Entrez Help

NCBI

Pomoc Entrez

Entrez skupia literaturę naukową, bazy danych DNA i sekwencji białkowych, informacje o strukturach 3D oraz domenach białek, zbiory danych o badaniach na temat populacji, informacje o ekspresji, zbiory kompletnych genomów i informacje taksonomiczne w ściśle powiązanym systemie. Jest to system zaprojektowany do przeszukiwania powiązanych z nim baz danych. Pomoc dotycząca komponentu literaturowego Entrez, znanego jako PubMed, jest również dostępna. Przejdź do Pomocy PubMed (PubMed Help).

Entrez, wyszukiwarka dla nauk przyrodniczych: globalna kwerenda

Na stronie Entrez znajduje się wyszukiwarka bazy danych Entrez Global Query (strona przeszukiwania przekrojowej bazy danych Entrez). Cała grupa poszczególnych baz danych Entrez jest zorganizowana na tej stronie, z bazami literaturowymi na szczycie, włączając w to: PubMed, PubMed Central, Journals, Books, OMIM i OMIA. Strona NCBI Site Search jest również wykazana. Bazy danych sekwencji składają się z baz: Nucleotide, Protein, Genome, Structure, and SNPs. Pozostałe bazy danych to Taxonomy, Gene, UniGene, HomoloGene, Conserved Domains, 3D Domains, UniSTS, PopSet, GEO Profiles, GEO Datasets, PubChem Bio-Assay, PubChem Compound, PubChem Substance, Cancer Chromosomes, Probe, MeSH, Journals i NLM Catalog. Linki do popularnych stron NCBI jak PubMed, Human Genome, Map Viewer i BLAST są na pasku narzędzi. Jest tam również link do bazy danych GenBank prowadzący do nukleotydowej bazy danych.

Używając Entrez Global query, przeszukiwanie wszystkich baz danych Entrez jest wykonywane poprzez wprowadzenie szukanego pojęcia lub frazy w oknie dialogowym kwerendy "Search across databases". Wciśnij przycisk Go by uruchomić wyszukiwanie, lub naciśnij przycisk Enter na klawiaturze. Przycisk CLEAR kasuje treść zapytania w okienku dialogowym kwerendy; użyj go do rozpoczęcia nowego wyszukiwania. Rezultaty znalezione w każdej bazie danych są pokazane na stronie

Global Query. Naciśnij na numerze rezultatu lub sąsiadującą nazwę bazy danych aby przenieść się do wybranego rezultatu. Zobacz link do dokumentu Global Query Help, który jest na prawo od przycisku CLEAR.

Bazy danych

Umbrella Nucleotide Database

Tradycyjną nukleotydową bazę danych NCBI można teraz również przeszukiwać jako trzy bazy danych: „EST” (obejmującą sekwencje EST), „GSS” (obejmującą sekwencje GSS) i "CoreNucleotide" (która zawiera resztę sekwencji nukleotydowych). To umożliwi szybsze przeszukiwanie i uzyskiwanie dokładniejszych rezultatów dla rejestru nukleotydowych sekwencji. Kiedy wyszukiwanie jest zakończone w nukleotydowej bazie danych, rezultaty wyszukiwania Entrez są również pokazywane dla trzech komponentów nukleotydowej bazy na Search statistic Line. Komponent nukleotydowej bazy razem obejmuje wszystkie dane sekwencji z GenBank, EMBL, i DDBJ, elementy International Collaboration of Sequence Databases. W celu uzyskania szczegółowych informacji o bazach danych i ich wyszukiwaniu zobacz dokument Entrez Help.

Jako przykład zmiany formatu rezultatów zwróconych z przeszukiwania Entrez Nucleotide, rozważ proste wyszukiwanie używające kwerendy typu "mouse[organism]". Rezultaty są pokazywane razem z linkami do podzielonych informacji w GenBank EST division, the GenBank GSS division, and the CoreNucleotide subset. Użytkownicy mogą wybrać rezultaty z interesujących ich zbiorów nukleotydowych przez kliknięcie właściwego linku wyników.

Nowe składnikowe bazy danych są zawarte w schemacie połączeń Entrez i Links wewnątrz i pomiędzy bazami danych mogą być wybierane jak dotychczas z różnych zbiorów danych. Popularne strategie wyszukiwania takie jak Limits, Preview/Index, History, and MyNCBI mogą być używane w obrębie każdej poszczególnej bazy danych.

Specjalne pola wyszukiwania są dostępne dla każdej z nowych składowych baz danych, które dostrzec można w sekcji „Add Term(s) to Query or View Index:” zakładki Preview/Index. Nowe pola dostępne dla składowej bazy danych EST zawierają EST ID, EST Name, and Library. Baza danych GSS posiada następujące

nowe pola: GSS ID, GSS Name, Library Class, and Library Name. CoreNucleotide zawiera te same 23 pola wyszukiwania jak tradycyjna nukleotydowa baza danych.

Wszystkie poprzednie funkcje Entrez pozostały, takie jak Clipboard (Schowek) gdzie wyszukiwania mogą być tymczasowo zapisane i MyNCBI gdzie wyszukiwania mogą być zapisane na nieokreślony czas by mogły zostać uruchomione przez użytkownika w wybranych przez niego przedziałach czasu. Strony Details and History są dostępne dla indywidualnych planów wyszukiwania. Aby zobaczyć szczegóły wyszukiwania, interesująca nas baza danych musi zostać wybrana podczas wykonywania wyszukiwania z głównej strony Nucleotide.

Opatentowane sekwencje są włączone poprzez porozumienie z U.S. Patent and Trademark Office (USPTO) oraz inne międzynarodowe biura patentowe poprzez współpracujące międzynarodowe bazy danych.

The Umbrella Nucleotide database również zawiera rekordy Reference Sequence (RefSeq). RefSeqs są wspieranymi przez NCBI niepowtarzalnymi zbiorami sekwencji. Naciśnij ten link dla uzyskania więcej informacji dotyczących projektu RefSeq.

Białkowa baza danych (Protein Database)

Białkowa baza danych zawiera informacje z translacyjnych regionów kodujących sekwencji DNA w GenBank, EMBL i DDBJ, jak również sekwencje białkowe przekazywane do Protein Information Resource (PIR), SWISS-PROT, Protein Research Foundation (PRF), and Protein Data Bank (PDB) (sekwencje z rozwiązanych struktur).

Genomowa baza danych (Genome Database)

Genomowa baza danych zapewnia widok na różne genomy, kompletne chromosomy, mapy sekwencji z contigs, i połączone mapy genetyczne i fizyczne.

Baza danych struktur (Structure Database)

Baza danych struktur lub Molecular Modeling Database (MMDB) zawiera informacje doświadczalne z krystalograficznego i NMR oznaczania struktur. Informacje dla MMDB są otrzymywane z Protein Data Bank (PDB). NCBI wiąże dane strukturalne z informacjami bibliograficznymi, bazami danych sekwencji i taksonomią.

Użyj Cn3D, przeglądarkę struktur NCBI 3D dla łatwej, interaktywnej wizualizacji struktury molekularnej z Entrez.

Domeny 3D (3D Domains)

Domeny 3D zawierają domeny białkowe z Entrez Structure Database. Zobacz 3D Domains.

Domeny konserwatywne (Conserved Domains)

Konserwatywne domeny jest bazą danych domen białkowych. Źródłowe bazy danych dla Konserwatywnych Domen to Pfam, Smart i COG. Zobacz CDD Help.

UniSTS

UniSTS jest zunifikowanym, niepowtarzalnym przedstawieniem znakowanych końców sekwencji (STSs). UniSTS łączy marker i informacje o mapowaniu z różnych, ogólnodostępnych zasobów. Źródła danych obejmują dbSTS, RHdb, GDB, różne ludzkie mapy (genetyczna mapa Genethon, genetyczna mapa Marshfield, mapa Whitehead RH, mapa Whitehead YAC, mapa Stanford RH, fizyczna mapa NHGRI chromosomu 7 i fizyczna mapa chromosomu X) i różne mapy myszy (mapa Whitehead RH, mapa Whitehead YAC i mapa MGD Jackson Laboratory). Zobacz UniSTS.

Gen (Gene)

Gene zapewnia jednolite środowisko kwerendy dla genów określone przez sekwencje i/lub w Map Viewer NCBI. Można pytać o nazwy, symbole, dodatki, publikacje, terminy GO, numery chromosomów, numery E.C., i wiele innych cech związanych z genami i kodowanymi produktami. Zobacz Gene i Gene Help.

UniGene

UniGene jest systemem doświadczalnym dla automatycznie partycjonowanych sekwencji GenBank w niezbędny zestaw genowo ukierunkowanych grup. Każda grupa UniGene zawiera sekwencje, które reprezentują unikatowy gen, jak również powiązane informacje takie jak typy tkanek, w których gen ulegał ekspresji i lokalizacja na mapie. Zobacz UniGene i jego wskazówki zapytań i FAQs.

HomoloGene

HomoloGene jest systemem dla automatycznej detekcji homologów pośród zbioru eukariotycznych genów. Zobacz HomoloGene i wskazówki zapytań (Query Tips).

SNP

SNP jest centralnym magazynem danych dla obu pojedynczych baz substytucji nukleotydów i krótkich delecji i polimorfizmów insercji. Dla znalezienia strony i dostępnych pól szukania i przykładów wyszukiwań, zobacz SNP.

Baza danych PopSet (PopSet Database)

Baza danych PopSet zawiera dopasowanie sekwencji dostarczone jako zespół wynikający z badań populacji, filogenezy lub mutacji. Te dopasowania opisują takie zdarzenia jak ewolucja i zmiany w populacji. Baza danych PopSet zawiera dane zarówno o sekwencjach nukleotydowych jak i białkowych. Zobacz PopSet.

Taksonomiczna baza danych (Taxonomy Database)

Taksonomiczna baza danych zawiera nazwy wszystkich organizmów, które są reprezentowane w genetycznej bazie danych NCBI przez przynajmniej jedną sekwencję nukleotydową lub białkową. Dla uzyskania kontekstu taksonomicznej bazy danych zobacz Taxonomy i Taxonomy FAQs.

Profile GEO (GEO Profiles)

Ta baza danych przechowuje profile ekspresji genów i profile mnóstwa cząsteczek gromadzone w magazynie Gene Expression Omnibus (GEO). Zobacz GEO QuickQuery(szybkie zapytanie) i FAQs.

Zbiory danych GEO (GEO Datasets)

Baza danych przechowuje zbiory danych dotyczące ekspresji genów i mnóstwa cząsteczek uzyskanych z repozytorium Gene Expression Omnibus (GEO) . Zobacz GEO FAQs.

