

KIWIEL



**POLSKA AKADEMIA NAUK**  
**Instytut Badań Systemowych**

# **WSPOMAGANIE DECYZJI**

# **SYSTEMY EKSPERCKIE**

pod redakcją

**Romana Kulikowskiego i Lucyny Bogdan**

Warszawa 1995

# **WSPOMAGANIE DECYZJI**

## **SYSTEMY EKSPERCKIE**

pod redakcją

**Romana Kulikowskiego i Lucyny Bogdan**

Warszawa 1995

Wydano z wykorzystaniem dotacji  
KOMITETU BADAŃ NAUKOWYCH

Materiały konferencji: "Analiza Decyzyjna, Systemy Ekspertyczne, Zastosowania Systemów Komputerowych",  
Warszawa, 25-27 maja 1994r.

Komitet Programowy Konferencji:

Andrzej Ameljańczyk, Zdzisław Bubnicki, Wiesław Grudzewski, Olgierd Hryniewicz, Janusz Kacprzyk, Lech Kruś, Roman Kulikowski (przewodniczący), Kazimierz Mańczak, Ireneusz Nykowski, Zdzisław Pawlak, Roman Słowiński, Andrzej Straszak, Andrzej Weryński, Andrzej Wierzbicki.

Wykonano z oryginałów tekstowych dostarczonych przez autorów

© Instytut Badań Systemowych PAN, Warszawa 1995

ISBN 83-85847-85-5

# ZASTOSOWANIE SYSTEMU EKSPERTOWEGO SCANKEE DO SYMULOWANIA PROCESU ANALIZY CHEMICZNEJ

Barbara Guzowska-Świder  
Katedra Informatyki Chemicznej  
Politechnika Rzeszowska,  
Al. Powstańców Warszawy 6, 35-959 Rzeszów

W artykule przedstawiono bazę wiedzy - wygenerowaną za pomocą systemu SCANKEE - przeznaczoną do symulowania analizy jakościowej związków chemicznych na podstawie danych otrzymanych metodą spektroskopii w podczerwieni.

## WSTĘP

Analizę jakościową związków chemicznych tzn. określanie rodzaju pierwiastków z których jest on zbudowany, a także rozpoznawanie struktury lub fragmentów strukturalnych (podstruktur, grup) wchodzących w skład cząsteczki analizowanego związku chemicznego przeprowadza się stosując najczęściej metody spektralne np. magnetyczny rezonans jądrowy ( $^1\text{H-NMR}$ ,  $^{13}\text{C-NMR}$ ), spektroskopię w podczerwieni (IR), spektrometrię masową (MS) i inn. Metoda spektroskopii w podczerwieni polega na badaniu oddziaływania promieniowania podczerwonego z cząsteczką, która absorbuje (jeśli są spełnione pewne warunki) określoną porcję energii i przenosi się na wzbudzony poziom energetyczny. Mierząc zaabsorbowaną energię w funkcji częstości otrzymuje się wykres charakterystyczny dla danej sub-stancji, zwany widmem IR. Widmo składa się z pasm, które opisuje się za pomocą parametrów spektralnych: położenia pasma (na osi odciętych zazwyczaj w jednostkach częstości tzw. liczbach falowych [ $\text{cm}^{-1}$ ], jest to najważniejszy parametr charakteryzujący pasmo), natężenia pasma (na osi rzędnych w jednostkach np. transmitancji procentowej [T%]) oraz kształtu pasma (może być wyrażony szerokością pasma np. w połowie jego wysokości). Jeżeli założy się, że cząsteczka chemiczna jest zbiorem prawie niezależnych fragmentów strukturalnych, to jej widmo IR można traktować jako kombinację pasm, z których każde jest związane z odpowiednią grupą. Podstawą wykorzystania spektroskopii IR do analizy jakościowej jest fakt, że te same fragmenty strukturalne znajdujące się w różnych cząsteczkach chemicznych absorbują promieniowanie w wąskim zakresie częstości. Np. grupa karbonylowa  $>\text{CO}$  w cząsteczce  $\text{CH}_3\text{-CH}_2\text{-CO-CH}_2\text{CH}_3$  absorbuje przy  $1748\text{ cm}^{-1}$ , w cząsteczce  $\text{CH}_3\text{-CO-CH}_2\text{-CL}$  przy  $1742\text{ cm}^{-1}$ , natomiast w cząsteczce  $(\text{CH}_3)_3\text{-CO-CHCl}_2$  przy  $1750\text{ cm}^{-1}$ .

