

Wielomodułowy system  
rozpoznawania tekstów w notacji  
muzycznej

Marcin Luckner

[mluckner@gik.pw.edu.pl](mailto:mluckner@gik.pw.edu.pl)

# *Hierarchical OCR System For Text In Musical Scores*

- Homenda, Luckner
- Fuzzy Logic, Soft Computing and Computational Intelligence
- Eleventh International Fuzzy Systems Association World Congress
- Beijing, China

From the 20th Century-Fox Film "BUTCH CASSIDY AND THE SUNDANCE KID"

# RAINDROPS KEEP FALLING ON MY HEAD

Arranged by  
JOHN BRIMHALL

Words by HAL DAVID  
Music by BURT BACHARACH

Moderately (♩. ♩ = ♩<sup>3</sup> ♩)

F 4  
2  
Rain - drops keep fall - in' on my head, and just like the guy whose feet are  
did me some talk - in' to the sun, and I said I did - n't like the  
Rain - drops keep fall - in' on my head, but that does - n't mean my eyes will

*p*

Fmaj7 1  
F7 4  
1

5

Raindrops Keep Falling on My Head - 2 - 1

© 1969 BLUE SEAS MUSIC, INC., JAC MUSIC CO., INC. and WB MUSIC CORP.  
This Arrangement © 1997 BLUE SEAS MUSIC, INC., JAC MUSIC CO., INC. and WB MUSIC CORP.  
All Rights Reserved

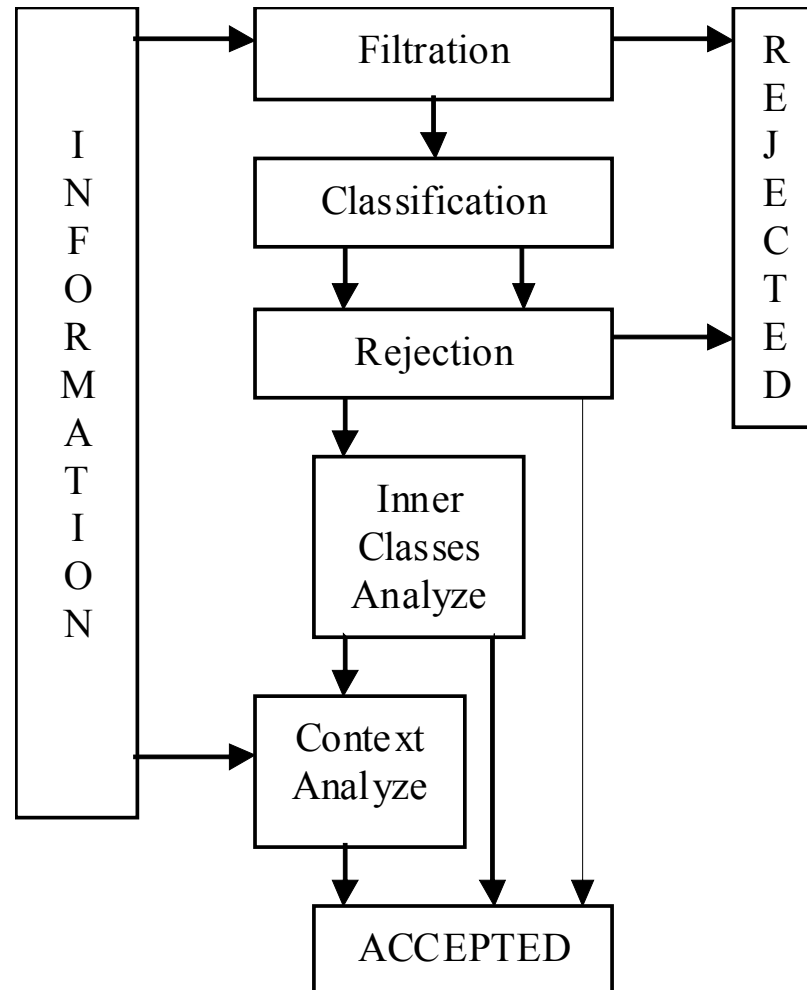
# Zbiór badawczy

- 18,381 znaków
- 62 klasy  $2*26 + 10$  (A, a, 0)
- >100 na literę (wyjątki O, Q, X, Z)
- Times New Roman, Arial, Courier, Verdana
- kursywa, pogrubienie, szeryfowe
- **Angielski**, Francuski, Niemiecki, Włoski, Łacina