GENSAT

Projekt GENSAT dąży do odwzorowania ekspresji genów w ośrodkowym układzie nerwowym myszy, używając zarówno hybrydyzacji in situ i transgenicznych metody myszy. Strona domowa GENSAT pozwala odnaleźć GENSAT FAQ w celu uzyskania dodatkowych informacji.

Chromosomy raka

Chromosomy raka zawierają trzy cytogenetyczne bazy danych: baza danych Mitelman NCI dotycząca aberracji chromosomowych w raku, NCI powracających aberracji chromosomowych w raku, NCI i NCBI SKY/M-FISH & baza CGH. Kariotyp SKY/M-FISH i dane CGH mogą być przeszukane jednocześnie. Podobieństwo uwidacznia się cytogenetycznie i na różnych poziomach specyficzności pokrewieństwa. Oglądając witrynę internetową Chromosomy raka uzyskujemy dodatkowe informacje.

Pubchem Compound

Baza danych PubChem zawiera potwierdzone informacje chemiczne przedstawiające opisane substancje w PubChem Substance. Strona internetowa PubChem pozwala uzyskać dodatkowe informacje.

PubChem Substance

Baza danych PubChem Substancja zawiera opisy chemicznych próbek z różnych źródeł i cytaty linków do PubMed, białka o strukturze3 D i biologiczne rezultaty uwidaczniające rezultaty, które są dostępne w PubChem BioAssay. Witryna internetowa PubChem Substancja pozwala uzyskać i poszukać dodatkowych informacji.

PubChem BioAssay

Baza danych PubChem BioAssay zawiera bioaktywne ekrany substancji chemicznych opisanych w PubChem Substance. To dostarcza wszystkich wyszukiwanych opisów każdego pomiaru aktywności biologicznej, wliczając w to opisy warunków i opis właściwy dla tej procedury. Witryna internetowa PubChem BioAssay pozwala uzyskać i poszukać dodatkowych informacji.

PubMed Central

PubMed Centralny (PMC) jest cyfrową Biblioteką Narodową zawierającą w archiwum medyczne czasopisma z zakresu nauk przyrodniczych. Dostęp do pełnego tekstu artykułów w PMC jest wolny za wyjątkiem, gdy czasopismo wymaga prenumeraty dla dostępu do ostatnich artykułów. Oglądając PubMed Central, PubMed CentralHelp i PubMed Central FAQ.

Journals

Baza danych zawierająca czasopisma może być przeszukiwana przy użyciu tytułów czasopism, skrótu Medline, NLM ID, skrótu ISO albo ISSN. Baza danych zawiera czasopisma w całej bazie danych Entrez np. PubMed, Nucleotide, Protein.

MeSH

MeSH jest Medyczną Biblioteka Narodową, która jest używana do włączania do katalogu artykułów w PubMed. MeSH terminologia dostarcza zgodnego sposobu by uzyskać informacje, które mogą korzystać z innej terminologii ale w takim samym pojęciu.

Bookshelf

Bookshelf zawiera kolekcję biomedycznych książek, które są połączone z Entrez. Przewodnik NCBI jest również dostępny z Bookshelf. Zobacz Bookshelf i Books FAQs

Baza danych OMIM

Baza danych OMIM jest katalogiem ludzkich genów i chorób genetycznych. Zobacz OMIM i pomoc OMIM.

Baza danych OMIA

Baza danych OMIA jest bazą genów, przejętych w spadku i występujących u innych gatunków zwierzęcy (inny niż człowiek i mysz) zredagowanych przez profesora Frank Nicholas z Uniwersytetu Sydney, Australia, z pomocą wielu innych ludzi przez lata. Baza danych zawiera tekstowe informacje i odniesienia, jak również link do istotnych rekordów z OMIM, PubMed, Gen, i fenotypowe bazy danych NCBI. Przejrzyj witrynę internetową OMIA.

Bazy danych prób

Próby są publicznym zapisem kwasu nukleinowego przeznaczonych do użytku w wielkim wyborze biomedycznych projektów badawczych, razem z informacjami o dystrybutorach odczynnika, skuteczność prób i obliczonym podobieństwem ciągu. Zobacz Probe i Probe Query Tips.

Przeszukując Entrez używamy Globar Query

Na Globar Query strona Entrez umożliwia dokładnie przeglądać stronę jednego typu albo więcej wyszukiwań w innych okna. Rezultaty dla każdej bazy danych są pokazane i może być wybrany dla wybranych bazy danych. Strona Globar Query jest pokazana poniżej.

Strona Global Query

Przeszukiwanie bazy Globar Query może być również przekierowane do NCBI i na stronę główną NCBI przez wchodzenie przez wyszukiwane pliki. Jedna baza danych może być wybrana ze wszystkich baz danych określonego menu na stronie głównej NCBI.

Bardziej złożone strategie przeszukiwania są wykonywane przy użyciu operatorów Boolean i kombinacji jednego lub więcej limitów.

Baza danych Neighbors i Interlinking — co sprawia, że Entrez jest potężniejszy niż wiele innych rekordów związanych z innymi zapisami danej bazie danych, obydwa w danej bazie danych (taki jak Nukleotyd) i między bazami danych. Linki w bazie danych są zatytułowane “ sąsiedzi ” (np., Nukleotydowi sąsiedzi).

Nukleotydy, których sekwencje są opublikowane. Rekordy białkowych sekwencji są związane z nukleotydowymi sekwencjami, których białka zostały poznane.

Zobacz Displaying i Saving Results w celu uzyskania większej ilości informacji.

Operatory Boolean

Operatory Boolean używane w Entrez to:

AND: użycie tego operatora umożliwia przeszukanie dwóch okresów w Entrezie w celu znalezienia wszystkich dokumentów, które zawierają te dwie informacje naraz.
OR: użycie tego operatora umożliwia przeszukanie dwóch okresów w Entrezie w celu znalezienia dokumentów, które zawierają oba okresy.
NOT: użycie tego operatora umożliwia przeszukanie dwóch okresów w Entrezie w celu znalezienia dokumentów, które zawierają jedną ale nie zawierają drugiej informacji.

Zasady przeszukiwania Entreza i składnia do użycia operatorów Boolean:

1. Operatory Boolean AND, OR, NOT muszą być zanotowane w UPPERCASE (np. Operator OR odpowiedni element)
2. Procesy Entrez z wszystkimi operatorami Booleany odbywają się od lewej do prawej strony. Porządek, w którym Entrez poszukuje oświadczenie może być przerobione przez załączanie pojedynczych pojęć w nawiasie. Okresy wewnątrz nawiasów okrągłych są przetworzone po raz pierwszy jako jednostka a następnie włączą do ogólnej strategii. Np. znalezienie g1p3 AND(odpowiedni element promotora OR) odbywa się w Entrazie w Ring w odpowiednim okresie przez odpowiedni element OR na początku i ANDing jako zwykły zbiór dokumentów z g1p3.
3. Kliknij na przycisk Details, aby zobaczyć jak Entrez przetłumaczył i przeszukał Twoją strategię

4. Zobacz Writing Advanced Search Statements aby uzyskać więcej informacji o użyciu operatorów Boolean i Entrez Search Field Qualifiers.

Wykorzystanie nawiasów okrągłych może zmieniać twoje poszukiwania wynika znacznie. Porównywać numer płyt aportowanych w Nukleotydowej bazie danych począwszy od lutego 2006 w każdym przypadku poniżej.

Example
glp3 AND (response element OR promoter) retrieves three records
glp3 AND response element OR promoter retrieves 354,554 records.

Wyszukiwanie graniczne lub wyszukiwanie fraz

Zależne terminy są automatycznie łączone (ANDed). Następują po pytaniu (kwerendzie), które nie jest umieszczone w podwójnym cudzysłowie: 16 S RNA - wyszukuje zupełnie wszystkie rekordy przy pomocy terminów 16 S i RNA (16 S AND RNA). Zobacz operatory Boole'a, aby uzyskać więcej informacji na temat łączenia terminów z operatorami Boole'a.

Aby dokładnie przybliżyć terminy graniczne wyszukiwania, albo czynnik Entrez do wyszukiwania fraz, należy zastosować podwójne cytaty (” ”) wokół frazy. Terminy w zacytowanej frazie są przeszukiwane i umieszczone w zacytowanej frazie ”i” następny jeden do drugiego. To jest siła, sens wyszukiwania fraz.

Stosowanie cytatów czynnika Entrez do sprawdzania listy fraz, w stosunku do których poszukiwane terminy są dopasowane, to nie jest prawdziwe wyszukiwanie graniczne. Jeżeli poszukiwana fraza nie jest na liście, do Entrez wraca pewna informacja: ”Zobacz szczegóły. Brak pozycji znalezionych” i równocześnie w nukleotydowej bazie danych pojawia się wiadomość : ” Podana fraza nie została znaleziona. Zobacz szczegóły. Brak pozycji znalezionych.”

Chociaż fraza szukana jest przydatna, to powinna być ostrożnie używana, ponieważ umieszczając poszukiwane terminy w cudzysłowia, ograniczamy odszukiwane

dokumenty tylko do tych skojarzonych z wymaganym tekstem wewnątrz cudzysłowia. Również pamiętaj, że nie wszystkie frazy będą umieszczone w spisie Entrez nukleotydowym lub w spisie białek. Dlatego można także ponownie uruchomić wyszukiwanie bez załączania wyrazów w cudzysłowie. Dokładnie sprawdź twoje wyniki, aby zobaczyć czy twoja fraza została znaleziona: W przykładzie ”16 S RNA”, dokumenty z wyrażeniem 16 S RNA są wyszukiwane, ale dokumenty z frazą 16 S RNA genu – już nie. PubMed prowadzi zacytowane wyszukiwanie w różny sposób a szczegóły na ten temat zobacz w dokumentach pomocy PubMed.

Porównanie liczby rekordów odnalezionych w nukleotydowych bazach danych począwszy od lutego 2006 w każdym przypadku poniżej.

Na przykład zacytowana fraza ”16 S RNA” odnajduje mniej dokumentów, porównując wyszukiwanie z 16 S AND RNA.

Przykład:

16 S rRNA odnaleziono 299,494;

”16 S rRNA” odnaleziono 85,500;

Wyszukiwanie ze względu na autorów

Wprowadź nazwisko autora w formacie: nazwisko oraz inicjały (e.g., johnson d). Nie używaj znaków interpunkcyjnych. Ten format zobowiązuje Entrez do wyszukiwania tylko pól autor. Entrez automatycznie usuwa nazwisko autora w celu uwzględnienia różnych inicjałów i określeń, takich jak Jr. albo 2nd. Jeżeli nazwisko jest jedynym terminem wpisywanym w pole wyszukiwania (e.g., johnson), Entrez będzie przeszukiwał wszystkie pola dla tej frazy. Można również użyć ogranicznika Entrez dla frazy: Johnson [AUTHOR]. To jest także informacja zobowiązująca Entrez do wyszukiwania tylko w polu autor.

