

Podstawy bioinformatyki

Dr Jan Paweł Jastrzębski

Katedra Fizjologii i Biotechnologii Roślin
Pok. 113 CB

www.uwm.edu.pl/bioinfo
jan.jastrzebski@uwm.edu.pl
bioinformatyka@gmail.com

wykłady

Nieobowiązkowe (własne materiały)
Wstęp do drugiej części bioinformatyki

ćwiczenia

Frekwencja 100% (zal)
Definicje bioinformatyczne (pkt)
Raport z ćwiczeń/projekt (pkt)
Kolokwium (ocena)

Plan wykładów

1 - Komputer, internet

informacje organizacyjne, definicja i jak działa komputer? co to jest internet? Historia internetu. Bazy danych - po co to jest i jak to wygląda? HTML

2 - bioinformatyka

definicje bioinformatyki, dziedziny i zakres bioinformatyki (-omiki oraz zagadnienia badawcze), pojęcie "in silico", historia bioinformatyki; Human Genom Project; bioinformatyczne bazy danych; projekty bioinformatyczne; niezbędne definicje (słownik).

3 - NCBI, EBI

przeszukiwanie biologicznych baz danych; wyszukiwarki; BLAST

4 - alignment (pairwise + MSA) + macierze substytucji

5 - analizy in silico

biologia eksperymentalna - biologia teoretyczna; właściwości fizyczne i chemiczne biomolekuł; analizy teoretyczne; hydrofobowość, skala hydrofobowości, profil hydrofobowości.

Plan ćwiczeń

- 1 - obsługa komputera
- 2 - podstawy HTML
- 3 - NCBI, EBI (modele danych)
- 4 - przeszukiwanie biologicznych baz danych - słowa kluczowe (PubMed, proteins)
- 5 - przeszukiwanie biologicznych baz danych - BLAST
- 6 - analizy in silico - alignment
- 7 - analizy in silico - właściwości fizykochemiczne

materiały

- **Podręczniki**
- **Internet** (serwisy bioinformatyczne, biotechnologiczne, biologiczne i medyczne bazy danych, narzędzia bioinformatyczne)
- **www.uwm.edu.pl/wbiol**
- **Wykłady + własne notatki**
- **Skrypty** (moje, studentów i udostępnione)
- **Multimedia** i materiały online

podręcznik

podręcznik

Bioinformatyka
Podręcznik do analizy genów i białek
pod redakcją
A.D. Basavanasia i B.F.F. Ouellette'a
WYDAWNICTWO NAUKOWE PWN

Bioinformatyka i ewolucja molekularna
Paul G. Higgins, Tom

Bioinformatics FOR DUMMIES
"An excellent book for beginners and occasional practitioners"
Updated to cover multiple new genomes and databases
2nd Edition
Jean Michel Claverie, PhD
Bioinformatics Department
Cedric Notredame, PhD
Department of Biology, University of

A Reference for the Rest of Us!
Jean Michel Claverie, PhD
Bioinformatics Department
Cedric Notredame, PhD
Department of Biology, University of

Biochemia
Lewin W. Rog
Jan L. Tymoczko
Leland Haynes

KfIBR, Wydział Biologii,
UWM w Olsztynie

dr Jan Paweł Jastrzębski

Podstawy Bioinformatyki
7

Rysować skale czasowe do multimediów!

KfIBR, Wydział Biologii,
UWM w Olsztynie

dr Jan Paweł Jastrzębski

Podstawy Bioinformatyki
8

multimedia

- History of Computers (od liczydeł) – ogólnie 10 minut
- BBC documentary history of computers (BBC) – 9 części
- Computer Pioneers (The Computer History Museum) – XX wiek

KfIBR, Wydział Biologii,
UWM w Olsztynie

dr Jan Paweł Jastrzębski

Podstawy Bioinformatyki
9

Materiały multimedialne

How Computers Work - Journey Into The Walk-Through Computer (dla dzieci)

Prawdziwa Historia Internetu:
- Bitwa Przeglądarek
- Wojna Wyszukiwarek

Polskie komputery – stracona szansa (PR 1)

KfIBR, Wydział Biologii,
UWM w Olsztynie

dr Jan Paweł Jastrzębski

Podstawy Bioinformatyki
10

do czytania

Bioinformatyka:

- rozdział ... (komputer)
- rozdział ... (internet)

Urządzenia techniki komputerowej część 1. Jak działa komputer (Krzysztof Wojtuszkiewicz, PWN)

Historia internetu, Domeny (Grzegorz Hodak, wykłady, Uniwersytet Wrocławski)

<http://www.mapainternetowa.com/> (bardzo ogólnie)

KfIBR, Wydział Biologii,
UWM w Olsztynie

dr Jan Paweł Jastrzębski

Podstawy Bioinformatyki
11

Definicja bioinformatyki

dyscyplina nauk biologicznych wywodząca się z biotechnologii (genetyki), zajmująca się stosowaniem narzędzi matematycznych i informatycznych do rozwiązywania problemów biologii (głównie biologii molekularnej) i zagadnień biotechnologicznych. Podstawowymi poddziedzinami bioinformatyki są: genomika, proteomika, transkryptomika i metabolomika.

in vivo – badania przyżyciowe; mało możliwości manipulacji
– w tkance; ograniczone możliwości manipulacji

in situ – w szkle; największe „naturalne” możliwości manipulacji

in vitro – w krzemiu; możliwość analizowania wszelkich, nawet pozornie niemożliwych układów

in silico

KfIBR, Wydział Biologii,
UWM w Olsztynie

dr Jan Paweł Jastrzębski

Podstawy Bioinformatyki
12

Narzędzia matematyczne i informatyczne

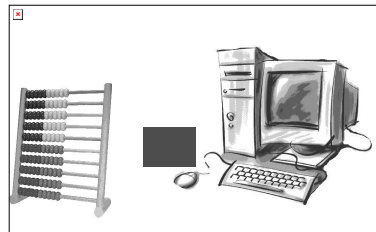
?