# Cele systemu

- Wysoka skuteczność klasyfikacji ( $>90\%$ )
- Odrzucanie dużej ilości znaków spoza zbioru rozpoznawanych
- Szybkość działania

# Schemat systemu



# Filtracja

- Odrzucanie “oczywistych” przypadków
- Istotna szybkość podejmowania decyzji
- Prosta (brak dodatkowej wiedzy)
- Kontekstowa (wymaga dodatkowej wiedzy)

# Filtracja prosta

- Bardzo małe obiekty (5x5 pikseli)
  - brak górnej granicy
- Znacznie szersze niż wyższe
  - nie odwrotnie: ‘l’, ‘I’, ‘i’
- Dominacja koloru tła



# Filtracja kontekstowa

The image shows a musical notation system with two staves. The top staff is a treble clef staff with a colon indicating a repeat sign. The bottom staff is a bass clef staff. The lyrics are written below the staves. The chord symbol 'Fmaj7' is written above the first staff. A circled '1' is written above the second staff. The lyrics are: 'head, sun, head,' under the first staff and 'and and but' under the second staff.

- Znajomość struktury partytury
- Wykluczenie niektórych obszarów z procesu rozpoznania
- Liryka ograniczona pięciolinia
- Brak kryteriów dla nagłówków

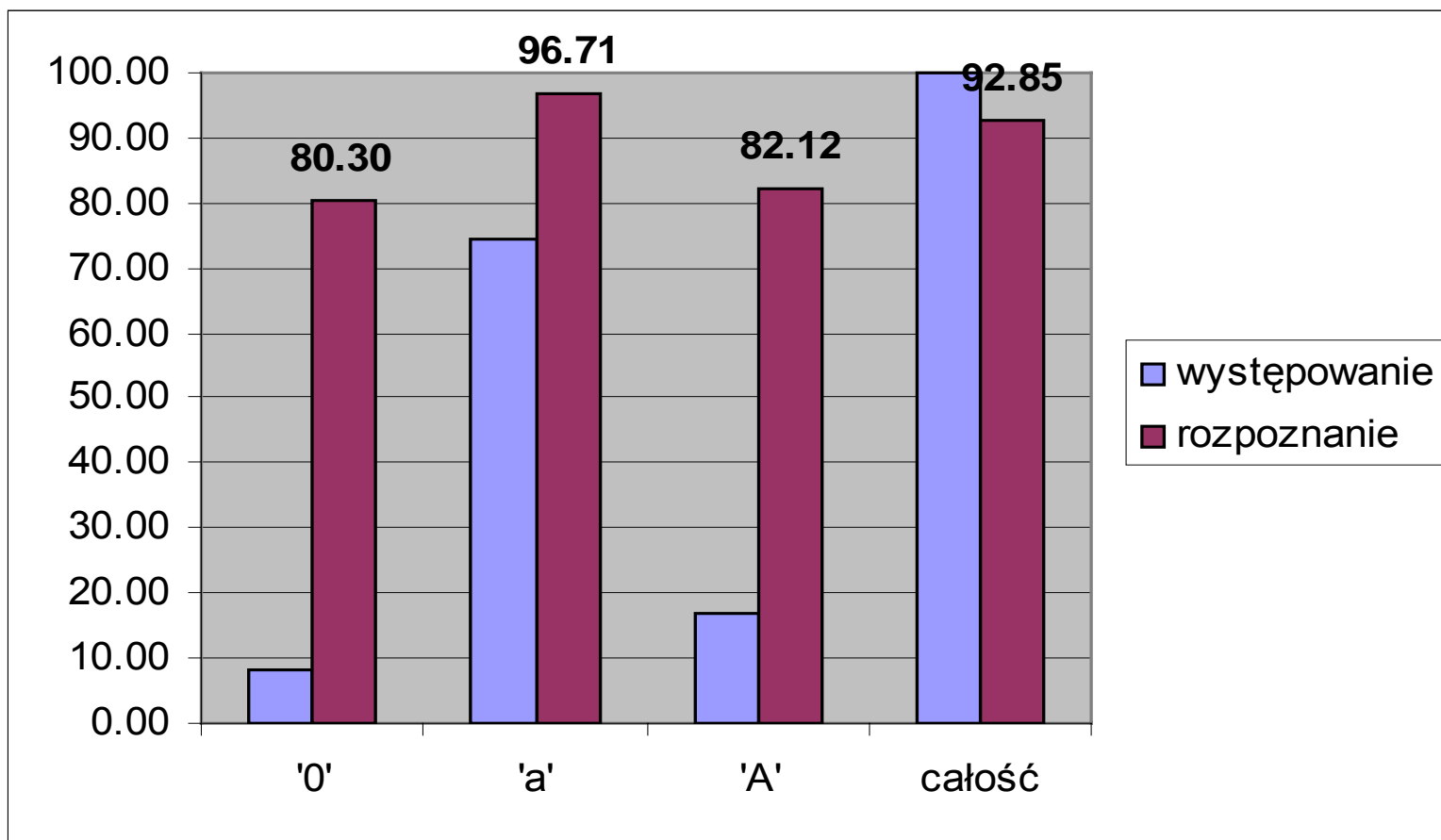
# Klasyfikacja

- Liniowa sieć neuronowa
- Liniowa szybkość klasyfikacji