Wyszukiwanie ze względu na wyjątkowe identyfikatory

Unikalne identyfikatory mogą być **numerami akcesyjnymi** (wstąpienia), które odnoszą się do kompletnego rekordu sekwencji, albo **numery identyfikacji kolejności**, które odnoszą się do poszczególnych sekwencji w rekordzie.

Format **numerów akcesyjnych** (wstąpienia) zmienia się, zależnie od źródła bazy danych.(Jak wspomniano wyżej w sekcji danych, każda domena danych Entrez zawiera rekordy od numerów różnych źródeł). Niektóre przykłady typowych formatów liczb wstąpienia znajdują się poniżej. Przykładowy rekord GenBank zawiera dodatkowe szczegóły dotyczące numerów akcesyjnych (wstąpienia). Kliknij tutaj, aby wyświetlić listę prefiksów numerów wstąpienia (akcesyjnych).

Typ rekordu	**Przykład formatu akcesyjnego (wstąpienia)**
GenBank\EMBL\DDBJ rekordy nukleotydowych sekwencji	Jedno oznaczenie literowe, mające za sobą 5 cyfr e.g.: U12345 Dwa oznaczenie literowe, mające za sobą 6 cyfr e.g.: AY123456, AF123456;
Rekordów sekwencji GenPept (zawierające aminokwasową sekwencję translacyjną z GenBank\EMBL\DDBJ rekordy mają kodować cechy regionu)	Trzy oznaczenia literowe i pięć cyfr e.g.: AAA12345;
Proteinowa sekwencja rekorów z SWISS-PROT i PIR	Wszystkie sześć charakterystycznych: Charakter/Format 1[O,P,Q] 2[0-9] 3[A-Z,0-9] 4[A-Z,0-9] 5[A-Z,0-9] 6[0-9] e.g.: P12345 i Q9JJS7
Zapisy proteinowych sekwencji z PRF	Seria cyfr (6 lub 7), mająca za sobą oznaczenie literowe, e.g.:1901178A
Zapisy nukleotydowych sekwencji RefSeq	Dwa oznaczenia literowe, podkreślnik i 6 cyfr e.g.: mRNA zapis (NM_*): NM_000492, genomic DNA contigs (NT_*): NT_000347 complete genome or chromosome (NC_*), NT_000907genomic region (NG_*) NG_000019,
RefSeq model (przewidywany) rekordów sekwencji dla genomu człowieka	Dwa oznaczenia literowe (XM,XP,XR), podkreślnik i 6 cyfr e.g.: XM_000483
RefSeq zapisy sekwencji proteinowych	Dwa oznaczenia literowe (np.), podkreślnik i 6 cyfr e.g.: 000483
Zapisy struktur białkowych	PDB wstąpienia posiadają zazwyczaj jedną cyfrę i trzy oznaczenia liczbowe e.g: 1TUP

	MMDB numer ID zawieraja zazwyczaj cztery cyfry e.g.: 3973 Rekord dla supresorów nowotworów P53 złożony z DNA może być uzyskiwany przez oba powyższe numery.

Istnieją dwa typy numerów identyfikacji sekwencji:

- Numery GI: seria cyfr, które są wyznaczane po kolei przez NCBI, przetwarzane dla każdej sekwencji;
- Numery wersji: składa się z numeru akcesyjnego (wstąpienia), mającego za sobą kropkę i numer wersji;

Na przykład, zapis RefSeq dla Homo sapiens – regulator przewodności (cftr) dla torbielowatej fibrozy przezbłonowej, ma numer wstępujący NM_000492. Rekord zawiera jedną sekwencję nukleotydową i jedną aminokwasową sekwencję translacyjną, które mają następującą sekwencje identyfikacyjną:

Sekwencja nukleotydowa:

GI: 6995995;
Numer wersji: NM_000492.2;

Aminokwasowa sekwencja translacyjna:

GI: 6995996
Numer wersji: NP_000483.2;

W przypadku zmiany kolejności w jakikolwiek sposób, rekord otrzymuje nowy numer GI, a numer wersji zostaje zwiększony o jeden. Przykładowy rekord GenBank zawiera dodatkowe szczegółowe informacje GI i numery wersji identyfikacji sekwencji.

Przeszukiwanie ze względu na masę cząsteczkową

NCBI wprowadził termin masy cząsteczkowej (wagi molekularnej) do przeszukiwania proteinowych baz danych Entrez, na wniosek masowej grupy spektrofotometrii w NIH. Dr Lewis Pannell zapewnił doradztwo techniczne.

Pole masy cząsteczkowej można przeszukiwać za pomocą pojedynczej masy:

2002 [Molecular Weight]

lub zakresu wagi:

2002:2009 [Molecular Weight].

Oba wyrażenia mogą zostać połączone z innymi terminami wyszukiwania Entrez, np. by ograniczyć wyszukiwania do danego organizmu:

002002:2009 [Molecular Weight] AND human [Organism].

Kwadratowe nawiasy mogą zawierać pełną pisownię w polu wyszukiwania, jak w powyższym przypadku, lub skrót [MOLWT] – małe lub duże znaki.

Należy również zauważyć, że w przypadku rozkładu produktu są one analizowane z cechami: masą cząsteczkową obliczoną dla każdego z produktu rozłamu, a nie masą całego białka. W ten sposób można wyszukać duże białka, podczas pytania o małe masy cząsteczkowe; należy sprawdzić cechy z tabeli rekordów proteinowych by zobaczyć czy mają one produkty rozkładu.

Jak obliczana jest masa cząsteczkowa: – jeżeli produkty rozkładu zostały odnotowane, masa cząsteczkowa jest obliczona dla każdego rozłożonego produktu, a nie dla całego białka. Rozszczepione produkty nie są konsekwentnie odnotowywane, ale dołożyliśmy wszelkich starań w celu wykrycia różnych stylów baz danych. Na przykład, produkt rozpadu nie jest odnotowywany jako „mapt” w GenBank ale jako „Region”z”/region_name= Mature chain" w SWISS-PROT.

Należy pamiętać, że oznacza to, że więcej niż jedna masa cząsteczkowa może wskazywać na jeden rekord białka!

Jeżeli tylko peptyd sygnalny został odnotowany, to zostanie on usunięty, a masa cząsteczkowa jest obliczana dalej z reszty białka.

Jeżeli na białku nie ma takich funkcji, wtedy masa cząsteczkowa obliczana jest w odniesieniu do całego białka. W tym przypadku sprawdzenie jest dokonywane dla uwzględnienia początkowej metioniny (Met), i nie jest to włączane przy obliczeniach, w przypadku stwierdzenia obecności.

Jeżeli stwierdzono występowanie zupełnie nieznanych aminokwasów (e.g.,"X"), masa cząsteczkowa nie jest obliczana.

Dwuznaczne aminokwasy są obliczane jako jedna z ich możliwych form:

B znaczy D albo N – masa cząsteczkowa obliczona jako D

Z znaczy E albo Q – masa cząsteczkowa obliczona jako E

Masa cząsteczkowa jest obliczana jako część procesu indeksowania rekordów białka w Entrez. Masa cząsteczkowa Entrez jest średnią masą cząsteczkową, nie monoisotopic. Masy są zaokrąglane do najbliższej liczby całkowitej. Masy występują tylko w indeksie Molecular Weight i nie są wyraźnie widoczne w rekordach sekwencji białkowych.

Wyszukiwanie zakresu (przedziału, zasięgu)

Wyszukiwanie zakresu może być wykonane na 4 elementach danych: numery dostępu [ACCN], długości sekwencji [SLEN], wagi cząsteczkowe [MOLWT], i dane [MDAT] i [PDAT]. Operatorem zakresu jest dwukropek (:), oraz odpowiedni kwalifikator pola powinien być zawarty w nawiasach kwadratowych po drugim określeniu.

Kwalifikatory pola nie rozróżniają małych i dużych liter więc zarówno [ACCN] jak i [accn] zadziała. Nie jest koniecznym załączanie wolnego miejsca pomiędzy wyszukiwanym terminem oraz kwalifikatorem pola, chociaż może być zrobione jeśli chcemy.

Przykładowe wyszukiwanie : - Zakres numerów dostępu:

AF114696:AF114714[ACCN].

Przypis: Niemożliwe jest aby wyszukać zakres sekwencyjnych numerów identyfikacyjnych (znanych jako numery GI I Version numbers).

Zakres długości sekwencji:
3000:4000[SLEN]

Zakres wag cząsteczkowych może być szukany w białkowej bazie danych
2002:2009[MOLWT]

Przypis: Dodatkowe informacje o Wyszukiwaniu za pomocą Wag cząsteczkowych są zawarte powyżej.

Wyszukiwanie zakresu może być również złożone z innych terminów wyszukiwania w Entrez, na przykład, ograniczanie przez organizm:

Wybierz białkową bazę danych, wejdź w wyszukiwanie w okienku zapytania:
2002:2009[MOLWT] i człowiek [ORGN]

W bazie danych zarówno nukleotydowej jak i białkowej :4000[SLEN] I człowiek [ORGN]

Aby stworzyć zakres, poszukaj 11 z 999,999 z sekwencji nukleotydów. Wprowadź:
11:999999[SLEN]

Zakres danych: 1998/02:2000/01/25[MDAT]

Ucinanie/Skracanie

Skracanie terminów wyszukiwania jest wygodnym sposobem znajdywania wszystkich rekordów (chodzi tutaj o te zapisy w bazach danych), które zawierają terminy

zaczynające się danym początkiem tekstu. Umieść gwiazdkę (*) na końcu wyszukiwanego terminu, aby znaleźć wszystkie rekordy z terminem, który rozpoczyna się tym ciągiem tekstowym. Dla przykładu, ucięte wyszukiwanie terminu „immunoglob*” pobierze wszystkie rekordy w bazie danych, które zawierają słowo immunoglobulin, immunoglobulins, immunoglobin, i immunoglobins. Wyszukiwanie Entrez w białkowej bazie danych:

Example
immunoglob* results in 99,760 records

Entrez wyszukuje pierwsze 600 opcji skróconego terminu. Jeśli ucięty termin daje w wyniku więcej niż 600 opcji, co jest możliwe z terminami takimi jak „bac*”, Entrez daje następujące ostrzeżenie:
„Wieloznaczne wyszukiwanie dla „bac*” użyło jedynie 600 pierwszych wariantów. Przedłuż słowo źródłowe dla wszystkich końcówek.”

Wyrażenia, które zawierają wolne miejsce w słowie po gwiazdce NIE BĘDĄ pobierane. Na przykład, jeżeli szukasz „chromo*” pobrane dokumenty będą zawierać terminy takie jak chromabacterium ale tak jak chromo helicase już nie.