(podać przykłady – wektor, chromosom, tablica, macierz, właściwości białek)

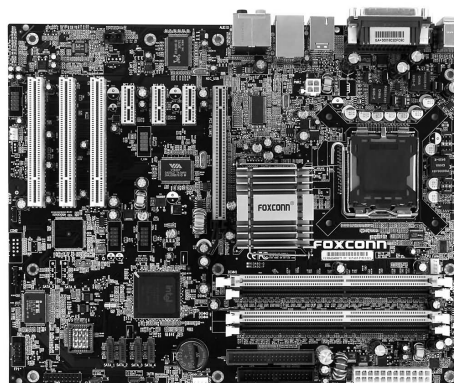
hardware

Komputer

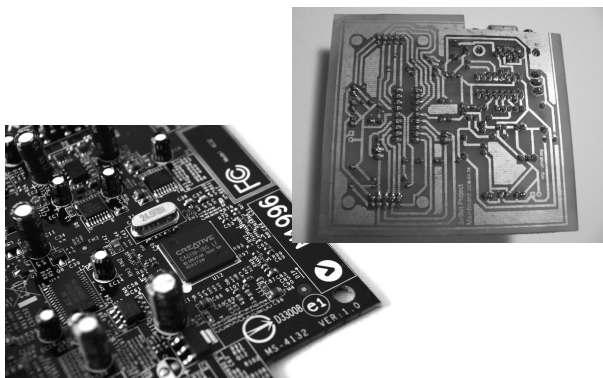
urządzenie elektroniczne służące do przetwarzania wszelkich informacji, które da się zapisać w formie ciągu cyfr, albo sygnału ciągłego.



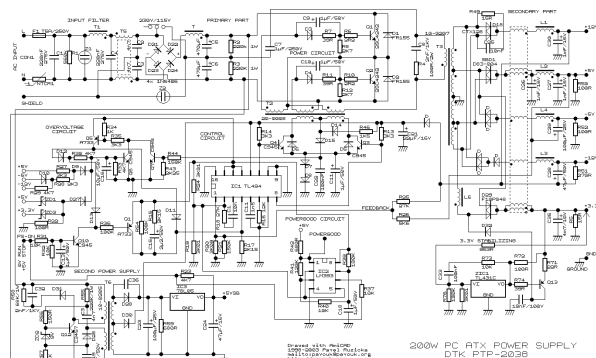
Płyta główna



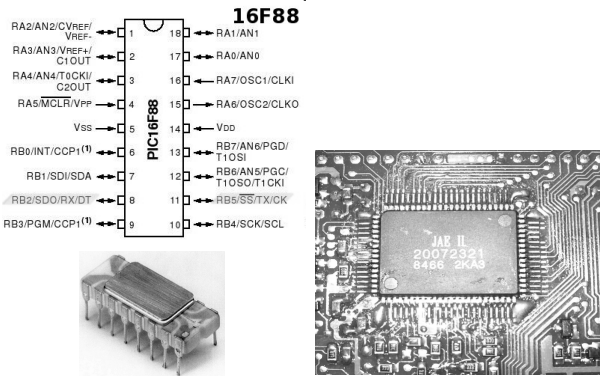
Układy elektroniczne / ścieżki



schemat



układ scalony / procesor

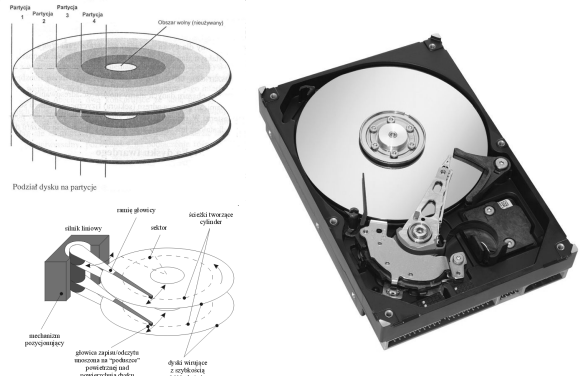


KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
19

Twardy dysk HDD

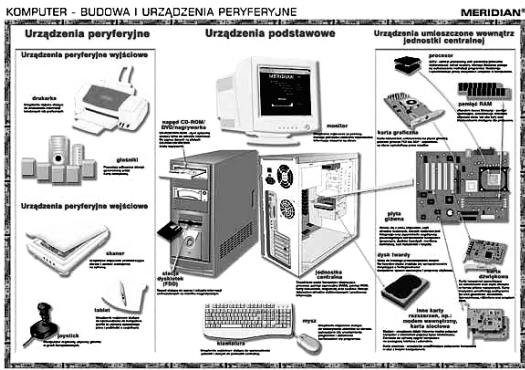


KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
20

Urządzenia peryferyjne



KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
21

System binarny

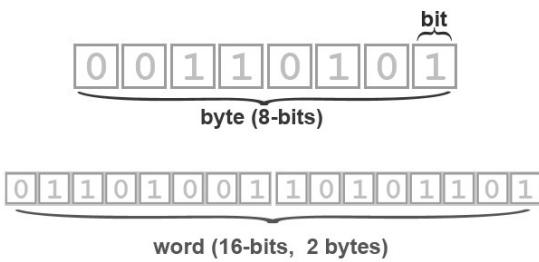
?

KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
22

sygnał, bit



KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

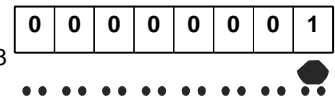
dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
23

Bit i bajt – ilość zajmowanej pamięci

1 bajt = 8 bitów

256 kombinacji $\rightarrow 2^8$



W ósmiobitowym systemie istnieje możliwość zapisu 256 różnych znaków, symboli, odcieni w jednej pozycji pamięci (np. w jednej zmiennej) np.:

256 odcieni koloru czerwonego (Red)
256 odcieni koloru zielonego (Green)
256 odcieni koloru niebieskiego (Blue)
RGB <255,0,255>

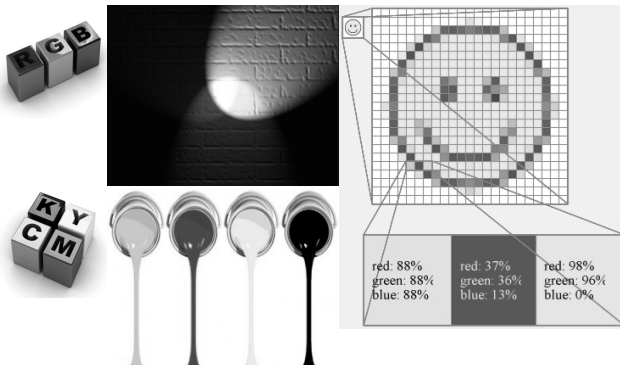


KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
24

Obraz na monitorze, kolory pikseli



KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

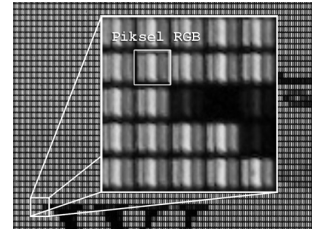
dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
25

Piksel

Piksel (ang. pixel = picture+element) jest to najmniejszy element obrazu bitmapowego. Jeden piksel to bardzo mały kwadrat o przeciętnym boku 0,28mm (rzadziej: prostokąt) widzialny z odległości użytkowej jako wypełniony jednolitym kolorem. Piksel stanowi także najmniejszy element obrazu wyświetlanego na monitorze komputera. Tryb pracy monitora, a konkretnie jego rozdzielczość to właśnie liczba pikseli jakie zawiera on w pionie i poziomie.

Wikipedia

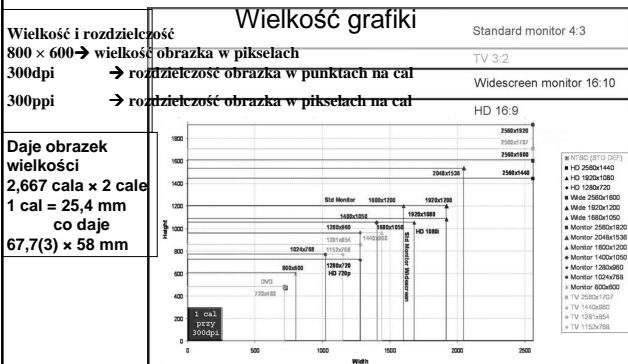


KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
26

Rozdzielczość / dpi / ppi



KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

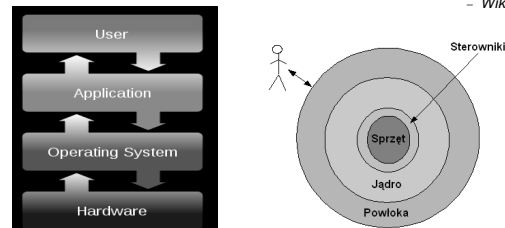
dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
27

SYSTEM OPERACYJNY NARZĘDZIA INFORMATYCZNE

- **System operacyjny** (ang. skrót **OS Operating System**) – oprogramowanie zarządzające sprzętem komputerowym, tworzące środowisko do uruchamiania i kontroli zadań użytkownika.

Wikipedia

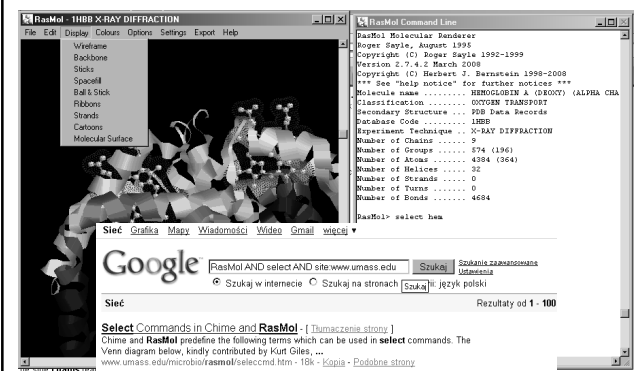


KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
28

Powłoka, INTERFACE, KOMENDY, OPERATORY, SKRYPTY



KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
29

Plik i katalog

Praca domowa

KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
30

Modele danych i ASN.1

Model danych jest to abstrakcyjny model (pojęcie/schemat) opisujący jak dane są reprezentowane i jak mają być używane.

