

Tadeusz Ratajczak

PRZYKŁADY ZASTOSOWANIA WNIOSKOWANIA Z BAZY PRZYPADKÓW W INFORMATYCZNYCH SYSTEMACH WSPOMAGAJĄCYCH PODEJMOWANIE DECYZJI

1. Wprowadzenie

Metodologię wnioskowania z bazy przypadków (ang. *case-based reasoning* – CBR) można określić jako filozofię rozwiązywania problemów praktycznych, porównywaną z regułowymi systemami eksperckimi czy sieciami neuronowymi. Zaleca się w niej wnioskowanie przez analogię do zdarzeń zaistniałych w przeszłości i wykorzystanie doświadczeń z tych zdarzeń. Metodologia ta ma kilka zalet, m.in.:

- systemy informatyczne z bazami przypadków nie wymagają zbudowania modelu matematycznego badanego fragmentu rzeczywistości; w konsekwencji wyniki przetwarzania systemu nie są obciążone błędami metody modelowania;
- systemy z bazami przypadków zawsze dają możliwość odczytania przesłanek, na podstawie których otrzymano ostateczny rezultat (też zalety nie posiadają np. sieci neuronowe);
- budując system wnioskujący z bazy przypadków, nie musimy wiedzieć, jak rozwiązać problem (też zalety nie posiadają np. regułowe systemy eksperckie);
- wnioskowanie z bazy przypadków jest odpowiednią metodologią projektowania systemów informatycznych w przypadku, gdy problem jest opisany za pomocą logiki rozmytej.

Celem niniejszej pracy jest pokazanie przykładów systemów informatycznych zbudowanych według metodologii wnioskowania z bazy przypadków, które miały za zadanie wspomaganie decydenta w procesie podejmowania decyzji. Przykłady zostały wybrane z różnych dziedzin, a wszystkie łączy fakt, że zastosowana metodologia CBR przyczyniła się do podniesienia skuteczności tych systemów. Systemy opisane w punktach 3 i 4 były opracowywane pod kierunkiem autora. W punkcie 5 zostały zwięźle scharakteryzowane systemy informatyczne zbu-

dowane według metodologii CBR, które do chwili obecnej są eksploatowane w trzech firmach znanych na rynku światowym.

W punkcie 2 jest scharakteryzowana metodologia CBR. W podsumowaniu artykułu zostały wymienione warunki, które musi spełniać modelowany wycinek rzeczywistości, aby od systemów informatycznych można było oczekiwać pozytywnych wyników.

2. Wnioskowanie z bazy przypadków

Wnioskowanie z bazy przypadków jest techniką rozwiązywania problemów, zakładającą posiadanie przez rozwiązującego problem dużego zbioru opisów sytuacji i sposobów postępowania w tych sytuacjach. Te opisy nazywamy przypadkami (ang. *case*). Opis przypadku zawiera lekcję z przeszłości i kontekst, w którym ta lekcja miała miejsce. Przypadek może być pewnym zdarzeniem, historią lub opisem pewnego problemu. Klasycznie przypadek składa się z trzech rodzajów informacji:

- opis problemu (epizodu zaistniałego w przeszłości),
- opis zastosowanej metody rozwiązania problemu,
- konsekwencje wywołane zastosowaniem rozwiązania problemu.

Gdy pojawi się nowy problem, jest on klasycznie rozwiązywany [Kolodner 1993; Case-based... 1993; Lenz i in. 1998] w czterech kolejnych krokach.

1. Po zdefiniowaniu problemu przez użytkownika przeszukiwana jest baza przypadków w celu znalezienia w niej problemu najbardziej podobnego (problemów najbardziej podobnych) do zagadnienia, które należy rozwiązać (ang. *retrieve*).

