

Kamil Połacik Adrian Zygmunt

KARTOTEKA WTÓRNA							
X\A	Rok założenia	Miasto	Liga	Stadion	Zdobyte tytuły ligowe	Wartość rynkowa klubu	Ilość obcokrajowców
x <sub>1</sub>	1905	Londyn	Premier League	Stamford Bridge	6	Bardzo wysoka	Średnia
x <sub>2</sub>	1924	Monako	Ligue 1	Stadion Ludwika II	8	Średnia	Średnia
x <sub>3</sub>	1886	Londyn	Premier League	Emirates Stadium	13	Wysoka	Średnia
x <sub>4</sub>	1899	Mediolan	Serie A	San Siro	19	Wysoka	Duża
x <sub>5</sub>	1908	Mediolan	Serie A	San Siro	19	Wysoka	Mała
x <sub>6</sub>	1880	Manchester	Premier League	Etihad Stadium	7	Bardzo wysoka	Mała
x <sub>7</sub>	1878	Manchester	Premier League	Old Trafford	20	Bardzo wysoka	Duża
x <sub>8</sub>	1902	Madryt	Primera División	Estadio Santiago Bernabéu	35	Bardzo wysoka	Średnia
x <sub>9</sub>	1899	Barcelona	Primera División	Camp Nou	26	Bardzo wysoka	Mała
x <sub>10</sub>	1900	Monachium	Bundesliga	Allianz Arena	32	Bardzo wysoka	Średnia
x <sub>11</sub>	1909	Dortmund	Bundesliga	Signal Iduna Park	8	Wysoka	Mała
x <sub>12</sub>	1970	Paryż	Ligue 1	Parc des Princes	10	Bardzo Wysoka	Średnia
x <sub>13</sub>	1921	Częstochowa	Ekstraklasa	MSP Raków	0	Niska	Mała
x <sub>14</sub>	1927	Rzym	Serie A	Stadio Olimpico	3	Średnia	Średnia
x <sub>15</sub>	1897	Turyń	Serie A	Allianz Stadium	36	Wysoka	Średnia
x <sub>16</sub>	1899	Lyon	Ligue 1	Groupama Stadium	7	Średnia	Mała
x <sub>17</sub>	1900	Rzym	Serie A	Stadio Olimpico	2	Średnia	Średnia

## Definicja Systemu:

$S = \langle X, A, V, r \rangle$

$X = \{X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}\}$

$A = \{\text{Rok założenia, miasto, liga, stadion, zdobyte tytuły ligowe, Wartość rynkowa klubu, Ilość obcokrajowców}\}$

$V_{\text{Rok założenia}} = \{1878, 1880, 1886, 1897, 1899, 1900, 1902, 1905, 1908, 1909, 1921, 1924, 1927, 1970\}$

$V_{\text{miasto}} = \{\text{Londyn, Monako, Mediolan, Manchester, Madryt, Barcelona, Monachium, Dortmund, Paryż, Częstochowa, Rzym, Turyn, Lyon}\}$

$V_{\text{liga}} = \{\text{Premier League, Ligue 1, Serie A, Primera División, Bundesliga, Ekstraklasa}\}$

$V_{\text{stadion}} = \{\text{Stamford Bridge, Stadion Ludwika II, Emirates Stadium, San Siro, Etihad Stadium, Old Trafford, Estadio Santiago Bernabéu, Camp Nou, Alinaz Arena, Signal Iduna Park, Parc des Princes, MSP Raków, Stadio Olimpico, Allianz Stadium, Groupama Stadium}\}$

$V_{\text{zdobyte tytuły}} = \{0, 2, 3, 6, 7, 8, 10, 13, 19, 20, 26, 32, 35, 36\}$

$V_{\text{wartość rynkowa klubu}} = \{\text{Niska, Średnia, Wysoka, Bardzo wysoka}\}$

$\text{Niska} < 100 \text{ mln \$} < \text{Średnia} < 400 \text{ mln \$} < \text{Wysoka} < 800 \text{ mln \$} < \text{Bardzo Wysoka} < 1100 \text{ mln \$}$