Na podstawie danych doświadczalnych zestawione zostały częstości charakterystyczne dla różnych fragmentów strukturalnych w tzw. tablice korelacyjne i opublikowane w licznych monografiach i artykułach [1-3]. Wiedza zawarta w tych tablicach jest wykorzystywana przez chemików do rozpoznawania z jakich fragmentów strukturalnych składa się cząsteczka analizowanego związku chemicznego. Wykorzystanie tej wiedzy wymaga jednak dużej umiejętności i doświadczenia, gdyż z jedną grupą związanych jest zazwyczaj kilka pasm, przy czym zakresy charakterystycznych częstości różnych grup nakładają się wzajemnie prowadząc do niejednoznacznej odpowiedzi. Uwzględnienie natężenia i kształtu pasm ułatwia proces identyfikacji struktury. Interpretacja widm w podczerwieni w celu określenia struktury związku chemicznego jest więc złożonym procesem myślowym wymagającym od chemika operowania zarówno danymi liczbowymi (zakresy charakterystycznych częstości grup) jak i danymi opisowymi związanymi z charakterystycznymi kształtami pasm spektralnych.

Opracowano wiele systemów informatycznych [4,5] wspomagających rozpoznawanie struktury związków chemicznych na podstawie danych spektralnych lecz żaden z nich nie spełnia swego zadania w sposób zadowalający. Podejmowane są więc nadal próby zastosowania w systemach informatycznych nowych elementów ułatwiających rozwiązanie tego problemu. Niniejszy artykuł przedstawia symulację procesu rozpoznawania struktury związku chemicznego na podstawie widma w podczerwieni realizowaną za pomocą systemu informatycznego SCANKEE.

## SYSTEM SCANKEE

System SCANKEE [6-8] jest narzędziem softwerowym typu shell do budowy systemów doradczo-decyzyjnych dla nauk przyrodniczych składającym się z luźno związanych modułów. Baza wiedzy (graficzna, tekstowa i regułowa) opisana w niniejszym artykule została wykonana za pomocą modułu budowy grafiki (Graphic Datapath Builder GDB) oraz modułu formalizacji wiedzy (Multimedial Knowledge Formalizer MKF). Wiedza zgromadzona w bazie jest wykorzystywana do rozwiązania problemu za pomocą modułu rozumującego - maszyny wnioskującej (Simulator of Expert Knowledge SEK).

## REGUŁOWO-GRAFICZNA BAZA WIEDZY

Wiedza w bazie jest reprezentowana za pomocą reguł syntaktycznych wyrażających relacje pomiędzy faktami. Słownik bazy zawiera m. inn. słowa kluczowe: JEST, JEŻELI, ORAZ, TO, NIE, INACZEJ, EKTRAN, POKAŻ. Reguły składniowe mają tradycyjną postać:

JEŻELI <warunki> TO <wnioski>

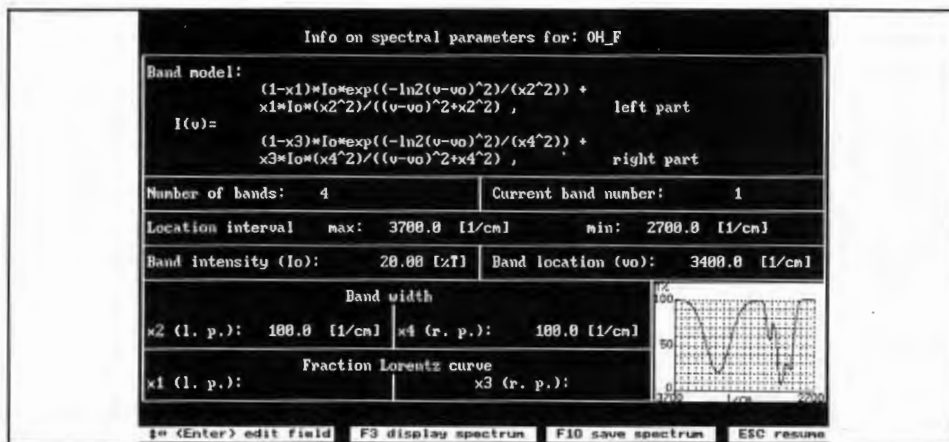
Warunki reguły dotyczą charakterystycznych zakresów częstości położenia pasm, natężenia pasm oraz ich kształtu, natomiast wnioski stanowią hipotezy o obecności bądź braku w cząsteczce identyfikowanych fragmentów strukturalnych lub podejmowane akcje (wyświetlanie komentarzy). Zasady budowy reguł ilustruje przykład reguły służącej do rozpoznania obecności w cząsteczce grupy hydroksylowej -OH związanej z pierścieniem benzenowym tzn. związków chemicznych będących pochodnymi fenoli:

