Emilia Kamińska, Adam Niewiadomski

Państwowa Wyższa Szkoła Zawodowa we Włocławku

Algorytmy rekomendacji produktów w systemach e-commerce

The alhorithms for product recommendation in e-commerce systems

---

Streszczenie

Systemy rekomendacyjne w serwisach internetowych są uważane za standard, z którego żaden producent nie jest w stanie zrezygnować, jeśli tylko chce zdobyć pozycję na rynku. Sklepy internetowe, portale aukcyjne, serwisy z multimediami przy każdym oglądanym przez użytkownika obiekcie wyświetlają obiekty (towary, filmy, zdjęcia) podobne lub zbliżone do bieżącego. Hasło „klienci którzy kupili ten towar oglądali również…" jest znane niemal każdemu użytkownikowi Internetu.
O ile jednak porównanie obiektów jest dla człowieka naturalne i intuicyjne, o tyle systemy komputerowe należy tej czynności nauczyć projektując i implementując odpowiednie algorytmy. Celem niniejszego artykułu jest zatem krytyczny przegląd wybranych algorytmów rekomendacji obiektów, opartych na obliczaniu podobieństwa oraz przeprowadzenie doświadczenia pokazującego działanie wybranych algorytmów na przykładzie bazy danych lokali mieszkalnych.

Słowa kluczowe: systemy rekomendowane, odległość Euklidesa, kosinus kąta między wektorami, odległość Hamminga, odległość Tanimoto

Abstract

Product recommendation systems on websites are considered a standard which no manufacturers can avoid in their effort to establish themselves on the market. Online shops, auction websites, multimedia portals- all of them display similar products (in the form of movie clips or photos) to the one the customer is checking out. The phrase:`Customers who bought this also bought...` is known to most Internet users.
However, whereas comparing things is natural and intuitive to human beings, computer systems need to be taught this through designing and implementing appropriate algorithms. Therefore, the aim of this paper is a critical review of select product recommen-

dation algorithms based on calculating probability, as well as conducting an experiment presenting how select algorithms work on the basis of a housing estate database.

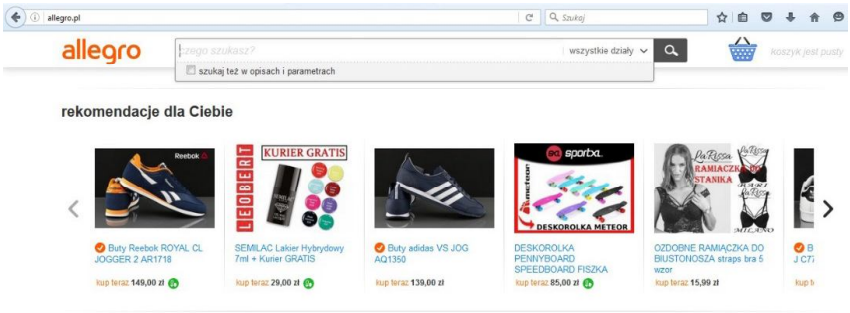
Keywords: recommender system, Euclidean distance, cosine similarity between two vectors, Hamming distance, Tanimoto distance

1. Wprowadzenie

W sytuacji, w której występuje zbyt wielka ilość informacji, by podjąć decyzję lub uzyskać sens danego tematu z pomocą przychodzą systemy rekomendowane. Systemy rekomendowane są specyficzną formą filtracji informacji, usiłującą zaprezentować treści najbardziej interesujące dla użytkownika. Rekomendować można m.in.: inwestycje finansowe, restauracje, samochody, ubrania, książki, wykłady czy chociażby nieruchomości. Z informatycznego punktu widzenia produktem rekomendowanym może być wszystko, co jesteśmy wstanie opisać za pomocą ciągu danych (wektora danych), reprezentującego wybrane cechy obiektu.

Zadanie porównania i selekcji obiektów na podstawie podobieństwa jest przez człowieka wykonywane odruchowo i intuicyjnie. Jednak system komputerowy, który „umie" tego dokonać, musi zostać odpowiednio zaprogramowany i wyposażony w umiejętności zbliżone do ludzkich. Należy zatem stwierdzić, iż zadanie porównywania i rekomendowania obiektów mieści się w obszarze badań nad sztuczną inteligencją.