Pojęcie MODEL DANYCH generalnie ma dwa znaczenia:

- A data model *theory* (teoretyczny) i.e. a formal description of how data may be structured and used.
- A data model *instance* (konkretny) i.e. applying a data model *theory* to create a practical data model *instance* for some particular application.

Model bazy danych

zbiór zasad, którymi należy się posługiwać podczas tworzenia bazy danych. W modelu danych określa się reguły, zgodnie z którymi dane umieszcza się w strukturach. Określone są również dozwolone operacje. Definiuje się strukturę danych poprzez specyfikację reprezentacji dozwolonych w modelu obiektów (encji) oraz ich związków. W informatyce głównymi modelami baz danych są:

- hierarchiczny model danych,
- relacyjny model danych,
- grafowy (sieciowy) model danych,
- obiektowy model danych,
- sieci semantyczne,

Wikipedia

ASN.1 (*Abstract Syntax Notation 1*) jest standardem ISO a nie modelem danych

KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
31

ASN.1 *Abstract Syntax Notation One*

- **ASN.1** (skrót od *Abstract Syntax Notation One* - abstrakcyjna notacja składniowa numer jeden) jest to standard służący do opisu struktur przeznaczonych do reprezentacji, kodowania, transmisji i dekodowania danych.

Dostarcza zbiór formalnych zasad pozwalających na opis struktur obiektów w sposób niezależny od konkretnych rozwiązań sprzętowych.

- Jest to standard ITU-T/ISO, po raz pierwszy został opisany w roku 1984 jako część dokumentu CCITT X.409/84. Następnie w 1988 wydano go jako samodzielny dokument ITU-T X.208. W roku 1994 wydano jego nową wersję w dokumentach ITU-T z serii X.680 (X.680-X.683). W roku 2002 wycofano dokument ITU-T X.208.

- Standard ASN.1 określa jedynie składnię abstrakcyjną informacji, nie określa natomiast sposobu jej kodowania w pliku. Metody kodowania informacji podanych w składni ASN.1 zostały opisane w kolejnych standardach ITU-T/ISO.

Wikipedia

KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
32

FORMATY PLIKÓW / ROZSZERZENIA NAZW PLIKÓW

Format pliku w informatyce to ustalony standard zapisu informacji w pliku danego typu.

Dysk:\katalog1\katalog2\sciezka_dostepu\nazwa_pliku.roz

c:\Program Files\RasMol\raswin.exe

<http://www.uwm.edu.pl/katedrafbr/index.php>

nazwa_pliku.rozszerzenie

KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
33

Formaty graficzne

Formaty grafiki rastrowej

BMP	-	bez kompresji
TIF/TIFF	-	kompresja bezstratna
GIF	-	kompresja z wyborem ilości kolorów, przezroczystości, animacje
PNG	-	kompresja bezstratna – miał wyprzeć GIF-a
JPG/JPEG-	-	kompresja stratna (nieodwracalnie)
DjVu	-	kompresja do 10x lepsza od JPEG

Formaty grafiki wektorowej

EPS	-	Encapsulated PostScript
PDF	-	Portable Document Format (Adobe)
SVG	-	Scalable Vector Graphics (open)
SWF	-	Flash Adobe (dawniej Macromedia)
CDR	-	Corel
WMF	-	Windows MetaFile

KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
34

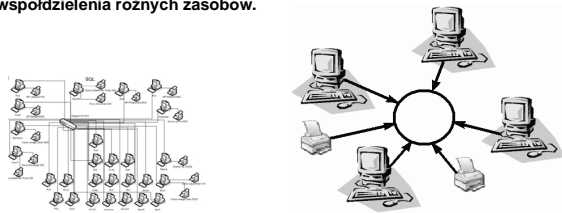
Sieć komputerowa i internet

Sieć komputerowa

- grupa komputerów lub innych urządzeń połączonych ze sobą w celu wymiany danych lub współdzielenia różnych zasobów.

Internet

- „międzysieć” ogólnosiwiatowa sieć komputerowa, czyli grupa komputerów lub innych urządzeń połączonych ze sobą w celu wymiany danych lub współdzielenia różnych zasobów.



KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
35

Sieć komputerowa i serwer

- **Sieć globalna, sieć rozległa (*Wide Area Network, WAN*)**
 - sieć komputerowa zasięgiem obejmująca duży obszar geograficzny (np. cały kraj); najpopularniejszą siecią rozległą jest internet. Zazwyczaj składa się z wielu połączonych sieci lokalnych.
- **Sieć lokalna (*Local Area Network, LAN*)**
 - najmniej rozległa postać sieci komputerowej obejmująca często kilka komputerów w jednym budynku.
- **Serwer**
 - program (potocznie również komputer, na którym zainstalowany jest program) umożliwiający udostępnianie lub wymianę danych między komputerami połączonymi w sieć komputerową

KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

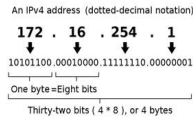
dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
36