JEŻELI pasmo\_1 JEST obecne ORAZ pasmo\_2 JEST obecne ORAZ pasmo\_3 JEST obecne ORAZ próbka\_1 JEST obecna ORAZ kształt\_1 JEST OH\_F

TO struktura JEST pochodna\_fenolu

Część warunkowa reguły zawarta między słowami JEŻELI... TO ujmuje warunki logiczne które muszą być jednocześnie spełnione, aby mogła zadziałać konkluzyjna część reguły. W części warunkowej reguły występuje pięć warunków:

- warunek "pasmo\_1 JEST obecne" oznacza, że widmo IR badanego związku chemicznego zawiera w zakresie 3120-3000  $\text{cm}^{-1}$  pasmo o natężeniu  $T\% < 70\%$  oraz w zakresie 900-600  $\text{cm}^{-1}$  pasmo o natężeniu  $T\% < 30\%$ ,
- warunek "pasmo\_2 JEST obecne" oznacza, że widmo IR zawiera w zakresie 1640-1580  $\text{cm}^{-1}$  lub/oraz 1520-1480  $\text{cm}^{-1}$  pasmo (pasma) o natężeniu  $T\% < 50\%$ ,
- warunek "pasmo\_3 JEST obecne" oznacza, że widmo IR zawiera w obszarze 1370-1300  $\text{cm}^{-1}$  pasmo o natężeniu  $T\% < 40\%$  oraz w obszarze 1250-1170  $\text{cm}^{-1}$  pasmo o natężeniu  $T\% < 20\%$ ,
- warunek "próbka\_1 JEST obecna" oznacza, że widmo zostało zarejestrowane dla związku w stanie ciekłym lub roztworze,
- warunek "kształt\_1 JEST OH\_F" oznacza, że widmo IR zawiera pasma w przedziale 3700-27000  $\text{cm}^{-1}$ , o kształcie charakterystycznym dla grupy hydroksylowej -OH związanej z pierścieniem benzenowym, których obraz jest zapamiętany na rysunku o nazwie OH\_F (Rys.1,2).

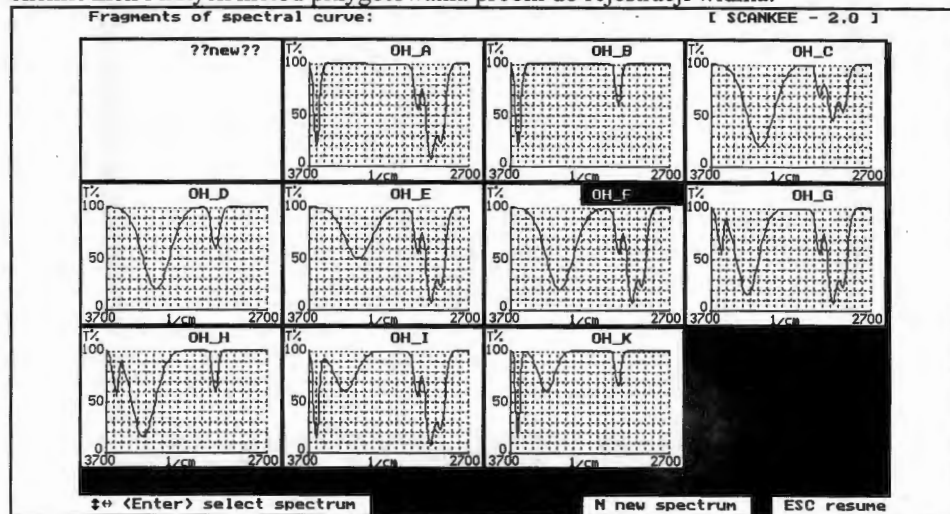


Rys.1. Ekran edycji obrazów krzywej spektralnej.