$V_{\text{Ilość obcokrajowców}} = \{\text{Mała, Średnia, Duża}\}$

$\text{Mała} \leq 15 < \text{Średnia} \leq 19 < \text{Duża} \leq 24$

## Algorytm Doyle'a

M = 4 – liczba grup

$\alpha = 0,5$  - współczynnik

S1	C1	F1	R1	P1	S2	C2	F2	R2	P2
X1	(Rok założenia ,1905)	1	3	24	X6	(Rok założenia ,1880)	1	3	24
X2					X7				
X3					X8				
X4	(Rok założenia ,1924)	1	3	24	X9	(Rok założenia ,1878)	1	3	24
X5	(Rok założenia ,1886)	1	3	24		(Rok założenia ,1902)	1	3	24
	(Rok założenia ,1899)	1	3	24		(Rok założenia ,1899)	1	3	24
	(Rok założenia ,1908)	1	3	24		(Miasto , Manchester)	2	2	25
	(Miasto , Londyn)	2	2	25		(Miasto , Madryt)	1	3	24
	(Miasto , Monako)	1	3	24		(Miasto , Barcelona)	1	3	24
	(Miasto , Mediolan)	2	2	25		(Liga, Premier League	2	2	25
	(Liga, Premier League)	2	2	25		(Liga, Primera División)	2	2	25
	(Liga,_Ligue 1)	1	3	24		(Stadion, Etihad Stadium)	1	3	24
	(Liga,_Serie A)	2	2	25		(Stadion, Old Trafford)	1	3	24
	(Stadion, Stamford Bridge)	1	3	24		(Stadion, Estadio Santiago Bernabéu)	1	3	24
	(Stadion, Stadion Ludwika II)	1	3	24		(Stadion, Camp Nou)	1	3	24
	(Stadion, Emirates Stadium)	1	3	24		(Tytuły, 7)	1	3	24
	(Stadion, San Siro)	2	2	25		(Tytuły, 20)	1	3	24
	(Tytuły, 6)	1	3	24		(Tytuły, 35)	1	3	24
	(Tytuły, 8)	1	3	24		(Tytuły, 26)	1	3	24

	(Tytuły, 13)	1	3	24		(Wartość , Bardzo wysoka)	4	1	26
	(Tytuły, 19)	2	2	25		(Ilość, Mała)	2	2	25
	(Wartość , Bardzo wysoka)	1	3	24		(Ilość, Duża)	1	3	24
	(Wartość , Średnia)	1	3	24		(Ilość, Średnia)	1	3	24
	(Wartość , Wysoka)	3	1	26					
	(Ilość, Średnia)	3	1	26					
	(Ilość, Duża)	1	3	24					
	(Ilość, Mała)	1	3	24					

S3	C3	F3	R3	P3	S4	C4	F4	R4	P4
X10 X11 X12	(Rok założenia , 1900)	1	2	25	X13 X14 X15 X16 X17	(Rok założenia , 1921)	1	3	24
	(Rok założenia , 1909)	1	2	25		(Rok założenia , 1927)	1	3	24
	(Rok założenia , 1970)	1	2	25		(Rok założenia , 1897)	1	3	24
	(Miasto , Monachium)	1	2	25		(Rok założenia , 1899)	1	3	24
	(Miasto , Dortmund)	1	2	25		(Rok założenia , 1900)	1	3	24
	(Miasto , Paryż)	1	2	25		(Miasto , Częstochowa)	1	3	24
	(Liga, Bundesliga)	2	1	26		(Miasto , Rzym)	2	2	25
	(Liga, Ligue 1)	1	2	25		(Miasto , Turyn)	1	3	24
	(Stadion, Allianz Arena)	1	2	25		(Miasto , Lyon)	1	3	24
	(Stadion, Signal Iduna Park)	1	2	25		(Liga, Ekstraklasa)	1	3	24
	(Stadion, Parc des Princes)	1	2	25		(Liga, Serie A)	3	1	26

	(Tytuły, 32)	1	2	25		(Liga, Ligue 1)	1	3	24
	(Tytuły, 8)	1	2	25		(Stadion, MSP Raków)	1	3	24
	(Tytuły, 10)	1	2	25		(Stadion, Stadio Olimpico)	2	2	25
	(Wartość , Bardzo wysoka)	2	1	26		(Stadion, Allianz Stadium)	1	3	24
	(Wartość , Wysoka)	1	2	25		(Stadion, Groupama Stadium)	1	3	24
	(Ilość, Średnia)	2	1	26		(Tytuły, 0)	1	3	24
	(Ilość, Mała)	1	2	25		(Tytuły, 3)	1	3	24
						(Tytuły, 36)	1	3	24
						(Tytuły, 7)	1	3	24
						(Tytuły, 2)	1	3	24
						(Wartość , Niska)	1	3	24
						(Wartość , Średnia)	3	1	26
						(Wartość , Wysoka)	1	3	24
						(Ilość, Mała)	2	2	25
						(Ilość, Średnia)	3	1	26