Lewostronne ucięcie jest niemożliwe (np. „*bacterium”).

Łączenie zbiorów

Użyj swojej historii wyszukiwań aby połączyć dokumenty pobrane z różnych wyszukiwanych terminów w innym czasie podczas twojej sesji poszukiwań. Jako przykład, wyszukaj bazę danych nukleotydów dla HIV. To wyszukiwanie pobierze ponad 188,000 dokumentów. Teraz wyszukaj bazę danych nukleotydów dla proteazy. To wyszukiwanie pobierze ponad 92,000 dokumentów. Następnie kliknij w historię dla nukleotydowej bazy danych.

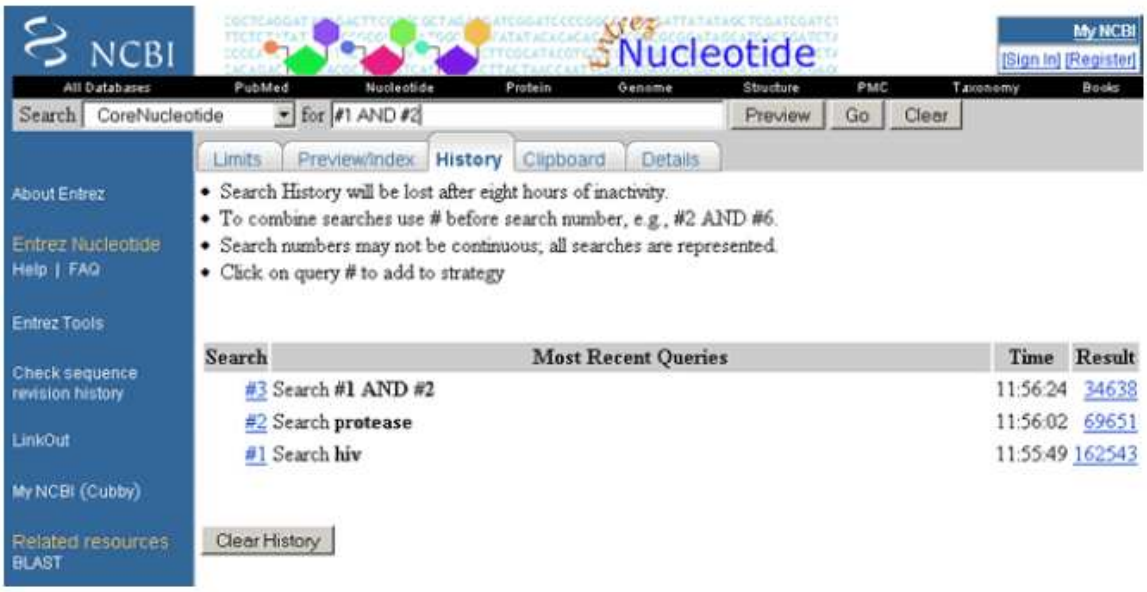
Rezultaty dla HIV i proteazy szukanych terminów są zapisane jako Search Sets#1 i #2 odpowiednio. W okienku wpisz #1 I #2 i wybierz „Go”. To wyszukiwanie połączy dokumenty z wyszukiwania #1 (HIV) z dokumentami z wyszukiwania #2 (proteaza) i pobierze tylko te dokumenty, które są w obydwu zespołach.

Zapamiętaj, ta historia jest jedynie dla nukleotydowej bazy danych i zostanie usunięta po 8 godzinach nieaktywności.

Numery wyszukiwań mogą nie być stałe; wszystkie wyszukiwania są zastępowane.

Zobacz Boolean operator i używanie swojej historii dla większej ilości informacji i przykładów.

Wzór Historii wyszukiwań w CoreNucleotide do łączenia wyszukiwanych terminów



Udoskonalanie Twoich wyszukiwań

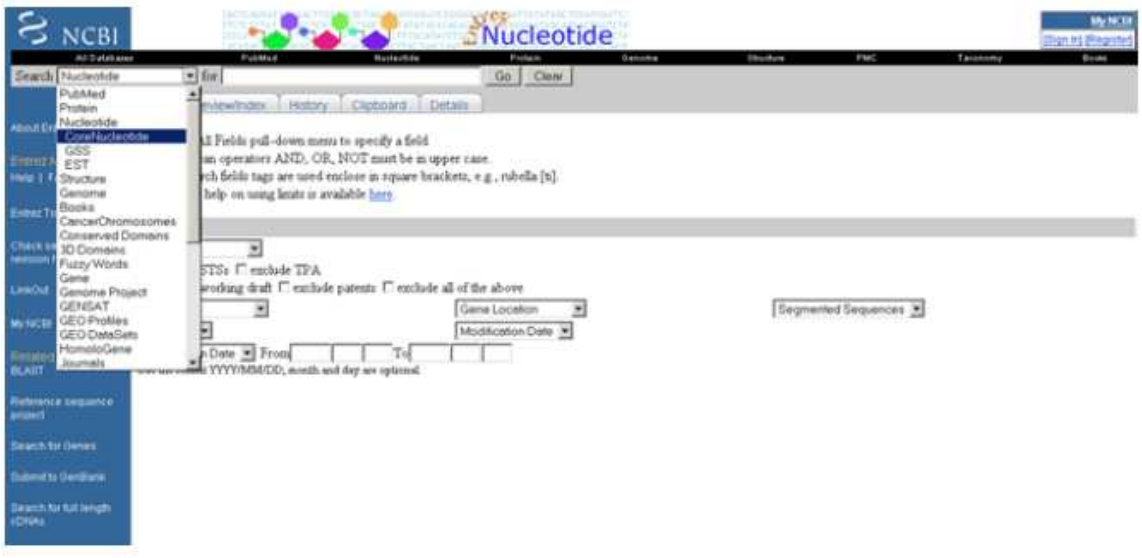
Czasami jest to niezbędne, aby udoskonalić(oczyścić) bilans wyszukiwań przez użycie Limitów, Podglądu/Indeksu i opcje historii danej bazy danych Entrez. Przejrzyj pola wyszukiwań i operatory Boolean w celu efektywnego użycia Entrez opcji limitów oraz podglądu/indeksu.

Limity

Limity umożliwiają ograniczenia poszukiwań do określonego podzbioru bazy danych. Limity mogą być ustanowione, aby ograniczyć wyszukiwanie do konkretnego pola baza danych (np. pola autora). Limity mogą być ustanowione tak by wyszukiwać wszystko oprócz konkretnego rodzaju danych (np. z wyjątkiem opatentowanych rekordów).Zamiennie limity mogą być ustawione tak, aby wyszukiwały tylko konkretnych rodzajów danych (np. DNA/RNA genomowy) lub aby wyszukiwały dane z określonego źródła bazy danych (np. EMBL). Limity danych oraz kolejnych długości są również możliwe.

Zawartość każdej bazy danych Entrez się różni, dlatego też limity dostępne dla każdej bazy danych również się różnią. Zobacz „Limity dostępne przez podsumowanie bazy danych” w sekcji podsumowań macierzy (zobacz Tabela 1, Tabela 2, Tabela 3, Tabela 4) tego wprowadzenia. Zobacz również dział Używanie Limitów tego dokumentu, aby pomogło Ci to w używaniu limitów w twoich poszukiwaniach.

Limity dostępne dla baz danych Component Nucleotide i CoreNucleotide



Używanie limitów

Limity są używane, aby udoskonalić (oczyścić) rezultaty wyszukiwań, aby otrzymać jedynie najodpowiedniejsze dokumenty. Innymi słowy limity usuwają niepotrzebne lub niechciane dokumenty. Ta sekcja dostarcza przykłady limitów w celu:

- Ograniczenie wyszukiwania do konkretnego pola bazy danych.
- Wykluczenie oczywistych rodzajów następstw
- Ograniczenie wyszukiwań do konkretnego typu cząsteczek
- Ograniczenie wyszukiwań do konkretnej lokalizacji genu
- Wyświetl tylko oryginał lub wyłącznie części połączonych zbiorów sekwencji (ciągów)
- Ogranicz poszukiwania do rekordów z konkretnej sekwencji (ciągu) bazy danych
- Ogranicz poszukiwanie przez dane
- Używanie więcej niż jednego limitu na raz

Zobacz podsumowanie macierze tabel: Tabela 1, Tabela 2, Tabela, 3 aby przejrzeć limity dostępne dla każdej bazy danych.

Ogranicz wyszukiwanie do konkretnej bazy danych.

Przykład: Jesteś zainteresowany jedynie sekwencją nukleotydową myszy:

1. Zacznij od strony głównej NCBI www.ncbi.nlm.nih.gov, ,,Nucleotide database'' z rozwijanego menu i wybierz przycisk Go aby dostać się do bazy danych nukleotydów.
2. Wybierz następnie ,,entitled Limits''
3. W sekcji ,,limited to'' wybierz ,,Organism'' z rozwijanego menu.
4. Wpisz ,,mouse'' bez cytatów w polu zapytania i wybierz następnie przycisk Go.

Na ekranie wyników, zauważ że pole wyboru obok ,,Limits'' jest zaznaczone, wskazujący, że ,, Limits'' są dobrane i aktywne. Pod polem wyboru selektywne i aktywne ,, limits'' są wyróżnione kolorem żółtym.(zakres: Organism). Możesz również nawiązać do wyników w CoreNucleotide, EST, czy GSS.

Zauważ: Pamiętaj zawsze odznaczyć pole wyboru w ,,limits'' po ukończeniu wyszukiwania, ograniczenia które wybrałeś pozostaną aktywne w następnych wyszukiwaniach.

Przykład: Jesteś zainteresowany sekwencją proteinową, która ma mniej niż 50 aminokwasów długości.

1. Zacznij od strony głównej NCBI www.ncbi.nlm.nih.gov, wybierz ,,Protein database'' z rozwijanego menu i wybierz przycisk Go aby dostać się do proteinowej bazy danych.

2. Wybierz ,,Limits''

3. W sekcji , limited to'' wybierz ,,Sequence Length'' z rozwijanego menu.

4. Wpisz "0:50'' bez cytatów w polu zapytania i wybierz następnie przycisk Go.

Na ekranie wyników, zauważ że pole wyboru obok ,,Limits'' jest zaznaczone, wskazujący, że ,, Limits'' są dobrane i aktywne. Pod polem wyboru selektywne i aktywne ograniczenia są wyróżnione kolorem żółtym.(zakres: Sequence Length).

Zauważ: Pamiętaj zawsze odznaczyć pole wyboru w ,,limits'' po ukończeniu wyszukiwania, ograniczenia, które wybrałeś pozostaną aktywne w następnych wyszukiwaniach.

