

Rozdział 1

Wstęp

1.1. Teza, cel i zakres pracy

Punkt wyjścia, okoliczności powstania problemu, niewystarczalność istniejących rozwiązań

Metoda badawcza

- Studia literaturowe
- Analiza budowy i działania istniejących produktów
- Projektowanie i prototypowanie nowatorskich rozwiązań
- Obliczenia i

Co, w którym rozdziale. Oto zakres materiału będący treścią pracy. Na początku zawarto przegląd metod tworzących podsumowania dokumentów tekstowych. Spojrzenie na to zagadnienie z perspektywy kilkunastu ostatnich lat zostało opisane w Rozdziale 1.

Rozdział ?? zawiera podstawowe definicje zbiorów i pojęć wykorzystywanych w rozprawie, takie jak zbiory klasyczne, zbiory rozmyte i zmienna lingwistyczna oraz ich rozszerzenia — intuicjonistyczne zbiory rozmyte i zbiory rozmyte drugiego rodzaju.

Rozdział 2 zawiera opis podsumowań Yager’a oraz rozszerzenia tych podsumowań tj. podsumowania wieloatrybutowe George’a i Srikantha oraz wskaźniki jakości oceniające poprawność podsumowania. Przedstawiono także dalsze prace oparte o teorię podsumowań lingwistycznych, w tym istotne wyniki uzyskane przez Kacprzyka [4] — [10].

W Rozdziale 3 znajdują się przykłady zastosowań opisanych wcześniej metod tworzenia podsumowań lingwistycznych oraz opis systemu komputerowego automatycznie generującego podsumowania wraz z wnioskami wynikającymi z testowania tego systemu.

Kolejne części to bibliografia, spis rysunków i tabel oraz skorowidz najważniejszych pojęć występujących w rozprawie.

Rozdział 2

Nowe sposoby generowania podsumowań lingwistycznych

Rozwój technologii komputerowych, a także sposobu i ilości zapisywanych danych wymusił nowe podejście do generowania podsumowań lingwistycznych. Obecnie w bazach danych przechowywane są przeróżne informacje zawierające dane w postaci numerycznej i tekstowej. Duża ilość tych ostatnich zapisywana jest w postaci zdań, wyrażeń, czy też różnej wielkości plików tekstowych. Dotychczasowy sposób tworzenia podsumowań nie jest w stanie zapewnić uniwersalności w przypadku tak różnorodnych sposobów reprezentacji danych w bazie. Kolejne podejścia do generowania podsumowań radzą sobie z informacjami w postaci tekstowej (szukane jest bowiem podobieństwo do wzorca podsumowania), jak również z danymi znajdującymi się w plikach tekstowych [12].

Generowanie podsumowań lingwistycznych nadal opiera się na koncepcji, którą zaproponował Yager [17, 18, 20, 19], czyli będą to zdania oznajmujące składające się z następujących określeń:

- określenia liczności Q ,
- przedmiotu podsumowania P ,
- interesującej cechy S ,
- wskaźnika jakości podsumowania T .

Przedstawiona tu metoda generowania podsumowania jest rozwinięciem idei Yager’a. Pozwala ona wykonać podsumowanie na różnych typach danych. Dobrze radzi sobie z danymi przedstawionymi w tradycyjnej numerycznej formie oraz z opisami przedstawionymi w języku naturalnym (zdania znajdujące się w bazie, czy też umieszczone w plikach tekstowych). Zatem użytkownik bazy ma możliwość otrzymania odpowiedzi na dowolne pytanie dobrze określone w sensie istniejącej bazy, w rodzaju „*Ilu pacjentów zostało przyjętych w wieku średnim?*” lub „*Ilu pacjentów zostało przyjętych w stanie ogólnym dobrym?*”. Odpowiedzi na te pytania są automatycznie generowane na podstawie podsumowań lingwistycznych, zatem odpowiedzią jest również zdanie w języku naturalnym zawierające określenie liczności przedstawione przy użyciu odpowiedniego zbioru rozmytego. Przykładowe podsumowania mogłyby mieć postać: „*Większość pacjentów została przyjęta w wieku średnim.*” lub „*Niewiele pacjentów została przyjęta w stanie ogólnym dobrym.*”.