Jeżeli powyższe warunki są spełnione, system nadaje konkluzji "struktura" wartość "pochodna\_fenolu". Nadanie wartości atrybutom o nazwie "pasmo\_?" nie wymaga podpowiedzi użytkownika, gdyż jest realizowane automatycznie na podstawie danych spektralnych (wartości położenia pasm i ich natężeń) wprowadzonych do komputera. Natomiast atrybutom związanym z kształtem pasm przypisywana jest odpowiednia wartość na podstawie podpowiedzi użytkownika, którego zadaniem jest wybranie spośród

proponowanych przez system obrazów fragmentów krzywej spektralnej takiego, który jest najbardziej zbliżony do eksperymentalnego. Zastosowanie w systemie możliwości wizualizacji zarówno kształtu jednego pasma (szerokie, wąskie, rozmyte, ostre, z przegięciem itp.) jak i sekwencji pasm (kilku kolejnych pasm, często nakładających się wzajemnie, o różnych natężeniach) jest bardzo dogodną formą komunikowania się systemu z użytkownikiem. Dodatkową zaletą tego rozwiązania jest uwolnienie eksperta, twórcy regułowej bazy wiedzy, od opracowywania długich tekstów drobiazgowo i zazwyczaj w sposób niezodawalający opisujących kształt pasm krzywej spektralnej. System SCANKEE umożliwia autorowi bazy wiedzy zastąpienie tych słownych opisów charakterystycznych pasm spektralnych, przez ich obrazy, które opracowuje się za pomocą modułu GDB.

Edycję rysunku pasm spektralnych (Rys.1.) - podczas przygotowywania bazy wiedzy - rozpoczyna się od podania wartości przedziału częstości występowania pasm dla danego elementu strukturalnego oraz liczby pasm w przedziale. Następnie tak dobiera się wartości natężeń pasm oraz wartości współczynników określających udział funkcji Gaussa i Lorentza w sumarycznej funkcji opisującej kształt danego pasma, aby kształt wysymulowanej krzywej był identyczny z obrazem widma podawanym dla tego elementu w literaturze lub znanym ekspertowi z doświadczenia. Na Rys.1. przedstawione są również pasma spektralne w zakresie  $3700-2700\text{ cm}^{-1}$  charakterystyczne dla grupy hydroksylowej związanej z pierścieniem benzenowym w cząsteczce zawierającej również fragmenty alkilowe, widma zarejestrowanego dla substancji w postaci ciekłej. Natomiast Rys.2. przedstawia fragmenty widma IR, w analogicznym zakresie liczb falowych, grupy hydroksylowej OH związanej z pierścieniem benzenowym dla innych klas związków chemicznych i innych metod przygotowania próbki do rejestracji widma.



Rys.2.

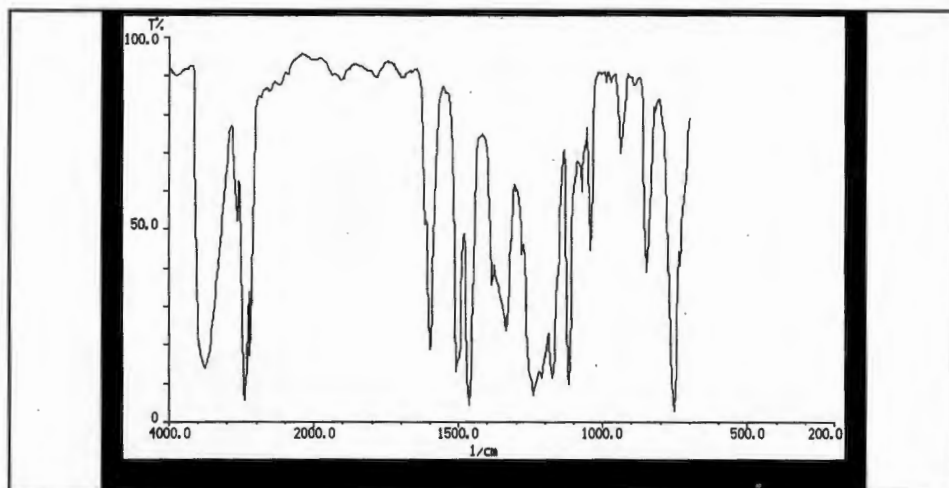
Obrazy fragmentów widma IR dla rodziny fragmentów strukturalnych - grup hydroksylowych -OH wchodzących w skład cząsteczek różnych klas związków chemicznych.

## SYMULACJA PROCESU ANALIZY CHEMICZNEJ

W celu rozpoznania struktury związku chemicznego na podstawie widma IR należy wyznaczyć parametry spektralne pasm tzn. ich położenie i natężenie. Jeśli wykonuje się analizę bez pomocy systemu komputerowego wówczas, korzystając z tablic korelacyjnych oraz opierając się na własnym doświadczeniu (szczególnie w odniesieniu do kształtu pasm), przeprowadza się proces myślowy, aby określić które fragmenty strukturalne mogą wchodzić w skład cząsteczki badanego związku. Natomiast, dysponując systemem SCANKEE oraz odpowiednią bazą wiedzy - uprzednio przygotowaną przez eksperta - wprowadza się do komputera parametry pasm spektralnych, a następnie w procesie konsultacji podejmuje decyzje odnośnie wyboru obrazu pasm krzywych wzorcowych najbardziej podobnego do badanego fragmentu widma analizowanej substancji.