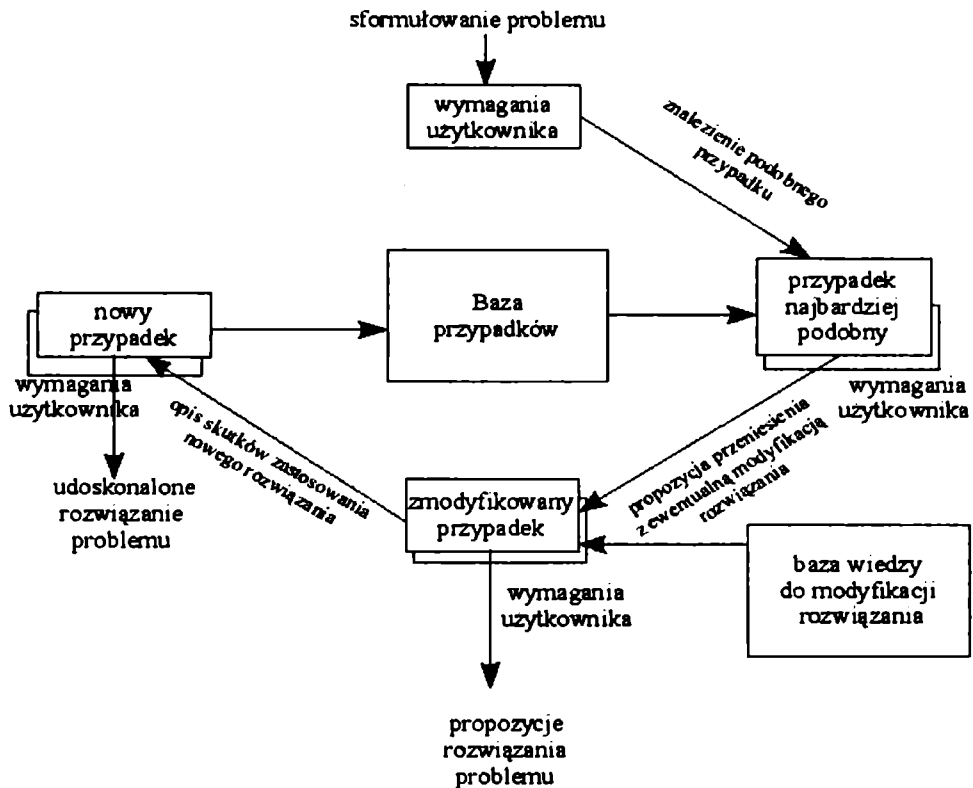
2. Następuje próba przeniesienia rozwiązania problemu najbardziej podobnego na problem rozwiązywany przez użytkownika (ang. *reuse*).

3. W większości przypadków „stare” rozwiązanie musi zostać zmodyfikowane (zaadaptowane), aby spełniało wymagania problemu rozwiązywanego (ang. *revise*).

4. Nowe rozwiązanie problemu wraz z opisem skutków jego zastosowania jest zapamiętywane w bazie jako kolejny przypadek (ang. *retain*).

Na rysunku 1 przedstawiono schemat wnioskowania z bazy przypadków. Jest to ogólny model wnioskowania i może on być odpowiednio zmodyfikowany w zależności od konkretnych wymagań.

Pierwszym krokiem projektowania systemu wnioskującego z bazy przypadków jest zaprojektowanie struktury przypadku. Oznacza to [Case-based... 1993; Kolodner 1993]: wyłonienie atrybutów mających najważniejszy wpływ na rozwiązanie problemu, określenie indeksu bazy, pokazanie, w jaki sposób będzie zapamiętane rozwiązanie problemu oraz co oznaczają konsekwencje wdrożenia rozwiązania i w jaki sposób one będą zapamiętane. Następnym krokiem projektowania systemu jest określenie kryterium podobieństwa przypadków, wykorzystywane do znajdowa-