Kolejne podrozdziały, to propozycje wykorzystania różnych mechanizmów do tworzenia podsumowań. W pierwszym opisano generowanie podsumowań na dokumentach tekstowych (jest to nowy element w podsumowaniach). W kolejnym — omówiono możliwość podsumowywania dokumentów mieszanych (zarówno danych numerycznych, jak i wyrażonych w języku naturalnym). W celu zmniejszenia niepewności wynikającej z porównywania tekstów wprowadzono dodatkową ilościową miarę, która ocenia niepodobieństwo porównywanych wyrażen. Następnie pokazano dwa nowe sposoby wyboru rekordów do generowania podsumowań: na podstawie selekcji ostrej i wnioskowania rozmytego. Ostatni z podrozdziałów przedstawia podsumowania z użyciem zbiorów rozmytych drugiego rodzaju.

Do ilustracji (Tabela 2.1) tych zagadnień wybrano przejrzysty przykładowy zbiór rekordów, na którym będą generowane podsumowania.

Lp.	Wiek pacjenta	Stan pacjenta podczas przyjęcia do szpitala
1.	20	Stan pacjenta dość dobry.
2.	16	Pacjent w stanie ogólnym dobrym.
3.	73	Stan ogólny dobry.
4.	59	Stan ogólny pacjenta dobry.
5.	19	Pacjent w stanie ogólnym dobrym.
6.	23	Pacjent w stanie ogólnym dobrym.
7.	12	Pacjentka w stanie ogólnym dobrym.
8.	27	Pacjent w stanie ogólnym bardzo dobrym.
9.	26	Stan ogólny pacjenta bardzo ciężki.
10.	65	Pacjentka w stanie ogólnym średnio dobrym.

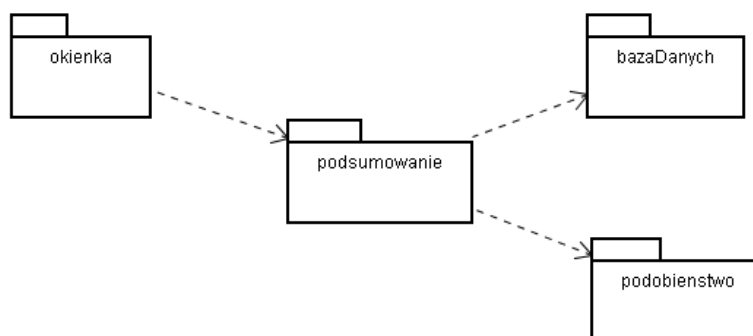
Tabela 2.1. Przykładowa baza danych pacjentów

Rozdział 3

Implementacja

System napisany jest w języku Java z wykorzystaniem pakietu `java.swing.*`, a zatem jest on niezależny od platformy, na której zostanie uruchomiony. Program, który wspomaga uruchomienie i obsługę tego programu to Eclipse — jest to platforma, która jest środowiskiem służącym do projektowania aplikacji, jak również umożliwiającym tworzenie własnych narzędzi.

W oparciu o poprzednie podrozdziały można wnioskować, że system składa się z trzech warstw: warstwy danych (bazy danych), warstwy biznesowej (umożliwiającej przetwarzanie danych) oraz warstwy prezentacji (służącej do komunikacji z użytkownikiem). System ten składa się zatem z klas, które odpowiadają za poszczególne zadania systemu — pobierania danych od użytkownika lub z bazy, ich przetwarzania oraz wysyłania danych na ekran. Poniżej zostanie przedstawiony opis systemu z wykorzystaniem notacji UML¹.



Rysunek 3.1. Pakiety systemu

System podzielony jest na pakiety (co widać na rysunku 3.1), które grupują klasy odpowiedzialne za działanie systemu. Klasy z danych pakietów komunikują się ze sobą w celu wymiany danych potrzebnych do wygenerowania podsumowania lingwistycznego. W pakiecie `okienka` znajdują się wszystkie klasy odpowiedzialne za komunikację pomiędzy użytkowni-

¹ UML — Unified Modeling Language, to graficzny język do obrazowania, specyfikowania, tworzenia i dokumentowania elementów systemów informatycznych [2].

kiem, a systemem (dokładny opis klas znajduje się poniżej). Klasy z tego pakietu odwołują się do klas z pakietu **podsumowanie**, które po otrzymaniu odpowiednich danych pobranych z bazy (poprzez klasy z pakietu **bazaDanych** i po przetworzeniu ich przez metody umieszczone w klasach z pakietu **podobienstwo** tworzą podsumowanie lingwistyczne wraz z wartościami poprawności podsumowania. Takie podsumowanie jest następnie wyświetlane użytkownikowi ponownie poprzez klasy z pakietu **okienka**.