Rysunek 1. Rekomendowane przedmioty dla użytkownika z serwisu internetowego Allegro



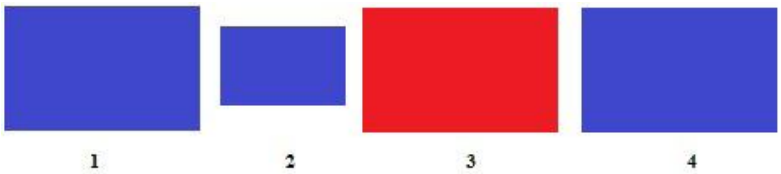
Źródło: opracowanie własne.

W niniejszej pracy zostanie podjęta próba zarekomendowania lokalu najbardziej podobnego do pożądanego przez użytkownika. W tym celu posłużono się bazą danych zawierających 199 lokali, z których każdy opisany jest 9-oma cechami. Cztery z nich są cechami opisanymi dwuwartościowo, pozostałe zaś liczbowo.

2. Miary podobieństwa

Określenie podobieństwa jest subiektywną oceną porównywanych obiektów. Załóżmy, iż mamy cztery obiekty przedstawiające prostokąty (Rys. 2).

Rysunek 2. Podobieństwo prostokątów



Źródło: opracowanie własne.

W jaki sposób ocenić, które z nich są podobne? Możemy powiedzieć, że prostokąt 1, 2 oraz 3 są podobne ze względu na kolor. Jak również stwierdzić, że prostokąty 1, 3 oraz 4 są podobne ze względu na wielkość. Dokonaliśmy tutaj określenia podobieństwa ze względu na cechy, którymi są tutaj kolor i wielkość, pozostałe zaś cechy (np. Długości poszczególnych bloków) zostały pominięte.

2.1. Podobieństwo ciągów binarnych (dwuwartościowych)

Obiekty, które poddajemy ocenie mogą być określone za pomocą wartości binarnych, np.: „tak”, „nie”; „0”, „1”; „zawiera”, „nie zawiera” itp. Jak można zauważyć dla takiej reprezentacji obiektów istnieją tylko dwie możliwe wartości reprezentowane przez wartości logiczne „prawda” i „fałsz”. Podobieństwo dla obiektów opisanych przy pomocy wartości dwuargumentowych może być wyznaczane w kategoriach częstości występowania

wartości pozytywnych oraz negatywnych w każdym z obiektów. Dla lepszego zrozumienia przedstawię przykład. Przykładem niech będą dwa lokale z następującymi cechami binarnymi (Tabela 1.).

Tabela 1. Lokale opisane za pomocą cech dwuargumentowych

	Nieruchomość lokalowa	**Lokal w dzielnicy X**	**< 5 piętra**	**Lokal nieremontowany**
Lokal 14	1	1	1	0
Lokal 35	1	1	1	1

Źródło: opracowanie własne.

Pierwszą miarą podobieństwa przy pomocy, której dokonamy określenia podobieństwa lokali przedstawionych w Tabeli 1 jest Odległość Hamminga.

Odległość Hamminga jest określeniem liczby pozycji, przy których występuje różnica w wartościach zmiennych. Przy założeniu, że porównywane obiekty posiadają tę samą długość. Wzór na odległość Hamminga przedstawia się następująco:

$$d(x_1, x_2) = \sum_{k=1}^{n} |x_{1k} - x_{2k}| \qquad \text{[Wzór 1]}$$

przy czym:
d – odległość Hamminga,
x_1, x_2 – porównywane obiekty (lokale),
n – liczba cech porównywanych obiektów,
k – porównywana cecha.

Tabela 2. Wyliczenia dla Odległości Hamminga

	k_1	k_2	k_3	k_4
Lokal 14	1	1	1	0
Lokal 35	1	1	1	1

Źródło: opracowanie własne.