IP i adres IP

- **IP (Internet Protocol)** – wewnętrzny protokół transmisji danych w formie pakietów.
- **TCP/IP (Transmission Control Protocol / Internet Protocol)**
- **Adres IP**
 - unikalna nazwa każdego urządzenia w sieci opartej na protokole TCP/IP wyrażona czterema oktetami oddzielonymi kropkami:
 - Oktet w praktyce oznacza 8 bitów, czyli 1 bajt i odpowiada jednej z cyfr od 0 do 255

130.14.25.1 = 130.014.025.001 – NCBI
 130.14 – domena NIH
 .25 – podsieć National Library of Medicine w NIH
 .1 – konkretny numer komputera w bibliotece



130.14.25.1 → „serwer nazw domen” → ncbi.nlm.nih.gov

Serwis internetowy i strona

- Serwis internetowy, witryna (**website**)
 - serwis informacyjny umieszczony w sieci; najczęściej wirtualny, interaktywny odpowiednik czasopisma, gazety, książki
- Strona internetowa
 - cyfrowy dokument kodowany w jednym z języków programistycznych zapewniających hipertekstowość (np. html, xml, php, flash itp.); kod interpretowany jest przez przeglądarkę internetową i wyświetlany w postaci odpowiednio sformatowanego tekstu; serwis internetowy składa się z serii połączonych tematycznie i fizycznie (hiperłączami) stron internetowych; odpowiednik kartki lub akapitu w książce, gazecie

Strona internetowa

```

7 <body>
8
9 <table width=250px border="1">
10 <tr><td>
11
12 <td> jest pogrubiony tekst</td>
13 </td> jest pochylony tekst</td>
14 <td> jest podkreślony tekst</td>
15
16 <td> jest wypunktowanie
17 <li>punkt 1</li>
18 <li>punkt 2</li>
19 <li>punkt 3</li>
20 </td></td></td>
21 </ul>
22
23 <center><td> jest wyśrodkowane</td></center>
24
25 <a href="adres odniesienia"> a to jest hiperłącze</a>
26
27 </td></tr></td>
    
```



Czy to jest tekst sformatowany, czy niesformatowany?

Strony domowe serwisów internetowych

PRZEGLĄDARKA INTERNETOWA / EDYTOR TEKSTU

Przeglądarka internetowa – program komputerowy, służący do pobierania i wyświetlania zawartości dokumentów z serwerów internetowych.

Edytor tekstu – program komputerowy ukierunkowany zasadniczo na samo wprowadzanie lub edycję tekstu, a nie na nadawanie mu zaawansowanych cech formatowania (do czego służy **PROCESOR TEKSTU**). W zależności od zastosowań, edytory tekstu nie mają w ogóle możliwości zajmowania się wyglądem i formatowaniem tekstu, skupiając się tylko na wprowadzaniu samych znaków, lub też mają te możliwości bardzo ograniczone.



TEKST SFORMATOWANY I NIESFORMATOWANY

TO JEST

Przykład tekstu sformatowanego

- Napisanego w edytorze typu WYSIWYG
- Edytorem takim jest np.: MSWord, OpenOffice, i inne

http, ftp i www

- **http** (*Hypertext Transfer Protocol* - protokół przesyłania dokumentów hipertekstowych)
 - to protokół sieci WWW. Za pomocą protokołu HTTP przesyła się żądania udostępnienia dokumentów WWW i informacje o kliknięciu odnośnika oraz informacje z formularzy. Zadaniem stron WWW jest publikowanie informacji - natomiast protokół HTTP właśnie to umożliwia.
- **ftp** (*File Transfer Protocol*)
 - protokół, który umożliwia przesyłanie plików z i na serwer poprzez sieć TCP/IP.
- **WWW** (*World Wide Web*)
 - (w skrócie określane jako WWW lub Web) jest hipertekstowym, multimedialnym, sieciowym (TCP/IP) systemem informacyjnym opartym na publicznie dostępnych, otwartych standardach IETF i W3C. Pierwotnym i w chwili obecnej nadal podstawowym zadaniem WWW jest publikowanie informacji.

Wikipedia

KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
43

Popularne protokoły wysokopoziomowe (aplikacyjne) i ich standardowe porty:

- BOOTP - serwer 67, klient 68
- DNS - 53
- Finger - 79
- FTP - 21
- Gopher - 70
- HTTP - 80, dodatkowe serwery, np. proxy, są najczęściej umieszczane na porcie 8080
- HTTPS - 443 (HTTP na SSL)
- IMAP - 143
- IMAP3 - 220
- Jabber
- IRC - 6667
- LDAP - 389
- LDAPS - 636 (LDAP na SSL)
- MySQL - 3306
- NNTP - 119
- POP3 - 110
- POP3S - 995 (POP3 na SSL)
- PostgreSQL - 5432
- Rsync - 873
- SMTP - 25
- SSH - 22
- Telnet - 23
- TFTP - 69
- WAP
- X11 - od 6000 do 6007
- XMPP

Numery portów reprezentowane są przez liczby naturalne z zakresu od 0 do 65535. Niektóre numery portów (od 0 do 1023) są ogólnie znane (*well-known port numbers*) i zarezerwowane na standardowo przypisane do nich usługi takie, jak np. WWW czy poczta elektroniczna. Dzięki temu możemy identyfikować nie tylko procesy, ale ogólnie znane usługi działające na odległych systemach.

KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
44

domena – adres WWW

- .com – domena komercyjna
- .edu – domena edukacyjna
- .gov – domena rządowa
- .mil – domena wojskowa
- .org – domena organizacji nieochodowej
- .pl – Polska
- .edu.pl – domena edukacyjna w Polsce

Forma ogólna URL:
protokół : //komputer .domena

130.14.25.1 → „serwer nazw domen” → ncbi.nlm.nih.gov

KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
45

Poczta elektroniczna

Adres e-mail posiada uniwersalną strukturę:

uzytkownik@komputer .domena

bioinformatyka@gmail.com

Received: from dodo.opmc.columbia.edu (dodo.opmc.columbia.edu [136.111.184.78]) by members.aol.com (6.9.3/4.9.3) with SMTP id PAA13177 for <scienceguy@aol.com>; Sun, 2 Jan 2000 17:51:22 -0500 (EST)
Received: (from phd@localhost) by dodo.opmc.columbia.edu (2000/27:001:0.0.0/000728:001:SMTP) id PAA03030 for scienceguy@aol.com; Sun, 2 Jan 2000 17:51:20 -0500 (EST)
Date: Sun, 2 Jan 2000 17:51:20 -0500 (EST)
Message-ID: <200001022251.PAA9030@dodo.opmc.columbia.edu>
From: phd@dodo.opmc.columbia.edu (PredictProtein) Nadawca, Adresat i Temat
To: scienceguy@aol.com
Subject: PredictProtein

PredictProtein Help
PHDsec, PHDacc, PHDdm, PHDtopology, TOP2S, MaxHom, EvalSec
Burkhard Rost

Table of Contents for PP help
1. Introduction
2. What is it?
3. How does it work?
4. How to use it?

<remainder of body truncated>

KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
46

Adres WWW / adres mailowy

- Adres WWW
 - **protocol://computer.domain**
 - http://ebiológ.pl
 - http://www.ebiológ.pl/index.html
 - ftp://ebiológ.pl/graf/
- Adres mailowy
 - **uzytkownik@komputer .domena**
 - jan.jastrzebski@uwm.edu.pl

KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
47

Języki hipertekstowe

- **html** *Hyper Text Markup Language, hipertekstowy język znaczników*,
 - to język składający się ze znaczników (ang. *tags*) stosowany do pisania stron WWW
- **php**
 - refleksywny skryptowy język programowania zaprojektowany do generowania dynamicznych stron internetowych
- **swf / flash**
 - technologia tworzenia animacji z wykorzystaniem grafiki wektorowej na zasadzie klatek kluczowych.

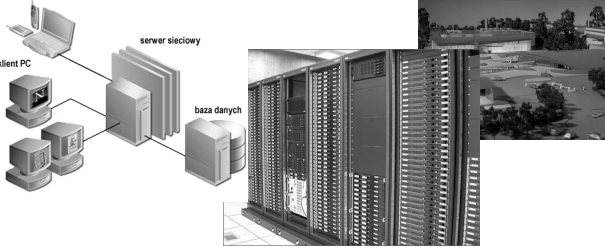
KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
48

Bazy danych

- **BAZA DANYCH** jest to uporządkowany zbiór danych o określonej strukturze, który zarządzany jest przez system DBMS.
- **DBSM - DataBase Management System**

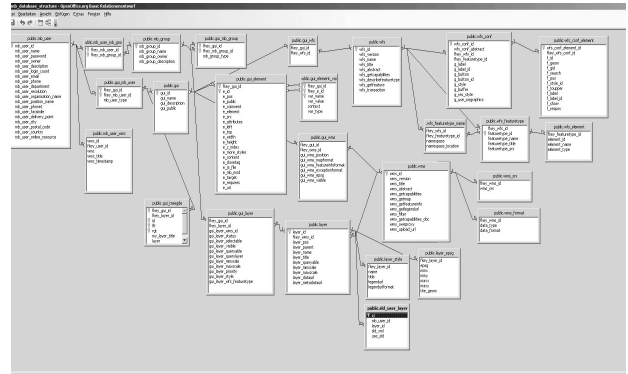


KiFBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
49

Struktura bazy danych



KiFBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
50

tabele

Tabela - jest podstawowym obiektem bazy danych stanowiącą zbiór informacji przedstawiona zwykle jako układ poziomych wierszy (rekordów) i kolumn (pól).