- Przyzwoite wyniki
  
- Symbole opisano 183 liczbami rzeczywistymi z przedziału  $[0,1]$
- Znaczący czas wyliczenia cech

# Wyniki klasyfikacji

Character	Count	Result [%]	Character	Count	Result [%]	Character	Count	Result [%]
0	102	15.7	a	984	98.5	A	265	99.2
1	142	43.7	b	254	68.1	B	155	94.2
2	105	94.3	c	161	63.4	C	115	66.1
3	254	96.9	d	539	100	D	235	92.3
4	157	99.4	e	1524	99.5	E	196	99
5	152	63.2	f	271	86.7	F	149	98.7
6	101	71.3	g	332	100	G	146	99.3
7	329	99.7	h	542	97.2	H	110	93.6
8	104	77.9	i	787	96.7	I	107	57.9
9	105	84.8	j	117	43.6	J	103	85.4
			k	149	96.6	K	94	98.9
			l	455	94.7	L	138	93.5
			m	728	99.7	M	106	98.1
			n	738	98.9	N	127	99.2
			o	846	99.3	O	5	0
			p	417	98.8	P	106	30.2
			q	117	97.4	Q	42	61.9
			r	826	99.5	R	111	89.2
			s	953	98.6	S	120	38.3
			t	1210	98.8	T	144	86.8
			u	577	97.9	U	155	82.6
			v	195	83.1	V	108	59.3
			w	333	99.7	W	112	18.8
			x	110	94.5	X	20	0
			y	433	98.6	Y	132	94.7
			z	116	98.3	Z	15	0

# Podsumowanie klasyfikacji



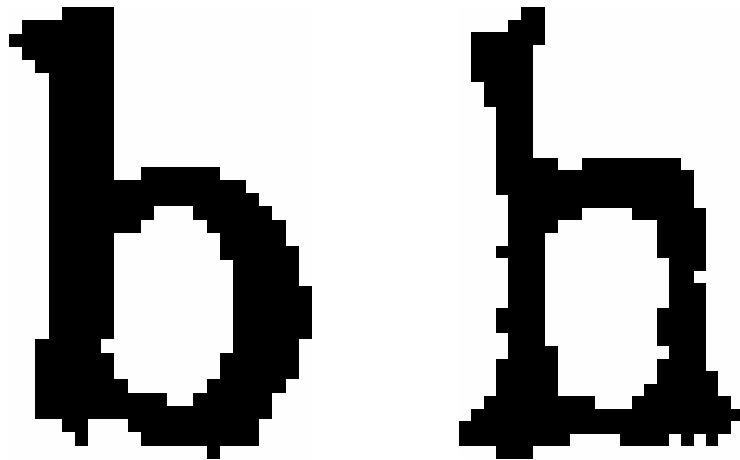
# Odrzucanie

- Znaczna liczba symboli spoza zbioru rozpoznawanych
- Odrzucane jezeli istotnie przekraczaja zakresy cech znane ze zbioru uczacego dla danej klasy
- W przypadku odrzucenia, brane pod uwage drugie przypisanie