Na podstawie Tabeli 2 możemy wyznaczyć odległość Hamminga dla lokali 14 i 35 na poziomie 1. Oznacza, to że oba lokale różnią się między sobą tylko na jednej pozycji porównywanych cech.

Drugą miarą podobieństwa określającą podobieństwo lokali jest odległość Tanimoto. Wzór na odległość Tanimoto przedstawia się następująco:

$$d\,(x_1, x_2) = \frac{\sum_{k=1}^{n} x_{1k} x_{2k}}{\sum_{k=1}^{n} x_{1k}^2 + \sum_{k=1}^{n} x_{2k}^2 - \sum_{k=1}^{n} x_{1k} x_{2k}} \qquad \text{[Wzór 2]}$$

przy czym:
d – odległość Tanimoto,
x_1, x_2 – porównywane obiekty (lokale),
n – liczba cech porównywanych obiektów,
k – porównywana cecha.

Odległość Tanimoto[1] określa w jakim stopniu porównywane obiekty są do siebie podobne.

Tabela 3. Podobieństwo lokali na podstawie Odległości Tanimoto.

	k_1	k_2	k_3	k_4	Σx_1^2 i Σx_2^2
Lokal 14	1	1	1	0	3
Lokal 35	1	1	1	1	4
licznik	1	1	1	0	

Źródło: opracowanie własne.

Odległość Tanimoto dla porównywanych lokali na podstawie Tabeli 3 kształtuje się na poziomie 0,75. Oznacza to, że oba lokale są do siebie podobne w 75%. Czyli trzy z czterech opisywanych cechy pokryły się ze sobą.

2.2. Podobieństwo wartości liczbowych

Załóżmy, że chcemy dokonać porównania dwóch lokali (obiektów), których wartości reprezentowane są w sposób liczbowy. Tak jak przedstawiono w Tabeli 4.

[1] Odległość Tanimoto można wyznaczać tylko dla obiektów, które zawierają cechy istotne z semantycznego punktu widzenia. Wyznaczanie podobieństwa dla cech nieistotnych nie ma sensu. Ponadto może prowadzić do niewiarygodnych wyników. Dlatego też, w przykładzie zaprezentowano m.in. Lokal 35, który posiada wszystkie porównywane cechy. [Kleć M., Podobieństwo i klasyfikacja dźwięku w kontekście rekomendacji muzycznej, Praca magisterska, Polsko-Japońska Wyższa szkoła Technik Komputerowych, Warszawa 2010, str. 17.].

Tabela 4. Lokale z cechami przedstawionymi w sposób liczbowy

	Liczba pokoi	Piętro lokalu	Wiek budynku [lata]	Dystans do centrum [min]	Odległość przystanku [m]
Lokal 5	5	3	4	51	375
Lokal 9	2	1	2	47	498

Źródło: opracowanie własne.

Obydwa lokale możemy przedstawić przy pomocy wektorów. Lokal 5 opisuje wektor [5,3,4,51,375] z kolei lokal 9 wektor [2,1,2,47,498]. Do wyznaczenia podobieństwa dwóch wektorów posłużymy się następującymi miarami: Odległość Euklidesa, Odległość Euklidesowa z wagami, Kosinus kąta miedzy wektorami oraz Kosinus kąta między wektorami z wagami.

Mogłoby się wydawać, iż teraz wystarczy podstawić oba wektory do wzorów na w/w miary podobieństwa, odczytać ich wartości i zinterpretować otrzymane wyniki. Nic bardziej mylnego. Jak można zauważyć w Tabeli 4 ostatnia cecha, tj. odległość przystanku komunikacji została opisana przy pomocy znacznie wyższych wartości liczbowych niż pozostałe cechy porównywanych lokali. Wyznaczenie podobieństwa w takim przypadku skutkowałoby otrzymaniem niewiarygodnych wyników, gdyż odległość przystanku miałaby znacznie większy wpływ niż pozostałe cechy. Aby temu zapobiec przed analizą dokonamy normalizacji wszystkich cech do wspólnego zakresu od 0 do 1. W tym celu posłużymy się normalizacją min/max.