Załóżmy, że nie znamy struktury związku chemicznego, którego widmo (Rys.3.) składa się z pasm (pierwsza z pary liczb oznacza położenie pasma [ $\text{cm}^{-1}$ ], liczba w nawiasie przedstawia natężenie pasma [T%]):

|           |           |           |           |           |
|-----------|-----------|-----------|-----------|-----------|
| 3520(14), | 3082(52), | 2986(5),  | 2876(17), | 1601(19), |
| 1509(13), | 1465(5),  | 1333(24), | 1241(7),  | 1213(12), |
| 1180(12), | 1121(10), | 1074(60), | 1043(45), | 941(70),  |
| 850(39),  | 758(3),   | 730(40).  |           |           |



Rys.3. Widmo IR analizowanej substancji.

Na podstawie tych danych, wprowadzonych do komputera, system sam nadaje wartości atrybutom "pasma\_1", "pasma\_2", "pasma\_3", natomiast w momencie, gdy próbuje ustalić wartość atrybutu "próbka\_1", na ekranie wyświetlana jest lista możliwych odpowiedzi z



których użytkownik wybiera prawidłową. W celu nadania wartości atrybutowi "kształt\_1" system wyświetla zestaw charakterystycznych - dla danej rodziny podstruktur - obrazów wzorcowych spośród których użytkownik powinien dokonać wyboru. W omawianym przypadku na ekranie wyświetlane są fragmenty widma IR w zakresie  $3700-27000\text{ cm}^{-1}$  dla grup hydroksylowych związanych z pierścieniem benzenowym w związkach należących do różnych klas. Dla badanej substancji do widma eksperymentalnego najbardziej podobny jest rysunek o nazwie OH\_F. Wybór tego obrazu powoduje odpalenie odpowiedniej reguły, a tym samym rozpoznanie, że cząsteczka badanej substancji zawiera grupę -OH połączoną z pierścieniem benzenowym. W analogiczny sposób rozpoznawane są inne fragmenty strukturalne uwzględnione w bazie wiedzy.

System SCANKEE jest dogodnym narzędziem do automatycznego tworzenia reguł bazy wiedzy przeznaczonej do analizy jakościowej związku chemicznego na podstawie jego widma otrzymanego metodą spektroskopii w podczerwieni, a także innymi metodami spektralnymi np. magnetycznego rezonansu jądrowego, nadfioletu, Ramana czy spektrometrii masowej.

## LITERATURA

- [1] Rao C.N.R., 'Chemical applications of infrared spectroscopy', Academic Press, New York and London, 1963
- [2] Szafran M., Dega-Szafran Z., 'Określanie struktury związków organicznych metodami spektroskopowymi', PWN, Warszawa, 1988
- [3] Wojtkowiak B., Chabanel M., 'Spektrochemia molekularna', PWN, Warszawa, 1984
- [4] Gray N.A.B., 'Computer assisted structure elucidation', John Wiley and Sons, New York, 1986
- [5] Fessenden R.J., Györgyi L., J. Chem. Soc. Perkin Trans. 2 1755, 1991
- [6] Hippe Z.S., Dębska B. and Mazur M., 'Knowledge Engineering System Environment Based on New Concepts', Proc. Intern. Conference System-Modelling-Control, Zakopane 1, 10, 1993
- [7] Hippe Z.S., Dębska B. and Mazur M., 'AI-tool for automatic design of expert systems for engineering sciences', Proc. 9th Intern. Conference on System Engineering, Las Vegas 1, 268, 1993
- [8] Dębska B., "Metodologia kreowania graficzno-regułowych baz wiedzy dla systemu ekspertowego SCANKEE", Mat. konf. Analiza Decyzyjna, Systemy Eksperckie, Zastosowania Systemów Komputerowych, Warszawa, 1994

**ISBN 83-85847-85-5**

---

**W celu uzyskania bliższych informacji i zakupu dodatkowych egzemplarzy  
prosimy o kontakt  
z Instytutem Badań Systemowych PAN  
ul. Newelska 6, 01-447 Warszawa  
tel. 36-19-01 w. 241 e-mail: [kotuszew@ibspan.waw.pl](mailto:kotuszew@ibspan.waw.pl)**