Eliminowanie określonych rodzajów sekwencji

Przykład: Jesteś zainteresowany nośnikiem mitochondrialnym, jednak nie chcesz żadnych opatentowanych sekwencji:

1. Wybierz ,,CoreNucleotide database'' z rozwijanego menu.
2. Wpisz "mitochondrial carrier" bez cytatów w polu zapytania.
3. W sekcji „limited to" zaznacz pole obok "exclude Patents'' i wciśnij przycisk ,,Go''.

Na ekranie wyników, zauważ że pole wyboru obok ,,Limits'' jest zaznaczone, wskazujący, że ograniczenia są dobrane i aktywne. Pod polem wyboru selektywne i aktywne ograniczenia są wyróżnione kolorem żółtym.(ograniczenia: ,,exclude Patents'').

W nukleotydowej bazie danych możesz wyeliminować EST, STS, GSS, robocze zarysy, TPA i sekwencje patentowe. W bazie danych ,,CoreNucleotide'' możesz wyeliminować STS, robocze zarysy, TPA i patentowe sekwencje. W proteinowej bazie danych możesz wyeliminować TPA i sekwencje patentowe.

Ograniczanie wyszukiwania do określonego typu molekularnego.

Przykład: jesteś zainteresowany sekwencjami rybosomalnego RNA Cryptosporidium.

1. Zacznij od strony głównej NCBI www.ncbi.nlm.nih.gov, wybierz ,, Nucleotide database'' z rozwijanego menu i następnie wybierz przycisk ,,Go'' aby dostać się do strony nukleotydowej bazy danych.
2. Wpisz "cryptosporidium" bez cytatów w polu zapytania.
3. Wciśnij ,,Limits''
4. W sekcji ,,limited to'' wybierz ,,Molecule'' z rozwijanego menu oraz wpisz rRNA i wciśnij ,,Go''.

Na ekranie wyników, zauważ że pole wyboru obok ,,Limits'' jest zaznaczone, wskazujący, że ograniczenia są dobrane i aktywne. Pod polem wyboru selektywne i aktywne ograniczenia są wyróżnione kolorem żółtym.(ograniczenie: ,, rRNA'').

Możesz również nawiązać do wyników w CoreNucleotide, EST, czy GSS.

Ograniczenie wyszukiwania do określonej lokalizacji genu

Przykład: Jesteś zainteresowany genami w chloroplastach kwitnących roślin:

1. Wybierz ,,Nucleotide database'' z czarnego lub rozwijanego menu.
2. Wpisz ,, flowering plants'' bez cytatów w polu zapytania.
3. Wciśnij ,,Limits''.
4. W sekcji ,,limited to'' wybierz ,, Gene Location'' z rozwijanego menu oraz wpisz ,,Chloroplast'' i wciśnij przycisk ,,Go''.

Na ekranie wyników, zauważ że pole wyboru obok ,,Limits'' jest zaznaczone, wskazujący, że ograniczenia są dobrane i aktywne. Pod polem wyboru selektywne i aktywne ograniczenia są wyróżnione kolorem żółtym.(ograniczenie: ,, Chloroplast'').

Możesz również nawiązać do wyników w ,,CoreNucleotide''.

Pokazanie tylko głównych lub tylko części segmentów sekwencji.

Przykład: Jesteś zainteresowany zwłóknieniem torbielowatym genu transmembranowego regulatora przewodnictwa (CFTR). Wiesz, że jest kilka

segmentów sekwencji związanych z genem CFTR. Jednakże jesteś zainteresowany jedynie wyświetleniem głównych rekordów z wielu segmentów związanych z genem CFTR:

1. Wybierz ,,Nucleotide database'' z rozwijajnego menu.
2. Wpisz "cftr" bez cytatów w polu zapytania.
3. Wciśnij ,,Limits''
4. W sekcji ,,limited to'' wybierz ,, Segmented Sequences'' z rozwijanego menu oraz wpisz ,, Show only master of set'' i wciśnij przycisk ,,Go''.

Na ekranie wyników, zauważ że pole wyboru obok ,,Limits'' jest zaznaczone, wskazujący, że ograniczenia są dobrane i aktywne. Pod polem wyboru selektywne i aktywne ograniczenia są wyróżnione kolorem żółtym.(ograniczenie: ,, Show only master of set'').

Zauważ, że ta opcja nie pozwala ograniczyć źródeł dokumentów tylko do zawierających segmenty sekwencji. Pozwala to po prostu kontrolować w jaki sposób segmenty sekwencji są wyświetlane.

Ograniczenie wyszukiwania do rekordów z bazy danych określonej sekwencji

Przykład: Jesteś zainteresowany jedynie proteinowymi sekwencjami fosfatazy cysteinowej przedłożonymi do PIR:

1. Wybierz ,,Protein database'' z czarnego lub rozwijanego menu.
2. Wpisz "cysteine phosphatase" bez cytatów w polu zapytania.
3. Wciśnij ,,Limits''
4. W sekcji ,,limited to'' wybierz "Only from" z rozwijanego menu wpisz następnie PIR i wciśnij GO.

Na ekranie wyników, zauważ że pole wyboru obok ,,Limits'' jest zaznaczone, wskazujący, że ograniczenia są dobrane i aktywne. Pod polem wyboru selektywne i aktywne ograniczenia są wyróżnione kolorem żółtym.(ograniczenie: ,, PIR'').

Ograniczenie wyszukiwania według daty

Przykład: Chcesz znaleźć jakieś sekwencje nukleotydowe świni dodane do bazy danych lub uaktualnione w przeciągu ostatnich 30 dni:

1. Wybierz ,,Nucleotide database'' z czarnego lub rozwijanego menu.
2. Wpisz "pigs" bez cytatów w polu zapytania.
3. Wciśnij ,,Limits''
4. W sekcji ,,limited to'' wybierz " Modification Date " z rozwijanego menu wybierz następnie i wciśnij GO.

Na ekranie wyników, zauważ że pole wyboru obok ,,Limits'' jest zaznaczone, wskazujący, że ograniczenia są dobrane i aktywne. Pod polem wyboru selektywne i aktywne ograniczenia są wyróżnione kolorem żółtym.(zakres: ,,Organism''; ograniczenie: ,,30days'').

Przykład: Chcesz odnaleźć wszystkie proteinowe sekwencje myszy lub człowieka, które zostały dodane lub uaktualnione w 1997 roku:

1. Wybierz ,,Protein database'' z czarnego lub rozwijanego menu.
2. Wciśnij ,,Limits''
3. Wpisz "mouse OR human" bez cytatów w polu zapytania.
4. Wciśnij ,,Limits''
5. W sekcji ,,limited to'' wybierz " Organism" z rozwijanego menu.
6. W sekcji ,,limited to'' wybierz " Modification Date" z rozwijanego menu a następnie wybierz „Modification Date". W polach daty, wpisz datę w następujący sposób: RRRR/MM/DD. Możesz przemieszczać z pola do pola w obrębie obszaru daty. Od daty 1997/01/01 do daty 1997/12/31. Wybierz GO.

Na ekranie wyników, zauważ że pole wyboru obok ,,Limits'' jest zaznaczone, wskazujący, że ograniczenia są dobrane i aktywne. Pod polem wyboru selektywne i aktywne ograniczenia są wyróżnione kolorem żółtym.(zakres: ,,Organism''; ograniczenie: ,, Modification Date, from 1997/01/01 to 1997/12/31'').

Używanie Więcej Niż Jednego Limitu Jednocześnie

Jak pokazano w dwóch ostatnich przykładach, możesz stosować więcej niż jeden limit naraz. Oto jeszcze jeden przykład używania wielokrotnych limitów w wyszukiwarce Entrez.

Przykład: Chcesz wyszukać translację białek ludzkich sekwencji nukleotydowych GenBank'u dodaną do proteinowej bazy danych (lub zaktualizowaną) w ciągu ostatnich 30 dni. Nie chcesz rekordów patent:

1. Wybierz Proteinową bazę danych z czarnego menu lub z rozwijanego menu Search.
2. Wybierz Limits.
3. Wpisz „human" bez cudzysłowie w głównym oknie zapytania.
4. Wybierz Limits.
5. W sekcji „Limited To:" wybierz Organism z rozwijanego menu.
6. W tym samym oknie, wybierz „exclude patents", wybierz GenBank z rozwijanego menu „Only from", na końcu wybierz „30 Days" z rozwijanego menu Modified in the last i przyciśnij Go.

Zauważ, że w oknie rezultatów zakładka Limits jest zaznaczona, wskazując że limity są wybrane i aktywne. Pod polem zakładek, wybrane i aktywne limity są podświetlone na żółto (np. Pole: Organism, Limity: Exclude patents, 30 Days, GenBank).

Używanie Indeksów

Indeksy są używane do przeglądania i/lub wyboru terminów którymi są opisane rekordy i/lub dane.

Ta sekcja dostarcza przykładów dla używania indeksów by:

- Sprawdzić indeksy pola Search

- Przejrzeć, Wybrać i przeszukać terminy
- Wybrać, łączyć, i wyszukać wielokrotne terminy
- Wybrać, łączyć i wyszukać wielokrotne terminy ze zbiorowych indeksów

Podgląd/ indeks

Indeksy są alfabetyczną listą terminów z dających się przeszukać baz danych. Kiedy indeksy są wyświetlane, umożliwiają przeglądanie terminów, którymi opisane są rekordy i/lub dane. Entrez pozwala Tobie nie tylko przejrzeć indeksy, możesz też wybrać terminy, by przeszukać bezpośrednio za ich pomocą.

Podobnie jak limity, indeksy dostępne dla konkretnej bazy danych są zależne od dających się przeszukać pól tej baz danych. Zobacz "Indexes Available by Database" w tabelach dokumentu summary matrices: Tabela 1, Tabela 2, Tabela 3.

Określone indeksy wybiera się z rozwijanego menu w sekcji "Add Term(s) to Query or View Index". Przeszukiwanie wykonaj się poprzez wpisanie szukanego terminu do pola zapytania i wybierz przycisk Index. Przejrzyj terminy przewijając je za pomocą przycisków Up i Down. Zobacz rozdział Using the Indexes by znaleźć pomoc jak używać indeksów.