W wyniku normalizacji danych otrzymamy wektory, których wartości cech zawierają się w przedziale od 0 do 1. Normalizacja przeprowadzana jest na podstawie poniższego wzoru:

$$x_i' = \frac{x_i - x_{i_min}}{x_{i_max} - x_{i_min}} \qquad \text{[Wzór 3]}$$

gdzie:

x_i – wartość znormalizowanej cechy,

x_{i_max} – jest maksymalną wartością występującą w zbiorze dla i-tej cechy,

x_{i_min} – jest minimalną wartością dla i-tej cechy.

Normalizacja nie uwzględnia rozkładu wartości danej cechy, w związku z tym w przypadku wystąpienia w danej cesze wartości znacznie różnych od

przeciętnej, nastąpi ściśnięcie tych wartości w bardzo wąskim przedziale. W przedstawianym przykładzie wartości minimalne i maksymalne przedstawia Tabela 5.

Tabela 5. Wartości minimalne i maksymalne porównywanych lokali.

	Liczba pokoi	Piętro lokalu	Wiek budynku [lata]	Dystans do centrum [min]	Odległość przystanku [m]
Minimum	1	0	0	0	10
Maksimum	5	5	21	60	985

Źródło: opracowanie własne.

Dla lokalu 5 przy cesze liczba pokoi normalizacja będzie przedstawiać się następująco:

$$x_1' = \frac{5-1}{5-1} = 1 \qquad \text{[Wzór 4]}$$

Znormalizowane wartości cech zostały przedstawione w Tabeli 6.

Tabela 6. Znormalizowane wartości cech dla obydwu lokali.

	Znormalizowana liczba pokoi	Znormalizowane piętro lokalu	Znormalizowany wiek budynku	Znormalizowana dystans do centrum	Znormalizowana odległość przystanku
Lokal 5	1,0000	0,6000	0,1905	0,8500	0,3744
Lokal 9	0,2500	0,2000	0,0952	0,7833	0,5005

Źródło: opracowanie własne.

Po dokonaniu normalizacji można dokonać dalszych działań w celu wyliczenia podobieństwa między lokalami. Pierwszym działaniem będzie obliczenie Odległości Euklidesa.

Odległość Euklidesa jest miarą stosowaną do obliczenia odległości między dwoma punktami (obiektami). Definiuje się ją jako pierwiastek z kwadratu różnicy między parami wartości zmiennych. Co przedstawia poniższy wzór.

$$d(x_1, x_2) = \sqrt{\sum_{k=1}^{n}(x_{1k} - x_{2k})^2} \qquad \text{[Wzór 4]}$$

przy czym:
d – odległość Euklidesa,
x_1 x_2 – porównywane obiekty (lokale),
n – liczba cech porównywanych obiektów,
k – porównywana cecha.

Tabela 7. Cząstkowe obliczenia Odległości Euklidesa.

	Znormalizowana liczba pokoi	**Znormalizowane piętro lokalu**	**Znormalizowany wiek budynku**	**Znormalizowana dystans do centrum**	**Znormalizowana odległość przystanku**
Lokal 5	1,0000	0,6000	0,1905	0,8500	0,3744
Lokal 9	0,2500	0,2000	0,0952	0,7833	0,5005
$(x_1 - x_2)^2$	0,5625	0,16	0,009082	0,004449	0,015901

Źródło: opracowanie własne.

Na podstawie Tabeli 7 obliczam Odległość Euklidesa, która wynosi 0,86714.

Przy Odległości Euklidesa można także wyliczyć Odległość Euklidesa z wagami. Różnica polega na założeniu, iż badane cechy nie są tak samo ważne, tj. nie mają takiego samego wpływu na wynik analizy. W związku z tym, wystarczy wprowadzić dodatkową zmienną we wzorze 4 co zostało przedstawione poniżej:

$$d(x_1, x_2) = \sqrt{\sum_{k=1}^{n} g_k(x_{1k} - x_{2k})^2} \qquad \text{[Wzór 5]}$$

przy czym:
– g_k stopień wpływu cechy na wynik analizy.