Bibliografia

- [1] Bandemer H., Gottwald S., *Fuzzy Sets, Fuzzy Logic, Fuzzy Methods with Applications*, John Willey and Sons, England, 1995.
- [2] Booch G., Rumbaugh J., Jacobson I., *UML przewodnik użytkownika*, Wydawnictwo Naukowo-Techniczne, Warszawa.
- [3] Buckley J.J., Eslami E., *An Introduction to Fuzzy Logic and Fuzzy Sets*, Physica-Verlag, Heidelberg, 2002.
- [4] Kacprzyk J., Zadrozny S., *On Linguistic Approaches in Flexible Querying and Mining of Association Rules*, In. Larsen H.L., Kacprzyk J., Zadrozny S., Andreassen T., Christiansen H., *Flexible Query Answering Systems*, Physica Verlag, Heidelberg, 475-484, 2001.
- [5] Kacprzyk J., Zadrozny S., *Computing with words in intelligent database querying: standalone and Internet-based applications*, Information Sciences, 34, 71-109, 2001.
- [6] Kacprzyk J., Zadrozny S., *Fuzzy querying for Microsoft Access*, In. Proceedings of the Third IEEE Conference on Fuzzy Systems, Orlando, USA, col. 1, 167-171, 1999.
- [7] Kacprzyk J., Zadrozny S., *Fuzzy queries in Microsoft Access: towards a 'more intelligent' use of Microsoft Windows based DBMSs*, In. Proceedings of the Second Australian and New Zealand Conference on Intelligent Information Systems-ANZIIS'94, Brisbane, Australia, 492-496, 1994.
- [8] Kacprzyk J., Zadrozny S., *FQUERY for Access: fuzzy querying for a Windows-based DBMS*, In. Bosc P., Kacprzyk J., *Fuzziness in Database Management System*, Physica-Verlag, Heidelberg, 415-433, 1995.
- [9] Kacprzyk J., Zadrozny S., *Flexible querying using fuzzy logic: An implementation for Microsoft Access*, In. Andreassen T., Christiansen H., Larsen H.L., *Flexible Query Answering System*, Kluwer Academic Publishers, Boston, 247-275, 1997.
- [10] Kacprzyk J., Zadrozny S., *Implementation of OWA operators in fuzzy querying for Microsoft Access*, In. Yager R.R., Kacprzyk J., *The Ordered Weighted Averaging Operators: Theory and Applications*, Kluwer Academic Publishers, Boston, 293-306, 1997.
- [11] Łęski J., Straszcka E., *Zbiory rozmyte i ich zastosowanie w diagnostyce medycznej*, In. Zajdel R., Kącki E., Szczepaniak P., Kurzyński (red.), *Kompendium informatyki medycznej*, Medica-press, Bielsko-Biała, 2003.
- [12] Ochelska J., Niewiadomski A., Szczepaniak P. S., *Linguistic Summaries Applied To Medical*

- Textual Databases*, Dept. of Electronics & Computer Systems University of Silesia, Ustroń, 2001.
- [13] Ochelska J., Szczepaniak P., Niewiadomski A., *Automatic Summarization on Standarized Textual Databases Interpreted in Terms of Intuitionistic Fuzzy Sets*, W: Soft Computing Tools, Techniques and Applications, Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2004.
- [14] Ochelska J., Szczepaniak P.S., *Textual Fuzzy Similarity and Sequence Kernels*, XIII KOnferencja Sieci i Systemy Informatyczne, Łódź, 299-304, 2005.
- [15] Rasiowa H., *Wstęp do matematyki współczesnej*, Wydawnictwo Naukowe PWN, Warszawa, 1998.
- [16] Rutkowska D., Piliński M., Rutkowski L., *Sieci neuronowe, algorytmy genetyczne i systemy rozmyte*, Wydawnictwo Naukowe PWN, Warszawa-Łódź, 1997.
- [17] Yager R. R., T. C. Robinson, *Linguistic Summaries of Data Bases*, Proc. IEEE Conference on Decision and Control, San Diego, 1981.
- [18] Yager R. R., KFord. M., Canas A. J., *On Linguistic Summaries of Data, w Information Processing and Management of Uncertainty in Knowledge-Based System*, 3rd International Conference, Paris, France, 1990.
- [19] Yager R. R., *On Linguistic Summaries of Data*, w Knowledge Discovery in Databases, Piatetsky-Shapiro, G. & Frawley, B. (eds.), Cambridge, 1991.
- [20] Yager R. R., *Linguistic summaries as a tool for database discovery*, Workshop on Fuzzy Database System and Information Retrival, Yokohama, Japan 1995.
- [21] Zadeh L.A., *Fuzzy Sets*, Information and Control 8, 1965.

Spis rysunków

3.1. Pakiety systemu	4
--------------------------------	---

Spis tabel

2.1. Przykładowa baza danych pacjentów	3
--	---