# Ponowna analiza symbolu

- Tylko dla klas o dużym ryzyku błędu
- Oparta na znajomości skuteczności procesu klasyfikacji (macierz błędnych klasyfikacji)
- Dodatkowo na analizie kontekstowej

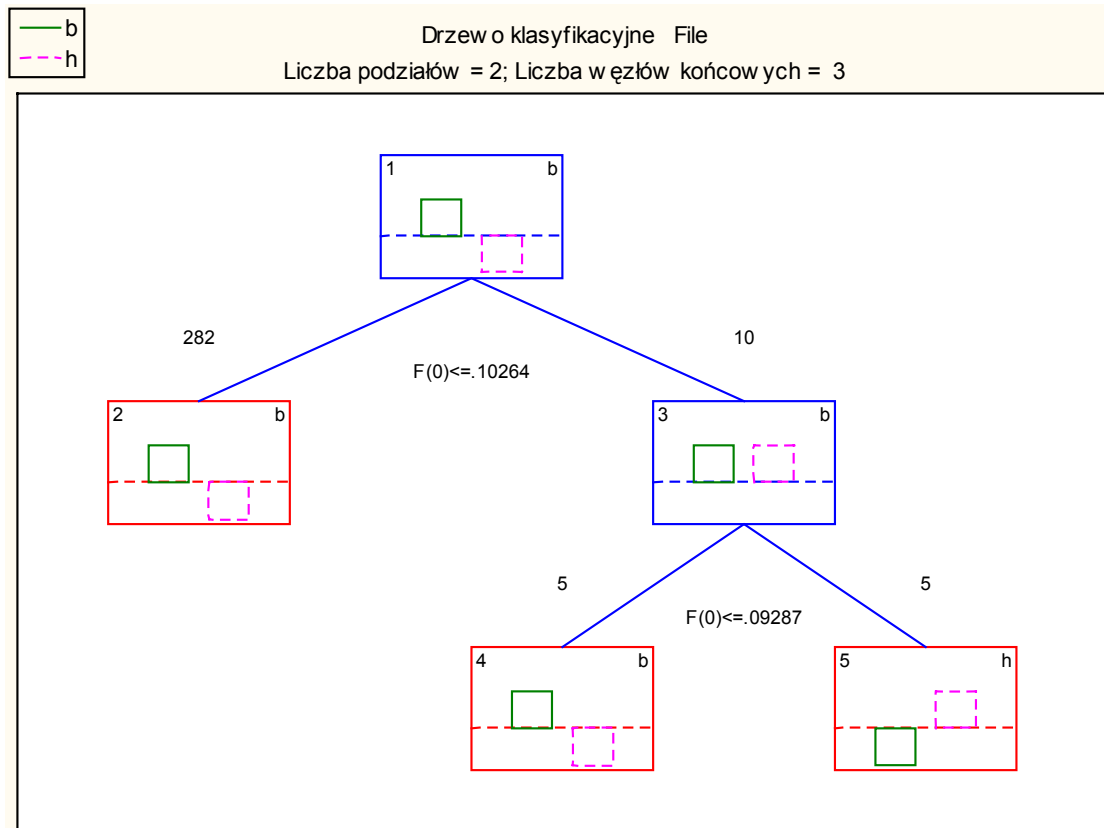
# Analiza błędów wewnątrz klas



- Tylko dla klas o dużym ryzyku błędu
- Błąd kumulowany w 1-2 klasach
- 28% 'b' rozpoznane jako 'h'