Załóżmy, że wartość cechy wiek budynku jest dla nas ważniejsza niż pozostałe cechy. Dlatego, też ustalimy wagę dla tej cechy równą 10, tzn. że ta cecha jest 10-krotnie bardziej ważniejsza niż inne brane pod uwagę.

Tabela 8. Cząstkowe obliczenia Odległości Euklidesa z wagami.

	Znormalizowana liczba pokoi	**Znormalizowane piętro lokalu**	**Znormalizowany wiek budynku**	**Znormalizowana dystans do centrum**	**Znormalizowana odległość przystanku**
Lokal 5	1,0000	0,6000	0,1905	0,8500	0,3744
Lokal 9	0,2500	0,2000	0,0952	0,7833	0,5005
$(x_1 - x_2)^2$	0,5625	0,16	0,009081	0,004449	0,015901

Źródło: opracowanie własne.

Na podstawie Tabeli 8 można stwierdzić, iż cząstkowe wyliczenia dla Odległości Euklidesa z wagami nie są znacząco różne niż przedstawione w Tabeli 7 cząstkowe wyliczenia dla Odległości Euklidesa bez wag. Jednakże w ogólnym rozrachunku różnica ta kształtuje sią na poziomie 0,045916 na niekorzyść Odległości Euklidesa z wagami.

Trzecią miarą, którą można zastosować do wyliczenia podobieństwa między lokalami jest Kosinus kąta między wektorami.

Kosinus kąta między wektorami reprezentuje podobieństwo między obiektami. Zaletą tej odległości jest fakt, iż jest ona ograniczona do przedziału [0,1]. Obiekty są do siebie podobne gdy wektory pokrywają się, wówczas Kosinus kąta między wektorami wynosi 1. Z kolei obiekty są różne gdy wektory są prostopadłe, wówczas Kosinus kąta między wektorami wynosi 0. Kosinus kąta między wektorami wylicza się na podstawie wzoru:

$$d\,(x_1, x_2) = \frac{\sum_{k=1}^{n} x_{1k} x_{2k}}{\sqrt{\sum_{k=1}^{n} x_{1k}^2}\sqrt{\sum_{k=1}^{n} x_{2k}^2}} \qquad \text{[Wzór 6]}$$

przy czym:

d – Kosinus kata między wektorami,

x_1, x_2 – porównywane obiekty (lokale),

n – liczba cech porównywanych obiektów,

k – porównywana cecha.

Dla lokali 5 i 9 Kosinus kąta między wektorami wynosi 0,836163.

Ponadto przy obliczaniu Kosinusa kąta między wektorami, można również zastosować wagi. Założenie jest takie samo jak miało to miejsce przy Odległości Euklidesa z wagami, tzn. zakładamy, iż jedna cecha (bądź

więcej) jest dla nas ważniejsza niż pozostałe. Co zostało przedstawione poniżej:

$$d\,(x_1, x_2) = \frac{\sum_{k=1}^{n} g_k x_{1k} x_{2k}}{\sqrt{\sum_{k=1}^{n} g_k x_{1k}^2}\sqrt{\sum_{k=1}^{n} g_k x_{2k}^2}} \qquad \text{[Wzór 7]}$$

przy czym:

– g_k stopień wpływu cechy na wynik analizy.

Dla lokali 5 i 9 Kosinus kąta między wektorami z wagą równą 10 dla wieku budynku wynosi 0,849535. Co oznacza, iż oba obiekty są bardziej do siebie podobne aniżeli według wartości Kosinusa kąta między wektorami bez uwzględniania wag.