Rys 1. Klasyczny cykl wnioskowania z bazy przypadków

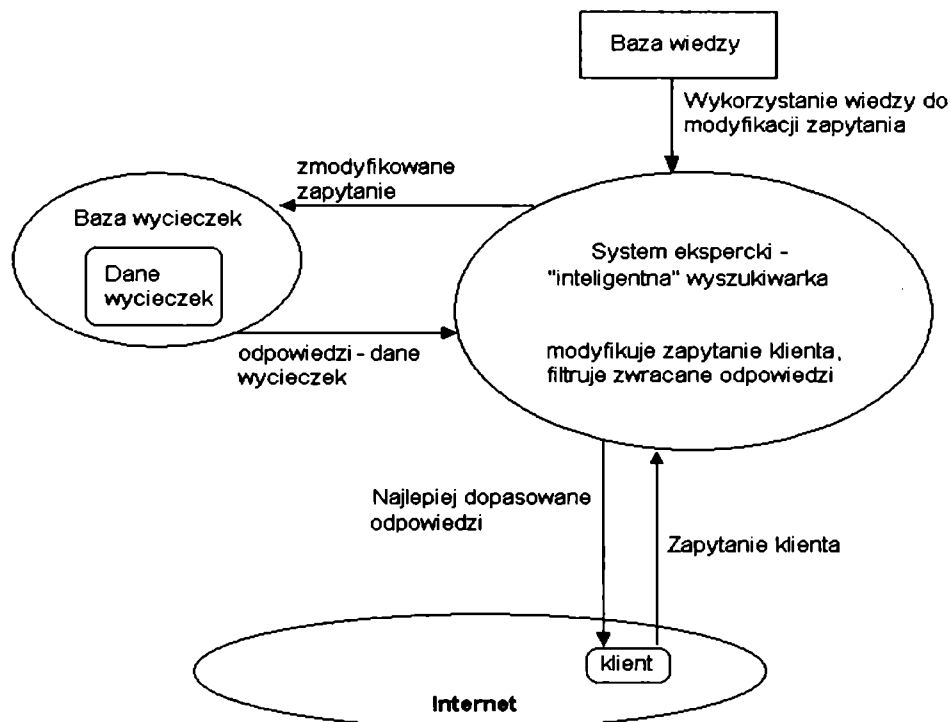
Źródło: opracowanie własne.

nia przypadku najbardziej podobnego. Kryterium to ustalamy, wskazując atrybuty, względem których rozważamy podobieństwo. Z kolei należy zdefiniować miarę podobieństwa przypadków, ażeby liczbowo móc ocenić stopień podobieństwa przypadków.

3. Wyszukiwanie wycieczek

Gdy klient nie ma jasno sprecyzowanej wizji wycieczki, a biura podróży oferują wiele różnorodnych wycieczek, to może on mieć problem ze znalezieniem właściwej oferty. W ostatnich latach biura podróży zwiększają sprzedaż poprzez wprowadzenie systemów B2C (*business to consumer*). Biura nie tylko zamieszczają na swoich stronach internetowych pełną ofertę organizowanych przez siebie wycieczek oraz informacje o aktualnych promocjach, ale też umożliwiają klientowi dokonanie rezerwacji poprzez Internet. Aby ułatwić klientowi wybór wycieczki,

portale biur podróży udostępniają wyszukiwarki. Niestety, wyszukiwarki te nie realizują na ogół zamierzonej funkcji. Wyświetlają one oferty spełniające jeden lub kilka warunków logicznych, a nie podejmują próby głębszej analizy potrzeb klienta. W [Kasperowicz 2004] podjęto próbę zbudowania internetowej wyszukiwarki wycieczek dla klientów, którzy nie mają jasno sprecyzowanej wizji swojego wypoczynku.