Indeksy Wszystkich Pól Wchodzących W Skład Nukleotydowej Bazy Danych

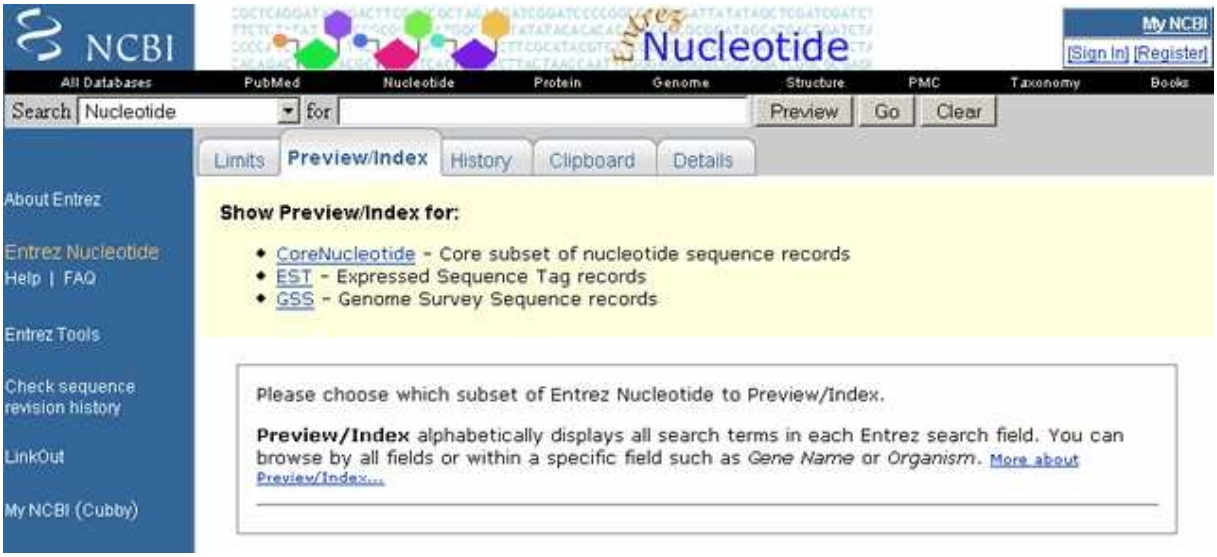
Dostępne indeksy dla nukleotydowej bazy danych są pokazane w rozwijanym menu "Add Terms to Query or View Index".

Dostępne indeksy dla nukleotydowej bazy danych są pokazane poniżej.

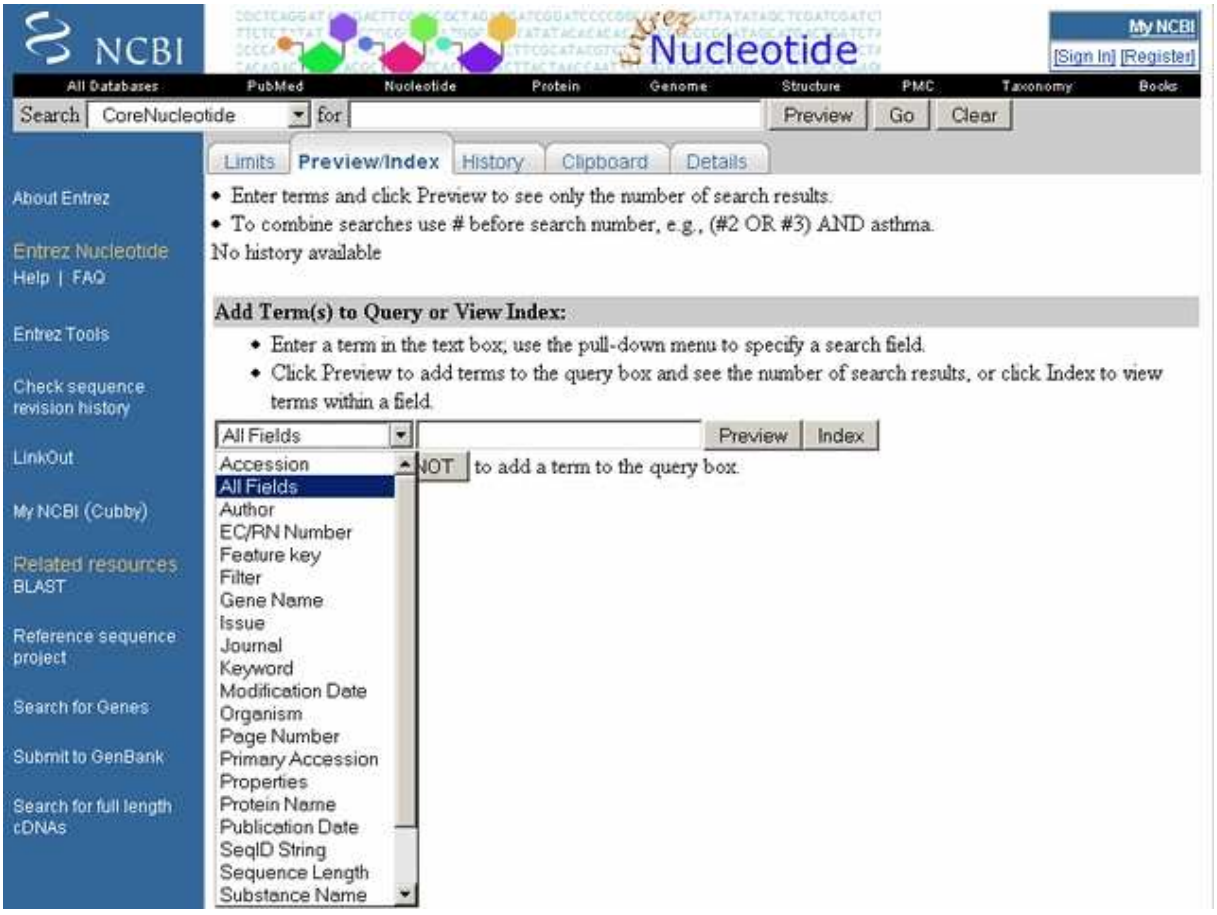
1. ze strony Nucleotide, kliknij zakładkę Preview/Index

2. dodaj termin do sekcji ‘add terms to Query or View Index’, kliknij na rozwijane menu by uzyskać listę możliwych indeksów (zobacz obraz poniżej)

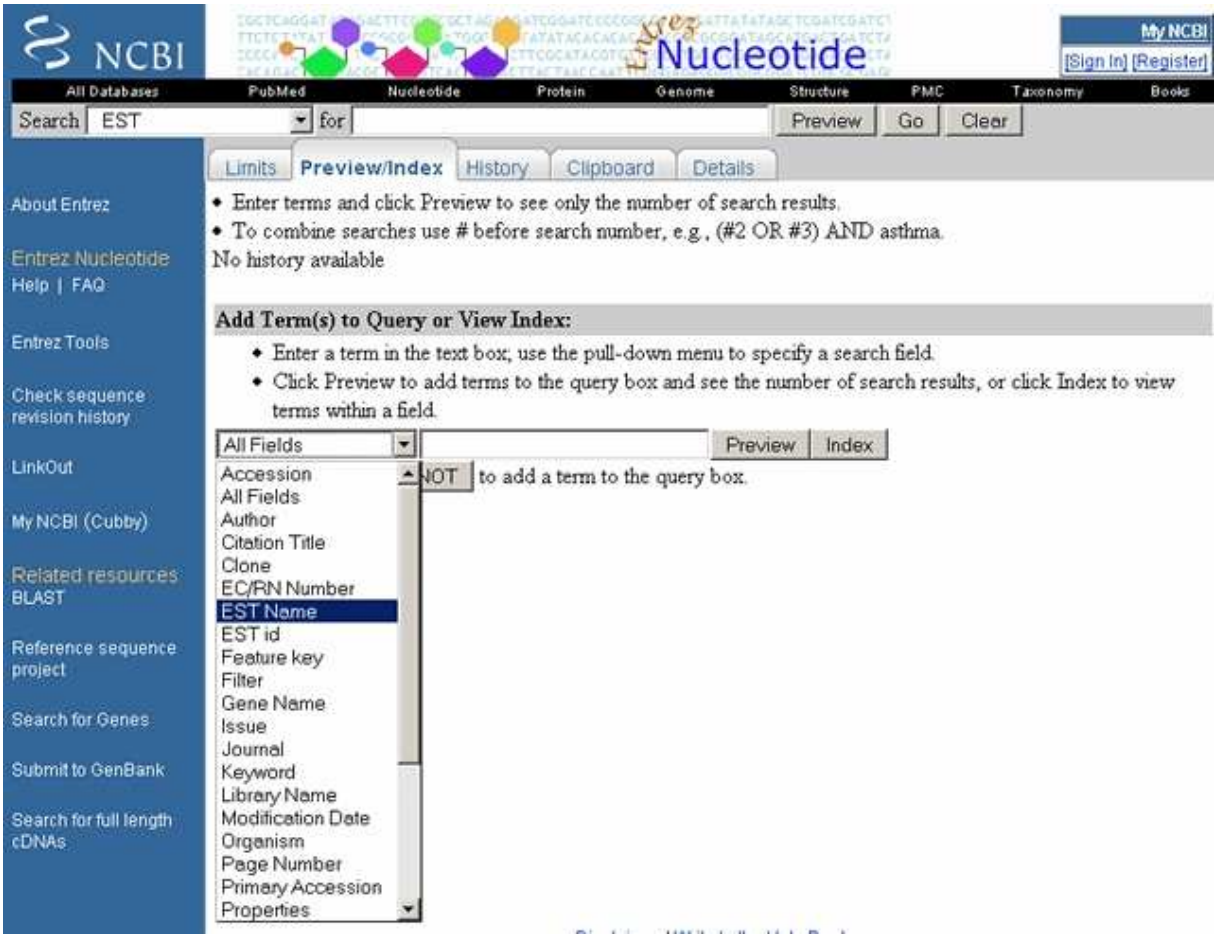
The Umbrella Nucleotide "Add Term(s) to Query or View Index" pull-down menu is shown



The CoreNucleotide "Add Term(s) to Query or View Index" pull-down menu is shown



The EST Nucleotide "Add Term(s) to Query or View Index" pull-down menu is shown



Sprawdzanie indeksów wyszukiwania

Przykład: Sprawdzenie rodzaju informacji spisanych we właściwym indeksie nukleotydowej bazy danych:

1. Wybierz nukleotydową bazę danych.
2. Wybierz Preview/Index.
3. Wybierz indeks Properties z menu rozwijanego Preview/Index.
4.Zostaw pole wyszukiwania Preview/ Index puste i wybierz przycisk "Index", aby wyświetlić alfabetyczną listę wszystkich terminów wewnątrz indeksu Properties.

Wpisy indeksu są wymienione w kolejności alfabetycznej i Entrez rozpocznie wyświetlanie indeksu przy pierwszym wejściu (tzn. Biomol genomu).

Użyj paska przewijania, aby wyświetlić więcej wpisów. Użyj przycisków dół/góra, aby wyświetlić następny zbiór zapisów w obu kierunkach. Pole wyszukiwania Properties i odpowiadający mu indeks są bardzo przydatne. To pole zawiera informacje na temat kategorii GenBank, do której należy rekord. Opisuje również rodzaj i położenie cząsteczki. Pole Properties również
opisuje takie rzeczy jak to, czy sekwencja jest częścią badanej grupy lub oddzielnego układu.

