$$

$$p(v, y) = 1 - \frac{1}{20 \cdot 2} = \frac{39}{40}$$

$$d(x, y) = \frac{39}{40} = 0,975$$

Odległość wierzchołka z od zbioru B

Najbliższym wspólnym przodkiem z oraz y jest v , a najbliższym wspólnym wierzchołkiem z oraz r jest w .

$$D(z, B) = \min(d(z, y), d(z, r))$$

$$d(z, y) = \max(p(v, z), p(v, y)) = \max(1 - \frac{1}{20 \cdot 5}, 1 - \frac{1}{20 \cdot 2}) = \max(\frac{198}{200}, \frac{195}{200}) = \frac{198}{200} = 0,99$$

$$d(z, r) = 1 - \frac{1}{5} = \frac{4}{5} = 0,8$$

$$D(z, B) = 0,8$$

Odległość zbioru A od zbioru B

Najbliższym wspólnym przodkiem r oraz z jest w , który jest jednocześnie najbliższym wspólnym przodkiem r oraz x ; najbliższym wspólnym przodkiem y i z , a także y i x jest v .

$$D(r, A) = \frac{4}{5}$$

$$D(y, A) = \frac{99}{100}$$

$$\max(D(r, A), D(y, A)) = \frac{99}{100}$$

$$D(z, B) = D(x, B) = \frac{4}{5}$$

$$D_H(A, B) = \max\left(\frac{99}{100}, \frac{4}{5}\right) = 0,99$$

Podsumowanie

Metrykę d można zastosować do określania podobieństwa artykułów klasyfikowanych hasłami tezausa. Każde hasło użyte do klasyfikacji artykułu należy w tym celu zamienić na tyle wierzchołków, ile posiada kodów semantycznych. Przydatność proponowanego rozwiązania można byłoby zweryfikować w praktyce poprzez zastosowanie do bazy PubMed (indeksowanej deskryptorami Medical Subject Headings NLM) lub Polskiej Bibliografii Lekarskiej (indeksowanej deskryptorami Polskiego Tezausa Medycznego Głównej Biblioteki Lekarskiej).

Bibliografia

- [1] Hopcroft J.E., Ullman J. D.: Wprowadzenie do teorii automatów, języków i obliczeń. - Warszawa : Wydawnictwo Naukowe PWN, 1994, 478 s. - Wyd. I. - ISBN 83-01-11298-0
- [2] Włostowska Ewa, Głębocki Jacek, Chrobak Ewa: Nowoczesne metody samodzielnego przeszukiwania bazy danych „Polska Bibliografia Lekarska” : propozycja konspektu szkolenia użytkowników naukowej informacji medycznej. - Forum Bibl. Med. 2009 R. 2 nr 1 (3), s. 283-291.
- [3] Rasiowa Helena: Wstęp do matematyki współczesnej. - Warszawa : Państwowe Wydawnictwo Naukowe, 1984, 302 s. - Wyd. VIII. - ISBN 83-01-01373-7
- [4] MeSH Tree Structures. https://www.nlm.nih.gov/mesh/intro_trees.html. [Dostęp 2018.09.18]
- [5] Duda Roman: Wprowadzenie do topologii : Topologia ogólna. - Warszawa : Państwowe Wydawnictwo Naukowe, 1986, 428 s. - Wyd. I. - ISBN 83-01-05714-9