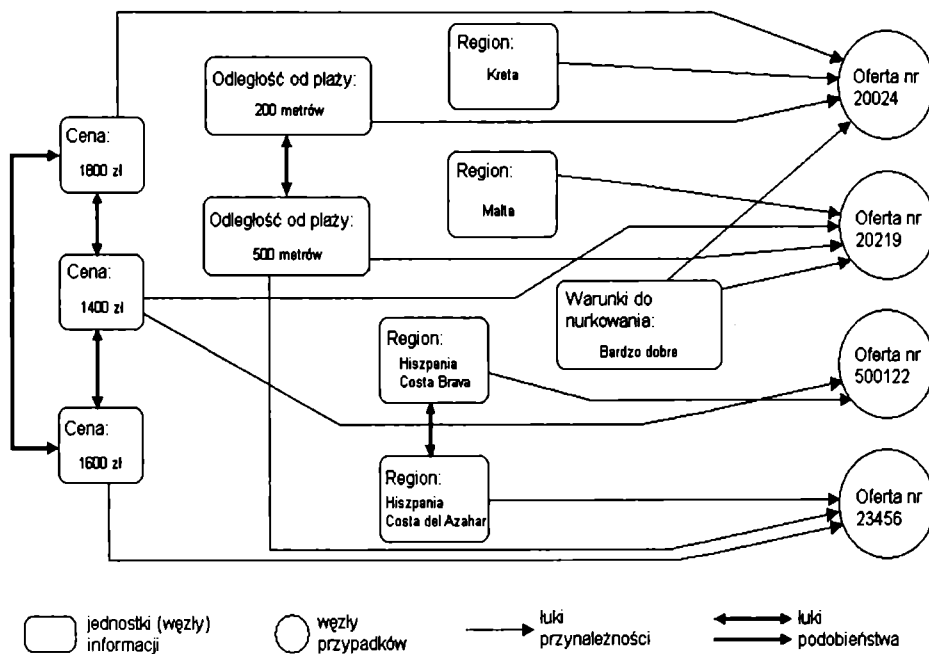
Rys. 2. Schemat wyszukiwarki wycieczek działającej jako system ekspercki

Źródło: [Kasperowicz 2004].

Ogólną koncepcję tej wyszukiwarki przedstawiono na rys. 2. Wyszukiwarka, będąca systemem informatycznym, realizuje następujące zadania:

- przechowuje zbiór dostępnych ofert,
- interpretuje życzenia i preferencje klienta oraz hierarchię ich ważności,
- odnajduje najbardziej odpowiadające klientowi oferty wycieczek,
- jeśli w bazie nie ma ofert w pełni odpowiadających wymaganiom klienta, wyszukiwarka znajduje oferty najbardziej zbliżone do wymagań klienta,
- filtruje skonstruowane odpowiedzi, aby wyświetlić klientowi tylko 3 najlepiej dopasowane oferty,
- umożliwia użytkownikowi pracę iteracyjną.

Rdzeniem systemu jest sieć wyszukiwania przypadków. Jest to graf z dwoma rodzajami węzłów i dwoma rodzajami łuków. Pierwszym typem węzłów są jednostki informacji zapisywane za pomocą pary <atrybut; wartość>, drugim typem są oferty biur podróży traktowane w systemie jako przypadki. Przypadek jest zbiorem jednostek informacji. Węzły jednostek informacji są połączone z węzłami przypadków za pomocą jednokierunkowych łuków przynależności. Połączenie to oznacza, że dany przypadek posiada atrybut o wartości zapisanej w jednostce informacji. Drugi typ połączeń stanowią jedno- lub dwukierunkowe łuki podobieństwa. Mogą one łączyć ze sobą poszczególne jednostki informacji. Połączenie to oznacza, że pomiędzy wartością atrybutu zapisaną w jednej jednostce informacji a inną wartością atrybutu zapisaną w drugiej jednostce informacji występuje podobieństwo. Dołączone do łuków wagi wyrażają stopień podobieństwa lub przynależności. Przykład sieci wyszukiwania przypadków przedstawiono na rys. 3.



Rys. 3. Przykład sieci wyszukiwania przypadków

Źródło: [Kasperowicz 2004].

Wyszukiwanie ofert odbywa się w trzech krokach.