Porównaj indeks Properties nukleotydowej bazy danych do indeksu Properties innych
baz danych. Indeks Properties nie jest dostępny dla strukturalnej bazy danych.

Przykład: Sprawdzenie rodzaju informacji zapisanej w indeksie Feature key Genomowej bazy danych.

1. Wybierz Genomową bazę danych.
2. Wybierz Indeks.
3. Wybierz indeks Feature key z menu rozwijanego View Index (Przeglądu Indeksu).
4. Wpisz "0" (zero) bez cudzysłowu w polu zapytania przeglądu indeksów i wybierz Pokaż (View).

Użyj paska przewijania, aby wyświetlić wpisy. Użyj przycisków W górę i w dół, aby wyświetlić następny zestaw wpisów w obu kierunkach. Pole wyszukiwania Feature key i odpowiadający mu indeks są również bardzo przydatne. To pole zawiera informacje na temat biologicznych cech sekwencji nukleotydów dołączonych przez przesyłających i pracowników bazy.

Przeglądanie, wybieranie i szukanie haseł.

Przykład: Aby wyszukać wszystkie sekwencje w dziale EST GenBank.

Wystarczy wybrać bazę danych EST i wprowadzić wyszukiwane hasła w polu wyszukiwania.

Wybieranie, kombinacje i szukanie haseł wieloskladnikowych

Przykład: Chcesz wszystkie zestawy dla dla ludzi, myszy i Drosophila:

1. Wybierz bazę danych PopSet.
2. Wybierz Index.
3. Wybierz indeks Organism z menu rozwijanego indeksu View.
4. Wpisz "human" bez cudzysłowu w polu zapytania View Index wybierz View.
5. Zobacz listę haseł i znajdź hasło "human".
6. Wybierz hasło "human" pojedynczym kliknięciem.
7. Wybierz hasło "human" jako termin wyszukiwania, klikając przycisk "AND". Należy pamiętać, że hasło to znajduje się teraz w polu wyszukiwania jako "human" [Organism].
8. Wpisz "mouse" bez cudzysłowu w polu zapytania View Index wybierz View.
9. Zobacz listę haseł i znajdź hasło "mouse".
10. Wybierz hasło "mouse" pojedynczym kliknięciem.
11. Wybierz "mouse" jako termin wyszukiwania poprzez kliknięcie "OR". Należy pamiętać, że pojęcie to znajduje się teraz w polu wyszukiwania z hasłem „human” (tj. "human" [Organism] OR
"mouse" [Organism]).
12. Powtórz kroki 8-11 dla Drosophila tak, że końcowe szukane wyrażenie w polu zapytania to: "human" [Organism] OR "mouse" [Organism] OR Drosophila "[Organism]
13. Wybierz Go dla realizacji tego hasła.

Wybieranie, Łączenie i Szukanie Wieloskładnikowych Haseł z Wieloskładnikowych Indeksów

Przykład: Chcesz wszystkie sekwencje proteinowe kinaz pochodzących od świń:

1. Wybierz Proteinową bazę danych.
2. Wybierz Indeks.
3. Wybierz indeks Organism z menu rozwijanego Indeksu View.

4. Wpisz "pig" bez cudzysłowu w polu zapytania View Index wybierz View.
5. Zobacz listę haseł i znajdź hasło "pig".
6. Wybierz "pig" przez pojedyncze kliknięcie.
7. Wybierz "pig" jako termin wyszukiwania, klikając przycisk "AND". Należy pamiętać, że pojęcie to obecnie znajduje się w polu wyszukiwania zapytań jako "pig" [Organism].
8. Wybierz indeks Tekst Word z menu rozwijanego Indeksu View.
9. Wpisz "kinase" bez cudzysłowu w polu zapytania View Index wybierz View.
10. Zobacz listę haseł i znajdź hasło "kinase".
11. Wybierz "kinase" przez pojedyncze kliknięcie.
12. Wybierz hasło "kinase" jako termin wyszukiwania, klikając przycisk "AND". Należy pamiętać, że pojęcie to obecnie znajduje się w polu wyszukiwania zapytań jako "kinase" [Text Word], a końcowe szukane wyrażenie w polu zapytania to: "pig" [Organism] AND "kinase" [Text Word]
13. Wybierz „Go" dla realizacji tego hasła.

PAMIĘTAJ, że Entrez służy do wyszukiwania złożonych wyrażeń za pomocą Opertorów Logicznych w określonej kolejności, jak opisano powyżej w rozdziale Operatory logiczne. Zawsze można sprawdzić przycisk Details, aby zobaczyć, jak końcowe szukane wyrażenie zostało wykonane.

Korzystanie z Historii

Historia stanowi zapis wyszukiwań podczas sesji wyszukiwania. Sekcja ta dostarcza przykładów wykorzystania historii wyszukiwania.

Historia

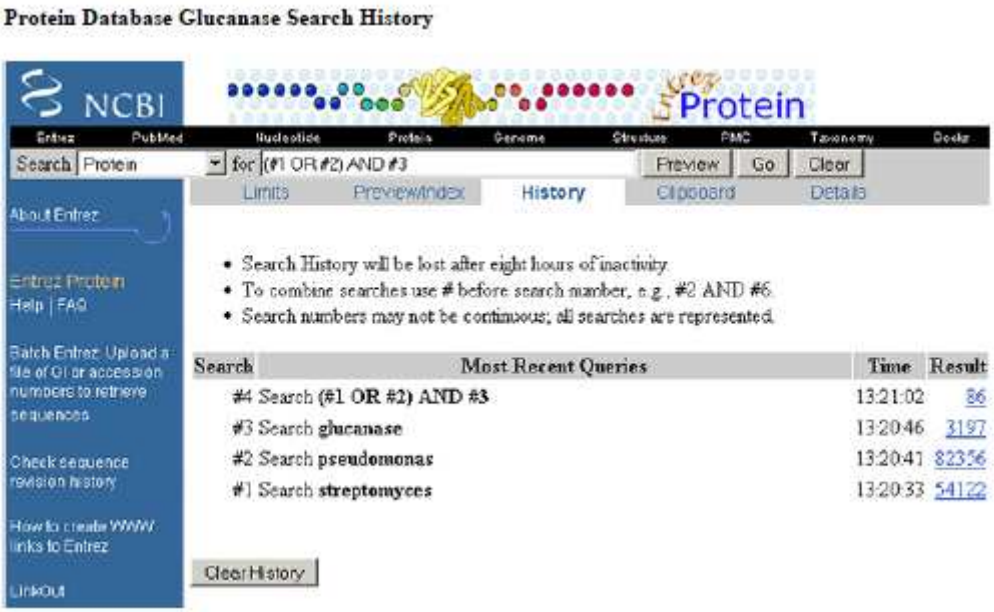
Korzystanie z opcji Preview z Preview/Index pozwala wyszukiwarce na wyświetlanie ostatnich trzech wyników dla kolejnych wyszukiwań. Wyszukiwarka może sprawdzić efekt każdego z kolejnych ograniczeń dodanych do wyszukiwania. Obserwacja objaśnienia historii w następnej sekcji dla opcji wyświetlenia całej historii wyszukiwania dla poszczególnych baz danych Enterez.

Historia stanowi zapis wyszukiwań podczas sesji wyszukiwania. Historie są specyficzne dla baz danych. Za każdym razem hasła są wpisane w oknie zapytań i wyszukiwania są realizowane, czas wyszukiwania hasła jest realizowany, a wyniki wyszukiwania ponumerowane i automatycznie zapisywane w historii dla tej bazy danych. Historia może być przypomniana, w każdym momencie trwania sesji wyszukiwania, ale historie są tracone po ośmiu godzinach bezczynności. Użyj Historii do przeglądu, weryfikacji, albo połączenia z wynikami wcześniejszych wyszukiwań. Zobacz sekcję Używanie Historii tego dokumentu, jako pomocy w użyciu historii wyszukiwania.

Sprawdzanie procesu wyszukiwania i łączenie wyników wyszukiwania.

Przykład: Wyszukaj wyniki dla Streptomyces, Pseudomonas, i glucanase, następnie użyj History do połączenia wyników wyszukiwania.

1. Wybierz proteinową bazę danych.
2. Wpisz "streptomyces" w główne okno i wybierz Go.
3. Wybierz Clear.
4. Wpisz "pseudomonas" w główne okno i wybierz Go.
5. Wybierz Clear.
6. Wpisz "glucanase" w główne okno i wybierz Go.
7. Wybierz History.
8. Sprawdź swoją historię wyszukiwania i wyniki. Należy zwrócić uwagę, że każda z wyszukanych pozycji jest ponumerowana. Zwróć też uwagę, że podany jest czas wyszukiwania oraz ilość wyników każdej z wyszukiwanych pozycji.
9. Połącz wyniki swoich wcześniejszych wyszukiwań, używając numerów swoich wyszukiwanych pozycji oraz operatorów boolowskich. Przykład: (#15 OR #16) AND #17. Wybierz Go. Zauważ: możesz kliknąć "#" w rezultatach wyszukiwania w History, a ukaże się menu 'Options'.
10. Wybierz History, aby jeszcze raz przejrzeć swoją historię wyszukiwania i wyniki.



Chociaż historie wyszukiwania są specyficzne dla danej bazy danych, to system numeracji użyty w historii wyszukiwania jest ciągły- występuje we wszystkich bazach danych podczas pojedynczej sesji wyszukiwania. Załóżmy, że właśnie zakończyłeś przeszukiwanie proteinowej bazy danych (Protein database), używając powyższego przykładu. Teraz chcesz przeszukać strukturalną bazę danych (Structure database) w tym samym celu. Nie możesz jednak użyć historii wyszukiwania z proteinowej bazy danych w bazie danych strukturalnej. Jednakże, jeśli zaczniesz przeszukiwać strukturalną bazę danych, Entrez ponumeruje kolejno przeszukiwane zbiory bazując na ostatnio wyszukiwanej kwerendzie w dowolnej bazie danych. A zatem, w tym przykładzie, pierwsza kwerenda wyszukana w strukturalnej bazie danych to pozycja #30. Następna wyszukiwana kwerenda zostanie oznaczona #31, itd. Entrez zachowa maksymalnie 100 kwerend za jednym razem.

Zauważ jeszcze, że jeżeli wyszukujesz kilkakrotnie tę samą kwerendę, w tej samej bazie danych, podczas jednej sesji wyszukiwania, to w historii wyszukiwania ta kwerenda będzie zachowana tylko raz. Aby przeprowadzić serię wyszukiwań w Entrezie, zobacz: Entrez tools.

Udoskonalenie wyników wyszukiwania

Przykład: Chcesz wyszukać jakąkolwiek sekwencję DNA mysiego antygenu fas.