$$P_i = b - r_i$$

b – wartość bazowa (max. Liczba cech +1) = 26 +1 = 27

Z S4

G(di,Pi)	P1	P2	P3	P4
X1	172	-	-	-
X2	170	-	-	-
X3	174	-	-	-
X4	174	-	-	-
X5	174	-	-	-
X6	-	173	-	-
X7	-	172	-	-
X8	-	171	-	-
X9	-	172	-	-
X10	-	-	178	-
X11	-	-	176	-
X12	-	-	177	-
X13	-	-	-	169
X14	-	-	-	176
X15	-	-	-	172
X16	-	-	-	171
X17	-	-	-	176

$$H1 = \max (G(di,P1)) = 174$$

$$H2 = \max (G(di,P2)) = 173$$

$$H3 = \max (G(di,P3)) = 178$$

$$H4 = \max (G(di,P4)) = 176$$

$$T = (\text{Min}(G(di,Pi)) + \max(G(di,Pi)))/2 = 178 + 169 / 2 = 347/2 = 173,5 = 174$$

$$T_j = \begin{cases} T & \text{dla } H_j \leq T \\ H_j - \alpha(H_j - T) & \text{dla } H_j > T \end{cases}$$

$$T1 = 174$$

$$T2 = 174$$

$$T3 = H3 - \alpha(H3 - T) = 178 - 0,5(178 - 174) = 178 - 2 = 176$$

$$T4 = H4 - \alpha(H4 - T) = 176 - 0,5(176 - 174) = 176 - 1 = 175$$

$$S1' = \{x_3, x_4, x_5\}$$

$$S2' = \{\emptyset\}$$

$$S3' = \{x_{10}, x_{11}, x_{12}\}$$

$$S4' = \{x_{14}, x_{17}\}$$

$$S5' = L = \{x_1, x_2, x_6, x_7, x_8, x_9, x_{13}, x_{15}, x_{16}\}$$

### Wyszukiwanie sekwencyjne

#### Pytanie ogólne:

t = (Liga, Serie A)

$$p_{\min} = 0,5$$

$$p(d, q) = \frac{\overline{\overline{d \cap q}}}{\overline{\overline{q}}}$$

Liczba wspólnych pojęć / liczba pojęć w pytaniu

$$p(t, x_i) = \frac{\overline{\overline{t \cap x_i}}}{\overline{\overline{t}}}$$

Obliczam współczynniki korelacji pytania ze wszystkimi dokumentami:

$$p(t, x_1) = 0/1 = 0$$

$$p(t, x_2) = 0/1 = 0$$

$$p(t, x_3) = 0/1 = 0$$

$$p(t, x_4) = 1/1 = 1$$

$$p(t, x_5) = 1/1 = 1$$

$$p(t, x_6) = 0/1 = 0$$

$$p(t, x_7) = 0/1 = 0$$

$$p(t, x_8) = 0/1 = 0$$

$$p(t, x_9) = 0/1 = 0$$

$$p(t, x_{10}) = 0/1 = 0$$

$$p(t, x_{11}) = 0/1 = 0$$

$$p(t, x_{12}) = 0/1 = 0$$

$$p(t, x_{13}) = 0/1 = 0$$

$$p(t, x_{14}) = 1/1 = 1$$

$$p(t, x_{15}) = 1/1 = 1$$

$$p(t, x_{16}) = 0/1 = 0$$

$$p(t, x_{17}) = 1/1 = 1$$

$$p(t, x_i) \geq p_{\min}$$

$$\sigma(t) = \{x_4, x_5, x_{14}, x_{15}, x_{17}\}$$

$$t \not\leq x_1$$

$$t \not\leq x_2$$

$$t \not\leq x_3$$

$$t \leq x_4 \quad \sigma(t) = \{x_4\}$$

$$t \leq x_5 \quad \sigma(t) = \{x_4, x_5\}$$

$$t \not\leq x_6$$

$$t \not\leq x_7$$

$$t \not\leq x_8$$

$$t \not\leq x_9$$

$$t \not\leq x_{10}$$

$$t \not\leq x_{11}$$

$$t \not\leq x_{12}$$

$$t \not\leq x_{13}$$

$$t \leq x_{14} \quad \sigma(t) = \{x_4, x_5, x_{14}\}$$

$$t \leq x_{15} \quad \sigma(t) = \{x_4, x_5, x_{14}, x_{15}\}$$

$$t \not\leq x_{16}$$

$$t \leq x_{17} \quad \sigma(t) = \{x_4, x_5, x_{14}, x_{15}, x_{17}\}$$



	Relevantne	Nierelevantne
Wyszukane	5	0 (b)
Niewyszukane	0 (c)	12

**Dokładność:**

$$D = \frac{a}{a + b}$$

**Kompletność:**

$$K = \frac{a}{a + c}$$

$$D = 5 / 5+0 = 5/5 = 100\%$$

$$K = 5 / 5+0 = 5/5 = 100\%$$

**Czas wyszukiwania:**

$$T = N * \tau_x$$

N – liczba dokumentów

$\tau_0$  – średnia czas przeglądu jednego obiektu, założenie  $\tau_0 = 1$  s

$$\tau = 17 * 1s = 17s$$

**Pytanie szczegółowe:**

t = (Liga, Serie A)(Ilość obcokrajowców, Średnia) + (Miasto, Rzym)

$$p_{\min} = 0,5$$

$$p(d, q) = \frac{\overline{\overline{d \cap q}}}{\overline{\overline{q}}}$$

Liczba wspólnych pojęć / liczba pojęć w pytaniu

$$p(t, x_i) = \frac{\overline{\overline{t \cap x_i}}}{\overline{\overline{t}}}$$

Obliczam współczynniki korelacji pytania ze wszystkimi dokumentami:

$$p(t, x_1) = 1/3 = 0,33$$

$$p(t, x_2) = 1/3 = 0,33$$

$$p(t, x_3) = 1/3 = 0,33$$

$$p(t, x_4) = 1/3 = 0,33$$

$$p(t, x_5) = 1/3 = 0,33$$

$$p(t, x_6) = 0/3 = 0$$

$$p(t, x_7) = 0/3 = 0$$

$$p(t, x_8) = 1/3 = 0,33$$

$$p(t, x_9) = 0/3 = 0$$

$$p(t, x_{10}) = 1/3 = 0,33$$

$$p(t, x_{11}) = 0/3 = 0$$

$$p(t, x_{12}) = 1/3 = 0,33$$

$$p(t, x_{13}) = 0/3 = 0,33$$

$$p(t, x_{14}) = 3/3 = 1$$

$$p(t, x_{15}) = 2/3 = 0,66$$

$$p(t, x_{16}) = 0/3 = 0$$

$$p(t, x_{17}) = 3/3 = 1$$

$$p(t, x_i) \geq p_{\min}$$

$$\sigma(t) = \{x_{14}, x_{15}, x_{17}\}$$

$t_1 = (\text{Liga, Serie A})(\text{Ilość obcokrajowców, Średnia})$

$$t \leq x_1$$

$$t \leq x_2$$

$$t \leq x_3$$

$$t \leq x_4$$

$$t \leq x_5$$

$$t \leq x_6$$

$$t \leq x_7$$

$$t \not\leq x_8$$

$$t \not\leq x_9$$

$$t \not\leq x_{10}$$

$$t \not\leq x_{11}$$

$$t \not\leq x_{12}$$

$$t \not\leq x_{13}$$

$$t \leq x_{14} \quad \sigma(t) = \{x_{14}\}$$

$$t \leq x_{15} \quad \sigma(t) = \{x_{14}, x_{15}\}$$

$$t \not\leq x_{16}$$

$$t \leq x_{17} \quad \sigma(t) = \{x_{14}, x_{15}, x_{17}\}$$

$$t_2 = (\text{Miasto, Rzym})$$

$$t_1 \not\leq x_1$$

$$t_2 \not\leq x_2$$

$$t_3 \not\leq x_3$$

$$t_4 \not\leq x_4$$

$$t_5 \not\leq x_5$$

$$t_6 \not\leq x_6$$

$$t_7 \not\leq x_7$$

$$t_8 \not\leq x_8$$

$$t_9 \not\leq x_9$$

$$t_{10} \not\leq x_{10}$$

$$t_{11} \not\leq x_{11}$$

$$t_{12} \not\leq x_{12}$$

$$t_{13} \not\leq x_{13}$$

$$t_{14} \leq x_{14} \quad \sigma(t) = \{x_{14}\}$$

$$t_{15} \not\leq x_{15}$$

$$t_{16} \not\leq x_{16}$$

$$t_{17} \leq x_{17} \quad \sigma(t) = \{x_{14}, x_{17}\}$$

$$\sigma(t) = \sigma(t1) \cup \sigma(t2) = \{x_{14}, x_{15}, x_{17}\} \cup \{x_{14}, x_{17}\} = \{x_{14}, x_{15}, x_{17}\}$$