3. Zastosowanie wybranych algorytmów rekomendacji – wybór lokalu na podstawie podanych kryteriów

Załóżmy, iż mamy za zadanie zarekomendować lokal najbardziej podobny do szukanego przez użytkownika. Przy czym lokal ten musi spełniać odpowiednie warunki. Po pierwsze musi posiadać następujące cechy: nieruchomość lokalowa, położony w dzielnicy X, położony w niskiej zabudowie wielorodzinnej (do pięciu kondygnacji) oraz musi być nieremontowany. W celu wyodrębnienia tylko tych lokali, które spełniają warunki konieczne zastosowano następujący sposób ich oznaczenia: 1 – oznacza, iż lokal spełnia warunek konieczny, 0 – nie spełnia warunku koniecznego. Dlatego też, aby z bazy 199 lokali wybrać tylko te spełniające powyższe warunki posłużymy się podobieństwem wartości binarnych. Toteż zastosujemy Odległość Tanimoto. Po zastosowaniu Odległości Tanimoto zostało nam już jedenaście lokali spełniających warunki konieczne. Co obrazuje Tabela 9.

Ponadto użytkownik, określił nam cechy opcjonalne, które byłyby wskazane aby lokal ten został uznany za pożądany przez użytkownika, tzn. wskazane by było gdyby lokal posiadał 3 pokoje, byłby zlokalizowany na parterze, wiek budynku nie byłby większy niż 3 lata, odległość do centrum nie przekraczałaby więcej niż 30 minut oraz odległość do przystanku komunikacji miejskiej nie powinna być wyższa niż 80 metrów. Teraz z 11

lokali zarekomendujemy ten, który jest najbardziej podobny według odległości Euklidesa.

Tabela 9. Lokale spełniające warunki konieczne.

	Nieruchomość lokalowa	Lokal w dzielnicy >	< 5 piętra	Loakl nieremontowany	Tanimo
Poszukiwany	1	1	1	1	
Lokal 10	1	1	1	1	1
Lokal 12	1	1	1	1	1
Lokal 17	1	1	1	1	1
Lokal 35	1	1	1	1	1
Lokal 88	1	1	1	1	1
Lokal 101	1	1	1	1	1
Lokal 105	1	1	1	1	1
Lokal 122	1	1	1	1	1
Lokal 130	1	1	1	1	1
Lokal 132	1	1	1	1	1
Lokal 135	1	1	1	1	1

Źródło: opracowanie własne.

Tabela 10. Rekomendacja lokalu według Odległości Euklidesa

	Odległość Euklidesa
Lokal 10	0,60557807
Lokal 12	0,827174845
Lokal 17	0,538164554
Lokal 35	0,506892543
Lokal 88	0,921502688
Lokal 101	0,767310638
Lokal 105	0,702748112
Lokal 122	0,666663371
Lokal 130	0,653756907
Lokal 132	1,062133122
Lokal 135	[illegible]

Źródło: opracowanie własne.

Na podstawie Tabeli 10 można stwierdzić, iż lokal 135 zostałby zarekomendowany użytkownikowi, ponieważ jest najbardziej podobny do lokalu pożądanego przez użytkownika. Z kolei przy zastosowaniu Kosinusa kata

między wektorami lokalem rekomendowanym użytkownikowi byłby lokal 17. Co przedstawia Tabela 11.

Tabela 11. Rekomendacja lokalu według Kosinusa kąta między wektorami.

	Kosinus kąta między wektorami
Lokal 10	0,553597164
Lokal 12	0,064096965
Lokal 17	0,789232008
Lokal 35	0,739284613
Lokal 88	0,474016334
Lokal 101	0,627229042
Lokal 105	0,420951301
Lokal 122	0,39446685
Lokal 130	0,496865336
Lokal 132	0,595291022
Lokal 135	0,75182759

Źródło: opracowanie własne.

Załóżmy, iż dla użytkownika spośród cech opcjonalnych ważniejsze znaczenie przy wyborze lokalu ma wiek budynku oraz dystans do centrum. W związku z czym, zastosujemy teraz Odległość Euklidesa oraz Kosinus kąta między wektorami odpowiednio z wagami dla poszczególnych cech: liczba pokoi – 1, piętro lokalu – 1, wiek budynku – 10, dystans do centrum – 10 oraz odległość do przystanku – 1. Teraz według Odległości Euklidesa z wagami rekomendowanym lokalem byłby lokal 17 (Tabela 12).

Tabela 12. Rekomendacja lokalu według odległości Euklidesa z wagami.