Name	Last Name	Address	City	State
Algera	2550 arden		Madison, WI 53715	485-2330
Hilda	1128 View Ave		Madison, WI 53715	265-1118
Amey	2650 Woodlea Lane		Madison, WI 53715	222-9887
Mar & Sall	214 E Dean Ave		Madison, WI 53716	
Martha	2102 Spradling St		Madison, WI 53716	221-5863
Virginia	Erwin Rd	Box 398	Interlachen, FL 32148	1847-7524
Christina	2601 Lakes Edge Rd		Madison, WI 53716	638-3644
Clare	210 E Dean Ave		Madison, WI 53716	222-6293
Christina	2601 Lakes Edge Rd		Madison, WI 53716	638-3644
Cherie	Curran	6702 Gray Fox Dr	Springfield, VA 22152	703-866-982
Carl	Curran	6702 Gray Fox Dr	Springfield, VA 22152	703-866-398
Rick & Jackie	Curran	1209 San Victoria, NZ	Abingdon, MA 02711	505-275-915
Mary	Curran	1820 N. Randall Ave, Apt 89	Jacksonville, FL 32246	789-0294
Maria	Purn	6006 N. Nakoma Dr	Coconutville, FL 33906	414-542-972
Ed & Louise	Edelman	1920 Broward St	Dallas, TX 75202	
Dr. Eleanor	Edehick	9626 Elm St	Riverside, CA 92504	414-629-410
Estimate	Edwards	4620 Woodland St	Madison, WI 53711	220-0236
Ray & Dorothy	Erwin	309 Vernal Pl	Bakersfield, CA 93304	805-831-699
Erica	Edgarsson	182 Elm St	Madison, WI 53711	
Dorothy	Feldman	7210 Blueford Lane	Madison, WI 53717	631-8418
Basim & Sawa	Eraser	128-Alex St	Tallahassee, FL 32307	913-767-948
Marcus South	Erwood	1726 N. 9th St	Los Angeles, CA 90019	800-967-364

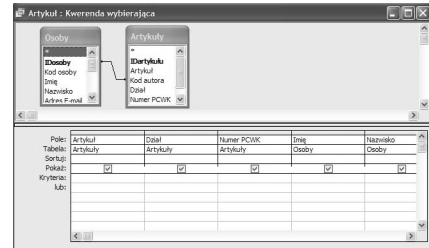
KiFBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
51

kwerendy

Kwerenda to obiekt bazy danych zawierający grupę rekordów po selekcji. Jest to żądanie okazania określonego zbioru danych. Kwerenda jest narzędziem, która zbiera dane z różnych tabel aby odpowiedzieć na pytanie zadane przez użytkownika. Jest podstawowym narzędziem analizy w bazie danych.



KiFBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
52

formularze

Formularz - jest to obiekt w którym umieszczamy formanty umożliwiające wprowadzanie, wyświetlanie i edycję danych.

KiFBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
53

raporty

Raporty - zawierają dane z tabel lub kwerend uporządkowane w żądany przez użytkownika sposób.

Nazwisko	Pierwsze imię	Drugie imię	Data urodzenia	Miejsce urodzenia
Frankowska	Marianna	Ewa	1989-02-15	Rawa
Brzoza	Waldemar	Przemysław	1989-04-08	Stasz
Konarski	Jan	Walerdy	1989-12-10	Prabuty

Liczba uczniów w klasie: 3
 Liczba chłopców: 2
 Liczba dziewcząt: 1
 Liczba uczniów mieszkających w mieście: 1
 Liczba uczniów mieszkających na wsi: 2

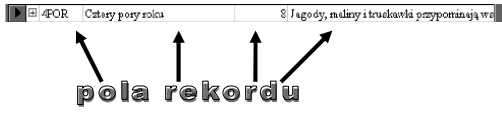
KiFBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
54

Rekord, pola i klucz

Rekord - zestaw informacji o pojedynczym elemencie tabeli bazy danych. W rekordzie powinno znaleźć się pole, które umożliwi jednoznacznie zidentyfikowanie rekordu, czyli klucz.



Klucz - atrybut nałożony na pole, zwykle w celu uniknięcia duplikowania się wartości. Kluczem identyfikującym może być kilka pól.

Rekord NCBI i pole rekordu

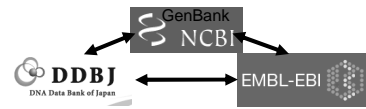
Tabela bazy danych

accession	organism	definition, protein name	sequence
ABK79072	Homo sapiens	hemoglobin	mvhlt...	...

„homo sapiens”[ORGANISM] AND hemoglobin[Protein Name]

BIOLOGICZNE BAZY DANYCH / serwisy bioinformatyczne

Biologiczne bazy danych są bibliotekami informacji z dziedzin nauk naturalnych. Dane gromadzone są z eksperymentów naukowych (zobacz pierwotne i wtórne dane oraz pierwotne i wtórne bazy danych), literatury (m.in. publikacje naukowe, książki, podręczniki) oraz analiz obliczeniowych (m.in. dane statystyczne, analizy bioinformatyczne). Biologiczne bazy danych zawierają informacje z takich dziedzin naukowych jak: genomika, proteomika, metabolomika, transkryptomika, mikromacierzowa analiza ekspresji genów, filogenetyka i tym podobne. Zbierane informacje dotyczą głównie funkcji i struktury genów, lokalizacji (zarówno jądrowej (chromosomalnej) jak i pozajądrowej), klinicznych efektów mutacji, podobieństwa sekwencji i struktury (białek i kwasów nukleinowych) oraz informacji postgenomowych.