1. Aktywowanie węzłów informacji wymienionych w zapytaniu klienta.
2. Powielenie aktywacji na węzły informacji połączone łukami podobieństwa z węzłami aktywowanymi w kroku 1.

3. Wyliczenie stopnia aktywacji i aktywowanie tych węzłów przypadków, które są połączone z aktywowanymi węzłami informacji.

Podstawowymi zaletami sieci wyszukiwania przypadków są: zdolność do operowania na pojęciach rozmytych oraz stosunkowo krótki czas znajdowania właściwego przypadku.

4. Prognozowanie popytu

W [Sroczyński 2002] metoda CBR została zastosowana do prognozowania k -krokowych średnich kroczących szeregu $x_{1t}, x_{2t}, \dots, x_{nt}$ dziennej sprzedaży ustalonego towaru t . Założono, że szereg $x_{1t}, x_{2t}, \dots, x_{nt}$ podlega trendom i że wartość sprzedaży towaru t w przeszłości ma wpływ na wartość jego sprzedaży w przyszłości.

$$Y = \{y_k, y_{k+1}, \dots, y_r, \dots, y_n\}, \quad \text{gdzie } y_r = \frac{\sum_{i=1}^k x_{r-i+1,t}}{k}.$$

W języku metody CBR danymi zadania, które należy rozwiązać, jest ciąg $Y_n = \{y_{n-r+1}, y_{n-r+2}, \dots, y_n\}$ składający się ze średnich kroczących y_r , odnoszących się do ostatnich r dni sprzedaży. Poszukiwane są wielkości $y_{n+1}, y_{n+2}, \dots, y_{n+l}$ (prognozy popytu na towar t na 1, 2, ..., l dni naprzód). Bazę przypadków tworzą wszystkie podciągi ciągu Y postaci

$$Y_p = \{y_{p-r+1}, y_{p-r+2}, \dots, y_p, y_{p+1}, y_{p+2}, \dots, y_{p+l}\},$$

w których r początkowych wyrazów definiuje zadanie prognozowania, a l końcowych stanowi znalezione prognozy popytu na odpowiednio 1, 2, ..., l dni naprzód. Podobieństwo między przypadkami Y_n oraz Y_p jest mierzone za pomocą metryki

$$d(Y_p, Y_n) = w_0 \sum_{i=0}^{r-1} (y_{p-i} - y_{n-i})^2 + w_1 \sum_{i=0}^{r-1} (y'_{p-i} - y'_{n-i})^2 + w_2 \sum_{i=0}^{r-1} (y''_{p-i} - y''_{n-i})^2,$$

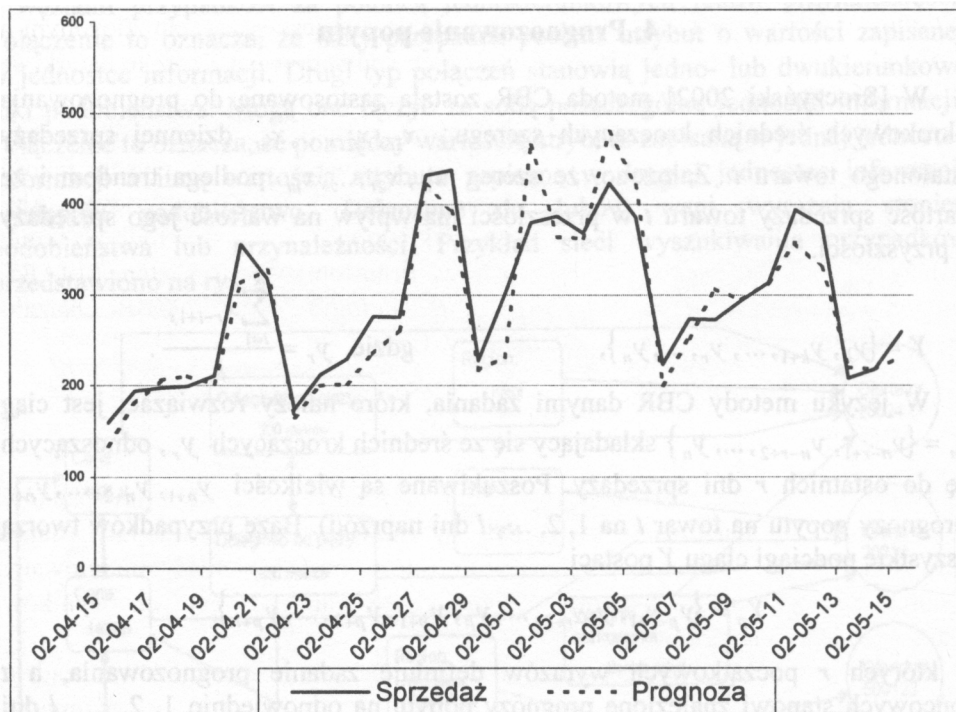
gdzie y'_j, y''_j oznaczają wartości odpowiednio pierwszej i drugiej pochodnej średniej kroczącej y_j , natomiast w_0, w_1, w_2 są arbitralnie dobranymi wagami.