b	254	<b>68.1</b>
h	542	<b>97.2</b>

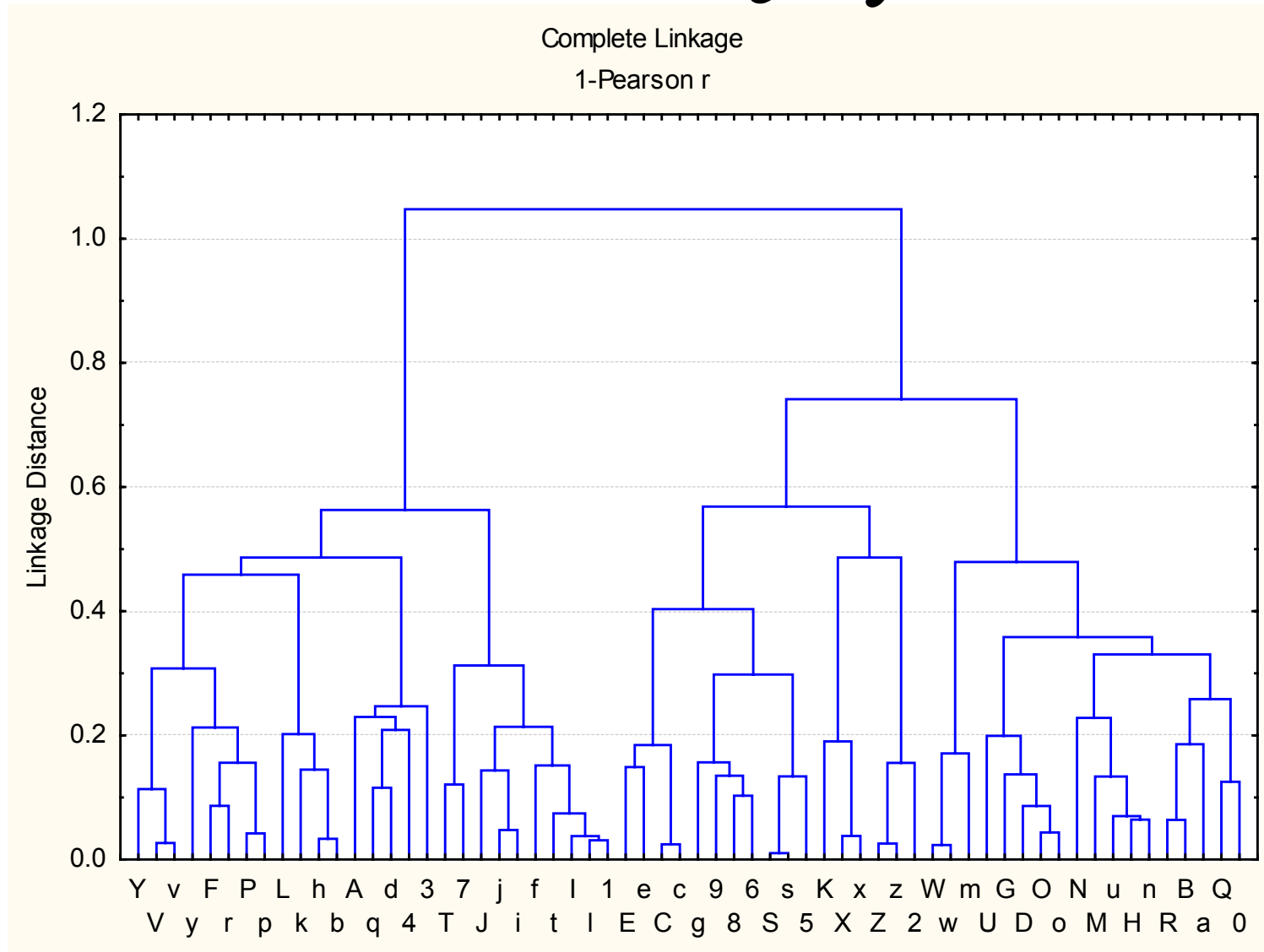
# Analiza klasy



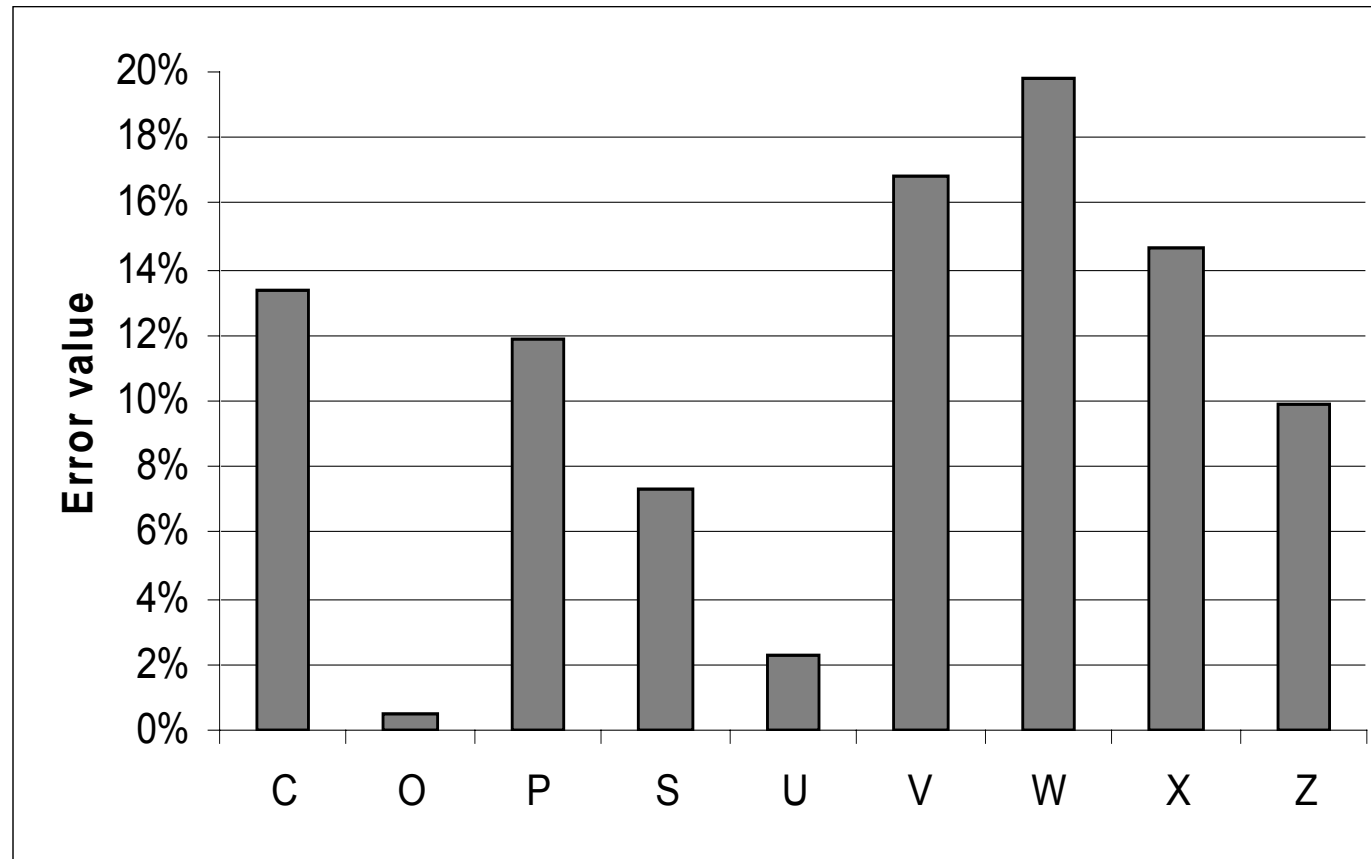
- Na podstawie liniowych reguł decyzyjnych
- maksymalnie kilkanaście cech
- układ drzewiasty
- znaczna redukcja błędów (<1%)



# Podobieństwa między klasami



# Wielkość liter



c	161	63.4	C	115	66.1
o	846	99.3	O	5	0
p	417	98.8	P	106	30.2
s	953	98.6	S	120	38.3
u	577	97.9	U	155	82.6
v	195	83.1	V	108	59.3
w	333	99.7	W	112	18.8
x	110	94.5	X	20	0
z	116	98.3	Z	15	0

# Analiza kontekstowa

- Położenie znaku
  - Odtworzenie wersów
  - Połączenie znaków (sylaby, wyrazy)
- Pozwala na ujednoczenie w procesie klasyfikacji znaków C, c; O, o; P, p; S, s; U, u; V, v; W, w; X, x; Z, z
- Redukcja wymiaru klasyfikatora do 53 klas

# Analiza wersu

- Budowa wersu na podstawie położenia liter
- Estymacja wysokości niskich i wysokich liter
  - metoda k-średnich ( $k=2$ )
  - Zamiana małych liter na wielkie i *vice versa*
- Określenie linii bazowej
  - Analiza położenia i rozróżnienie liter ‘p’ i ‘P’