MACIERZ

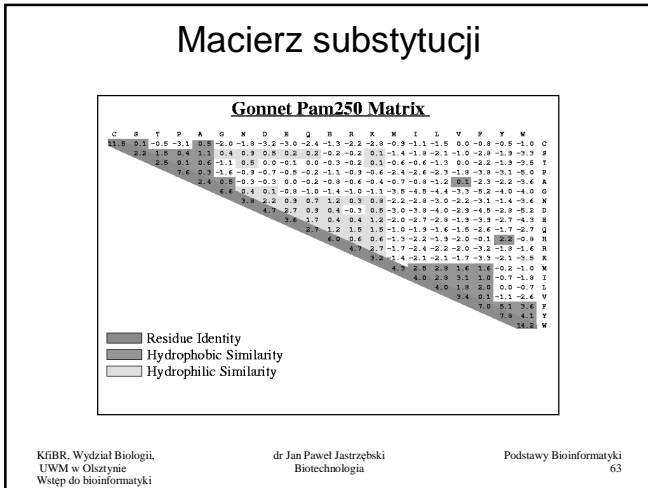
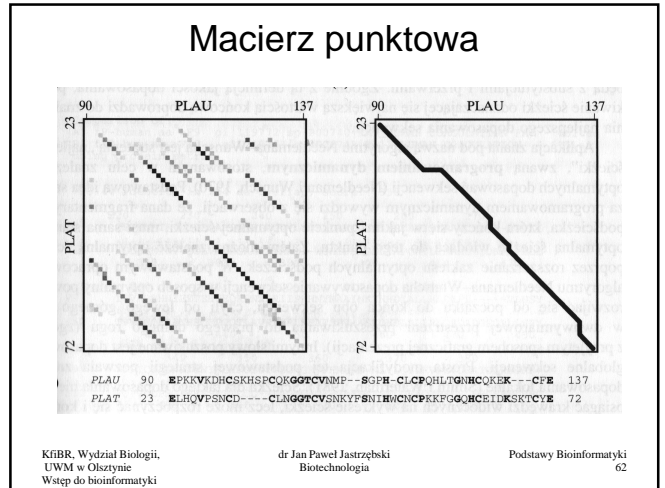
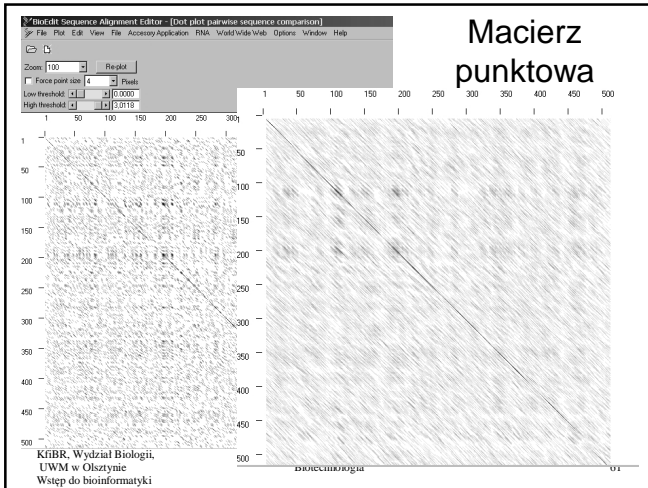
Macierz – układ zapisanych w postaci prostokątnej tablicy danych nazywanych *elementami* bądź *współczynnikami* będących elementami ustalonego zbioru, zwykle liczbowego.

INTERPOLACJA DANYCH

Dane [Grid of 3x3 squares]

INTERPOLACJA - "ZWIĘKSZANIE" ROZDZIELCZOŚCI

Jest to metoda matematyczna generowania brakujących danych w dokonanej serii pomiarów.



HTML

```

1 <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">
2
3 <HTML>
4
5 <HEAD>
6 <meta http-equiv="Content-type" content="text/html; charset=iso-8859-2" />
7 <meta name="Description" content="Schemat pliku HTML" />
8 <meta name="Author" content="Jan Paweł Jastrzębski" />
9 <title>TTTTT</title>
10 </HEAD>
11
12 <BODY>
13 <h1>nagłówek 1 rządek</h1>
14 <h2>nagłówek mniej ważny</h2>
15 <h6>itd...</h6>
16 </BODY>
17
18 </HTML>
19

```

KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
64

Strona internetowa

```

7 <body>
8
9 <table width=250px border="1">
10 <tr><td>
11
12 <b>to jest pogrubiony tekst</b>
13 <i>to jest pochylony tekst</i>
14 <u>to jest podkreślony tekst</u>
15
16 <ul>to jest wypunktowanie
17 <li>punkt 1</li>
18 <li>punkt 2</li>
19 <li>punkt 3</li>
20 </ul>
21
22 <center>to jest wyrównanie</center>
23
24 <a href="adres odniesienia">to jest hiperłącze</a>
25
26
27 </td></tr>

```

Podgląd w przeglądarce

KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki

dr Jan Paweł Jastrzębski
Biotechnologia

Podstawy Bioinformatyki
65

- ### Znaczniki HTML
- Znacznik ma następującą formę `<_>`, gdzie w miejsce `_` wstawia się odpowiedni ciąg znaków określający rodzaj znacznika, np.: `<a>`, ``, `<i>`, ``
 - Każdy otwarty znacznik BEZ WYJĄTKU musi zostać zamknięty:
 - ` link `
 - `
`
 - ``
- KfIBR, Wydział Biologii,
UWM w Olsztynie
Wstęp do bioinformatyki
- dr Jan Paweł Jastrzębski
Biotechnologia
- Podstawy Bioinformatyki
66

Znaczniki HTML

- Znaczniki posiadają atrybuty
 - ` link `
 - ``
 - `<option selected> wybrana opcja </option>`
- Znaczniki powinny być zawierane wewnątrz siebie, a nie naprzemiennie:
 - TAK `<a><i> text </i>`
 - NIE `<a><i> text </i>`

Polskie znaki



Koniec