Prognozowanie popytu na towar t na 1, 2, ..., l dni naprzód polega na tym, aby dla zadanego Y_n znaleźć w bazie przypadków taki Y_p , który spełnia warunek:

$$d(Y_p, Y_n) = \min_i d(Y_i, Y_n).$$

Rozwiązaniem są wartości $y_{p+1}, y_{p+2}, \dots, y_{p+l}$, oznaczające prognozowany popyt na towar t na 1, 2, ..., l dni naprzód.

Na rysunku 4 pokazano ilość sprzedanego piwa w małym, elbląskim sklepie spożywczym w okresie 15.04.2002 ÷ 15.05.2002 oraz wcześniej opracowaną – za pomocą omawianej metody – prognozę sprzedaży. Przykład ten dobrze ilustruje jakość metody.



Rys. 4. Porównanie prognozy popytu na piwo do rzeczywistej wielkości sprzedaży
Źródło: [Sroczyński 2002].

Błąd względny metody nie przekracza 23%. Warto zwrócić uwagę, że generowane prognozy wykazują tygodniową cykliczność charakterystyczną dla popytu na piwo oraz dłuższy wzrost popytu podczas długiego weekendu na początku maja.

5. Inne zastosowania metody CBR

Metodologię CBR zastosowały m.in. następujące firmy.

1. Deloitte & Touche. Firma wykorzystuje system TMFDT (*top management fraud diagnostic tool*), który pomaga audytorom oszacować prawdopodobieństwo oszustw zarządu kontrolowanej firmy. System TMFDT został zaprojektowany, aby wspomagać audytorów. System pokazuje audytorom przypadki, które są podobne

do sytuacji w kontrolowanej firmie. Dzięki temu audytorzy mają mniej kłopotu z wydaniem odpowiedniej opinii. Audyt jest procesem złożonym, łączącym w sobie wiele czynników. Z tego względu system doradczy, wnioskujący przez analogię do zaistniałych w przeszłości zdarzeń, jest pomocnym narzędziem w pracy audytora.

2. British Airways zastosowała system Caseline wykorzystujący CBR w celu zminimalizowania czasu potrzebnego do diagnozy silników samolotów Boeing 747-400. Wykrycie usterek w silnikach samolotów jest procesem złożonym i niezwykle trudnym. Opóźnienia linii lotniczych spowodowane przez te usterki są bardzo kosztowne.

3. Compaq. Dział marketingu tej firmy, produkującej komputery osobiste oraz wysokiej jakości serwery, zauważył, że dla firmy najważniejsze są: wysoka jakość wytworzonego produktu i szybkie, wysokiej jakości usługi serwisowe. Aby wspomóc konsultantów z punktów obsługi klientów, którzy nie byli w stanie obsłużyć lawinowo rosnącej liczby klientów, firma opracowała system SMART (*support management automated reasoning technology*) oparty na CBR. Okazało się, że system ten rozwiązuje 95% problemów firmy.