	Odległość Euklidesa z wagami
Lokal 10	1,642667785
Lokal 12	1,736620533
Lokal 17	[illegible]
Lokal 35	1,41531393
Lokal 88	1,611006106
Lokal 101	1,959276809
Lokal 105	1,711002156
Lokal 122	1,480072133
Lokal 130	1,917654321
Lokal 132	1,145338482
Lokal 135	1,407342716

Źródło: opracowanie własne.

Z kolei według Kosinusa kąta między wektorami z uwzględnieniem wag lokalem rekomendowanym byłby lokal 132 (Tabela 13).

Tabela 13. Rekomendacja lokalu według Kosinusa kąta między wektorami z uwzględnieniem wag.

	Kosinus kąta między wektorami z wagan
Lokal 10	0,309567582
Lokal 12	0,101500152
Lokal 17	0,84917155
Lokal 35	0,568416939
Lokal 88	0,469978874
Lokal 101	0,354759229
Lokal 105	0,305520241
Lokal 122	0,511105395
Lokal 130	0,306941182
Lokal 132	[illegible]
Lokal 135	0,577683769

Źródło: opracowanie własne.

Zestawienie lokali rekomendowanych według różnych miar wyznaczania podobieństwa między lokali przedstawia Tabela 14.

Tabela 14. Rekomendacja lokali według różnych miar wyznaczania podobieństwa.

	Odległość Euklidesa	Odległość Euklidesa z wagami	Kosinus kąta między wektorami	Kosinus kąta między wektorami z wagami
Lokal 10	0,60557807	1,642667785	0,553597164	0,309567582
Lokal 12	0,827174845	1,736620533	0,064096965	0,101500152
Lokal 17	0,538164554	0,930679715	0,789293008	0,84917155
Lokal 35	0,506892543	1,41531393	0,739284613	0,568416939
Lokal 88	0,921502688	1,611006106	0,474016334	0,469978874
Lokal 101	0,767310638	1,959276809	0,627229042	0,354759229
Lokal 105	0,702748112	1,711002156	0,420951301	0,305520241
Lokal 122	0,666663371	1,480072133	0,39446685	0,511105395
Lokal 130	0,653756907	1,917654321	0,496865336	0,306941182
Lokal 132	1,062133122	1,145338482	0,595291022	0,852346381
Lokal 135	0,484190007	1,407342716	0,75182759	0,577683769

Źródło: opracowanie własne.

4. Podsumowanie

W niniejszym artykule zostały opisane sposoby rekomendacji produktów z uwzględnieniem różnych miar wyznaczania podobieństwa między obiektami. Należy zwrócić uwagę, iż dobór miar wyznaczania podobieństwa zależny jest nie tylko od sposobu reprezentacji cech produktów rekomendowanych ale takżae od ich istotności dla użytkownika. Przeprowadzone doświadczenie dotyczące rekomendacji lokali pokazuje jak ważny jest wybór cech, które reprezentują w algorytmach porównywane obiekty: jest to „wiedza", która została przez człowieka przekazana systemowi komputerowemu, celem osiągnięcia działania jak najbardziej zbliżonego do naturalnych umiejętności ludzkich.

Literatura

1. Wójcik K., *Analiza porównawcza miar podobieństwa tekstów*, Taksonomia 17, Wydawnictwo UE, Wrocław 2010.
2. Kleć M., Podobieństwo i klasyfikacja dźwięku w kontekście rekomendacji muzycznej, Praca magisterska, Polsko-Japońska Wyższa Szkoła Technik Komputerowych, Warszawa 2010.
3. Kański K., Optymalizacja systemu rekomendacyjnego na podstawie bazy USOS, Praca magisterska, UW, Warszawa 2012.
4. Tuchowski J., Wójcik K., *Analiza porównawcza miar podobieństwa tekstów opartych na macierzy częstości i tekstów opartych na wiedzy dziedzinowej*, Taksonomia 19, Wydawnictwo UE, Wrocław 2012.
5. http://pqstat.pl/, z dnia 20.04.2016 r.
6. https://mi007.wikispaces.com/file/view/funkcje.pdf, z dnia 20.04.2016 r.