	Relevantne	Nierelevantne
Wyszukane	3	0 (b)
Niewyszukane	0 (c)	14

$$D = 3 / 3+0 = 3/3 = 100\%$$

$$K = 3 / 3+0 = 3/3 = 100\%$$

**Czas wyszukiwania:**

$$T = N * m * \tau_x + \tau_{pi}$$

N – liczba dokumentów

m – liczba iteracji

$\tau_0$  – średnia czas przeglądu jednego obiektu, założenie  $\tau_0 = 1$  s

$$\tau_{pi} = 0,5s$$

$$\tau = 17 * 1 * 1 + 0,5 = 17,5s$$

**Redundancja:**

$$R = \frac{\sum_{i=1}^m cardX_i - cardX}{cardX}$$

$$R = ((5+4+3+5) - 17) / 17 = 0$$

### **Zajętość pamięci:**

Opis jednego obiektu –  $7 \times 1B = 7B$

Ilość grup – 4 (Kolejno 5, 4, 3, 5 obiektów w każdej grupie)

Profile S1 – 25

Profile S2 – 21

Profile S3 – 18

Profile S4 – 26

1 profil – 3B

$$M = (5 \times 7B + 25 \times 3B) + (4 \times 7B + 21 \times 3B) + (3 \times 7B + 18 \times 3B) + (5 \times 7B + 26 \times 3B) = 110B + 91B + 75B + 113B = 389B$$

### **Wnioski:**

Zaletą wyszukiwania w naszym systemie jest brak redundancji oraz kompletność i dokładność odpowiedzi jaka uzyskaliśmy, a dokładniej oba współczynniki zarówno w pytaniu ogólnym jak i szczegółowym wynoszą 100%. Jak widzimy czas wyszukiwania odpowiedzi na pytanie ogólne jest zbliżony do czasu wyszukiwania na pytanie szczegółowe, więc bardziej nam się „opłaca” zadawać pytanie szczegółowe. W porównaniu do metod list prostych czas wyszukiwania pytania ogólnego jest taki sam, natomiast czas wyszukiwania odpowiedzi na pytanie szczegółowe jest o wiele mniejszy. W algorytmie Doyle’a – 17,5 s natomiast w MLP 51 s, a w modyfikacji przez grupowanie – 28,33.

W porównaniu do metody list inwersyjnych czas wyszukiwania na pytanie ogólne jest krótszy, ponieważ w Doyle’u wynosi 17 s, a w MLI 30 s, pytanie szczegółowe to 17,5 s Doyle oraz MLI w podstawowej 126 a w modyfikacji z zaznaczeniem części wspólnej - 58 s.

W metodzie łańcuchowej czas wyszukiwania odpowiedzi na pytanie ogólne jest krótszy, ponieważ wynosi on tylko 5 s, natomiast na pytanie szczegółowe jest zbliżony – 16 s. W modyfikacji łańcuchowania grup obiektów jest tak samo czyli ogólne – 5 s, szczegółowe – 16 s.

Wada algorytmu jest na pewno duża ilość danych przez co łatwo o pomyłkę.

Zajętość pamięci wynosi 389B jest to więcej niż w MLP oraz jej modyfikacji (119B oraz 125B), więcej niż w MLI (345B oraz 340B w modyfikacji) oraz więcej niż w podstawowym algorytmie MLŁ (340B) oraz zdecydowanie więcej niż w modyfikacji MLŁ 208B. Jak widzimy spośród wszystkich algorytmów ten zajmuje najwięcej pamięci.

Zaletą wyszukiwania sekwencyjnego jest, sterowanie parametrem  $p_{\min}$  dzięki któremu możemy zawężyć (większe  $p_{\min}$ ) bądź poszerzać (mniejsze  $p_{\min}$ ) zbiór odpowiedzi.