6. Podsumowanie

Nie we wszystkich dziedzinach system informatyczny z bazą przypadków może okazać się narzędziem pomocnym w rozwiązywaniu problemów. Janet Kolodner [1993] wymienia trzy warunki, które musi spełniać modelowany wycinek rzeczywistości, aby można było oczekiwać pozytywnych wyników zastosowania tych systemów.

- *Regularność*. Modelowanym wycinkiem rzeczywistości powinny rządzić pewne prawa; w związku z tym zjawiska tej rzeczywistości są przewidywalne. Innymi słowy: wykonanie po raz wtóry tych samych czynności w tych samych (lub bardzo podobnych) warunkach powinno prowadzić do takich samych (lub bardzo podobnych) wyników.
- *Powtarzalność zjawisk*. Zjawiska występujące w modelowanej rzeczywistości mają tendencję do powtarzania się.
- *Ciągłość modelowanej rzeczywistości*. Gdy nastąpią małe zmiany w modelowanym wycinku rzeczywistości, to dla otrzymania właściwego rozwiązania problemu w zmienionej rzeczywistości wystarczy dokonać małych zmian w starym rozwiązaniu problemu.

Z doświadczeń [*Case-based...* 1996; Kasperowicz 2004; Lenz i in. 1998; Sroczyński 2002] wiadomo, że kolejne trzy właściwości są istotne dla tego sposobu rozwiązywania problemów.

- Kluczową sprawą jest *podobieństwo* zdarzeń, sytuacji, czynności. Nie jest istotny sposób mierzenia podobieństwa. Istotne jest, że w dziedzinie, którą badamy, wydarzenia i sytuacje są podobne.

- Baza przypadków musi być *obszerna*. Zwróćmy uwagę, że sugerowane przez metodologię znajdowanie w bazie przypadku najbardziej podobnego jest zawsze realizowane. Od wielkości i jakości bazy zależy, czy znaleziony przypadek będzie bardziej czy mniej podobny do zadanego.
- Modelowanie musi dotyczyć *wąskiego wycinka* rzeczywistości. Aby móc wyposażyć komputer w wiedzę porównywalną z tą, którą posiada człowiek, dziedzina musi być bardzo wąska.

Literatura

- Case-based Reasoning: Experiences, Lessons&Future Directions*, red. D.B. Leake, Morgan Kaufmann Publishers, Inc., San Francisco (California) 1996.
- Kasperowicz M., *Witryna internetowa B2C dla biura podróży działająca w oparciu o bazę przypadków* (praca dyplomowa magisterska), Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, Gdańsk 2004.
- Kolodner J., *Case-based Reasoning*, Morgan Kaufmann Publishers, Inc., San Francisco (California) 1993.
- Lenz M. i in., *Case-based Reasoning Technology from Foundations to Applications*, Springer-Verlag, Berlin-Heidelberg 1998.
- Sroczyński P., *Zastosowanie metodologii wnioskowania z bazy przypadków do wyznaczania optymalnego zaopatrzenia sklepu* (praca dyplomowa magisterska), Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, Gdańsk 2002.

SOME EXAMPLES OF CASE-BASED REASONING APPLICATION IN INFORMATICS DECISION SUPPORTING SYSTEMS

Summary

Case-based reasoning as a methodology of expert systems designing is shortly described at the beginning of the paper. Next, some informatics decision supporting systems to which application of case-based reasoning methodology resulted in the system efficiency increase are characterized. Finally, the basic principles for designing case-based systems are specified.

Dr Tadeusz Ratajczak jest adiunktem w Katedrze Algorytmów i Modelowania Systemów Politechniki Gdańskiej
e-mail: tadra@eti.pg.gda.pl