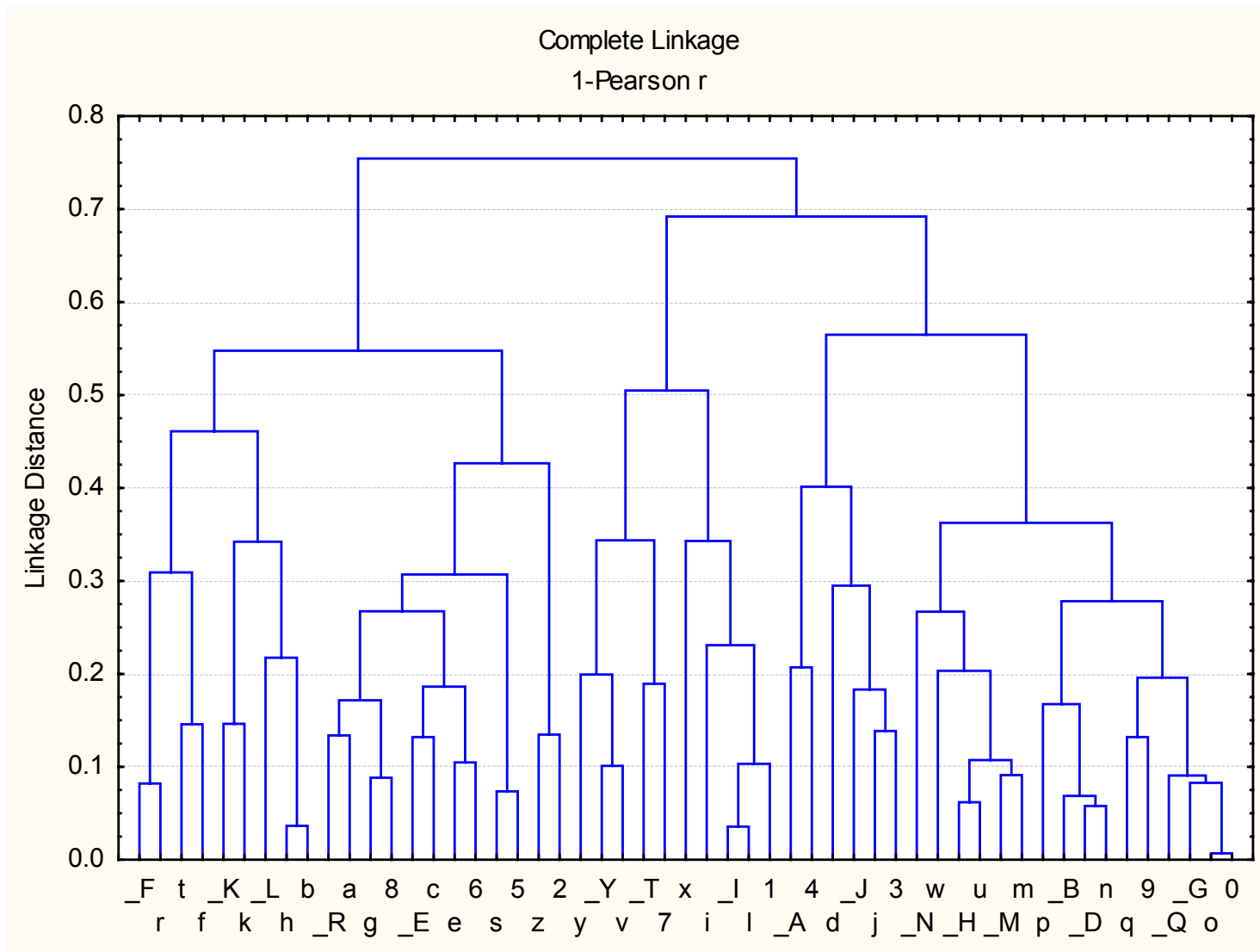
# Podział na słowa

- Wątpliwy dla analizy małe-wielkie litery
- Ułatwia analizę litery-cyfry (0, O; 1, 1)
- Wzmacnia wyznaczanie linii bazowej
- Tworzony przez analizę odstępów w wersji
  - k-mean (k=2) z modyfikacjami.
- Możliwość wsparcia słownikowego

# Nowe testy

- 30637 - przypadków w 53 klasach
- Klasyfikacja 94.20%
- Odrzucanie 96.10%
- Analiza klas 96.90%

# Podobieństwo symboli



# Pozostałe problemy

- Lokalizacja
  - ‘j’ ‘i’ jako całość
  - akcenty

- Podział ciągów
  - dwuliterowych
  - wieloznakowych

**fl oo ll**

**gra voce Arrang**