1. Wybierz nukleotydową bazę danych.
2. Wpisz w główne okno mouse[orgn] AND "fas antigen" (fas antigen w cudzysłowie) i wybierz Go.

3. Wyszuka się około 30 dokumentów. Nie chcesz przeglądać wszystkich, więc decydujesz się tylko na te sekwencje, które zawierają eksony lub introny.
4. Wybierz History.
5. Udoskonal swoje wyniki wyszukiwania używając numeru wyszukiwanej kwerendy oraz operatorów boolowskich, np. #1 AND (exon OR intron). Wybierz 'Preview'.
6. Wybierz History, aby jeszcze raz przejrzeć historię wyszukiwania i wyniki. Udoskonalenie procesu wyszukiwania zredukowało ilość wyszukanych dokumentów do 9.

Tworzenie zaawansowanych formuł wyszukiwania

Złożone formuły wyszukiwania mogą być używane w głównym oknie jakiejkolwiek bazy danych i stąd bezpośrednio podlegać procesom wyszukiwania, jeżeli tylko będziemy przestrzegać kilku prostych zasad oraz będziemy używać poprawnych syntaxów.

Wykonaj wyszukiwanie określając warunki wyszukiwania, ich zakres oraz używając operatorów boolowskich. Użyj następującego syntaxu:

term [field] OPERATOR term [field],
gdzie:
term(s) to warunki wyszukiwania, field(s) to pola wyszukiwania określone w Table 1, Table 2 i Table 3, a OPERATOR to operator boolowski. Pamiętaj, że operator boolowski jest normalnie odczytywany od strony lewej do prawej.

Przykład: Znajdź wszystkie artykuły dotyczące ludzkich sekwencji nukleotydowych zawierających pętle D.

W nukleotydowej bazie danych użyj następującego wyrażenia:

D-loop[FKEY] AND human[ORGN]

Przykład: Znajdź wszystkie artykuły dotyczące ludzkich sekwencji białkowych o długości pomiędzy 50 a 60 aminokwasów, które były umieszczone w bazie danych w roku 1999.

W proteinowej bazie danych użyj następującego wyrażenia:

human[ORGN] AND 50[SLEN]:60[SLEN] AND 1999[MDAT]

Przykład: Znajdź wszystkie artykuły dotyczące *Drosophila* opublikowane w Journal of Molecular Evolution.

W bazie danych PopSet użyj następującego wyrażenia:

j mol evol[JOUR] AND drosophila[ORGN]

Przedstawienie wyników wyszukiwania i ich zachowanie

Entrez wyświetla wyniki wyszukiwania jak pokazano poniżej:

Główne okno wyszukiwania zawiera zestawienie baz danych, w których dokonywał się proces wyszukiwania oraz wprowadzone warunki wyszukiwania (np. "Search Nucleotide for hiv protease").

Klikając na 'Links' w sąsiedztwie każdego docsum, rozwija się lista baz danych, które są związane z indywidualnym podsumowaniem dokumentu. Klikając na ‘Reports’ rozwija się lista formatowanych danych. Sequence Revision History dla konkretnego rekordu może być otwierane z linku 'Revision History'.

Opcje Display, Show, Sort and 'Send to'

Display- domyślny format to Summary ukazany w przykładzie powyżej.

Aby zmienić domyślny format należy wybrać alternatywny format z menu (rozwijanej listy). Format ten automatycznie będzie się pojawiał.

Dla bardzo dużego nukleotydowego rekordu, takiego jak genom, należy odzyskać streszczenie dokumentu dla NT_037704. Nie powinieneś wyświetlać w swojej przeglądarce pełnego raportu GenBank (z właściwościami oraz sekwencją), który znajduje się w rozwijanym menu Display. Instrukcje jak ściągnąć oraz zachować plik, są podane poniżej. Musisz być ostrożny jeśli próbujesz ściągnąć bardzo duże pliki takie jak NC_000001, które posiadają wiele właściwości i 247249719 sekwencji zasad. Otwarcie tak dużego pliku wymaga solidnego edytora tekstu i odpowiednio dużo miejsca na swoim komputerze.

1. Wybierz GenBank z menu Display albo kliknij na link NT_037704.
2. Wybierz 'Send to File', a następnie wybierz opcję Cancel jeśli się pojawi w oknie zachowywania plików.
3. Wybierz Gen Bank(Full) z rozwijanego menu Display.

4. W oknie pokaże się nazwa pliku: "sequences.gbwithparts". Wybierz żądany katalog/miejsce pliku i wybierz opcje Save.
5. Pełny rekord GenBank będzie zachowany na twoim komputerze.

Aby zobaczyć "graphical view", kliknij na numer identyfikacyjny, aby wyświetlić format raportu GenBank, a następnie wybierz 'Graphics' z menu Display. Pojawi się graficzna prezentacja Entrez.

Show- domyślna liczba wyświetlanych na stronie dokumentów to 20. Łączna liczba stron jest wyświetlana z prawej strony wyników wyszukiwania.

Aby zmienić liczbę wyświetlanych na stronie dokumentów, należy wybrać alternatywną liczbę z rozwijanego menu (np. 50). Raz wybrana, będzie się wyświetlała automatycznie.

W nukleotydowej bazie danych, wyniki mogą być sortowane wg numerów identyfikacyjnych. Wybierz 'Sort by' z rozwijanego menu i wybierz opcję accession. Send to- opcja menu, która umożliwia przesyłanie wyników wyszukiwania do postaci zwykłego tekstu, pliku lub schowka.

Zobacz formaty wyświetlania w Tabeli 5. dostępne dla: Nucleotide, CoreNucleotide, EST, GSS, Protein, PopSet, Genome, and Genome Project Databases.

Wybieranie dokumentów, przedstawianie ich lub udostępnianie ich linków

Zaznacz okienka po lewej każdego wyniku streszczenia dokumentu, którego użyłeś w celu wybrania poszczególnych dokumentów z całego zbioru dokumentów wyszukanych. Raz zaznaczone, dokumenty mogą być wyświetlane (w różnych formatach), wysłane do schowka lub na dysk lokalny. Dokumenty zostaną odznaczone przez kliknięcie w okienko.

Wybieranie dokumentów używając okienek zaznaczenia

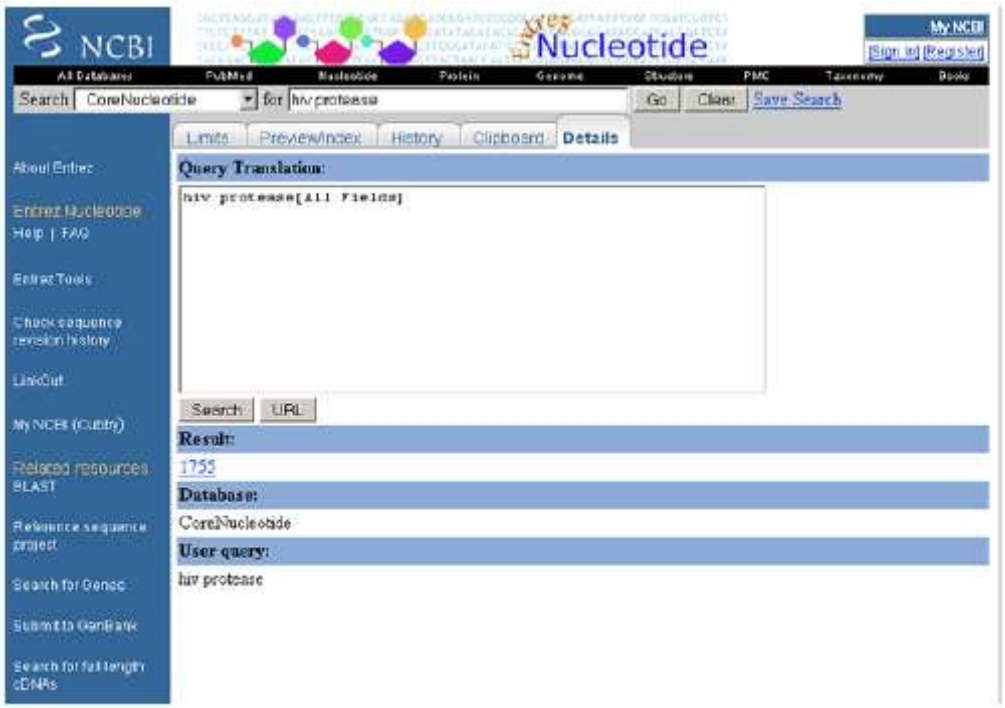
Aby **format FASTA** stał się łatwiejszy, użyto innych aplikacji, wybierz **Send to** w opcji tekstu. Opcja tekstu Send to użyje twojej przeglądarki do wyświetlenia sekwencji w formacie FASTA. Kopiuj i wklej sekwencję z przeglądarki do innej aplikacji. Widzisz również fragment, poniżej zapisywania na lokalnym dysku, o zapisywaniu większej ilości użytecznych formatach danych z Entreza.

Przycisk Szczegóły, Wyślij Do Schowka i Zapisz

Szczegóły – kliknij Szczegóły, aby wyświetlić swoją strategię szukania jako przetłumaczoną używając szukania Entrez i zasad syntax.

Okno Szczegóły zawiera także komunikat o błędzie kiedy jest przydatny. Zauważ, że informacja okna Szczegóły pokazuje szukane bazy danych, liczbę wyszukanych dokumentów (z odnośnikami do dokumentów) i twój pisemny komunikat/wyrażenie (tzn. nietłumaczony przez Entrez). W oknie Szczegóły możesz modyfikować i potwierdzać ponownie swoją strategię szukania. Potwierdź zmodyfikowaną kwerendę wciskając przycisk Szukaj.

Szczegóły



Schowek

Schowek jest tymczasowym miejscem zapisywania wyników szukania. Każda baza danych Entrez ma swój własny schowek. Szukane wyniki nie są zapisywane automatycznie. Schowek każdej bazy danych jest ograniczony do 500 pozycji, pozycje zapisane w schowku są tracone po 8 godzinach bezczynności/bierności. Pozycje mogą być wyświetlane i zapisywane ze Schowka. Zobacz Szczegóły, Wyślij Do Schowka i

Zapisz fragment dokumentu, aby pomóc w dodawaniu rekordów i używaniu ich w twoim Schowku.

Dodaj do Schowka – wybrane dokumenty 1, 3 i 5 ze zbioru wyników przez kliknięcie na okienku zaznaczenia, które sąsiaduje z numerem dokumentu. Kliknij przycisk „Schowek”. Zauważ, że te trzy pozycje zostały dodane do Schowka. Pamiętaj również, że Schowek jest ograniczony do 500 pozycji, i że te trzy pozycje zostaną utracone po 8 godzinach nieużywania w czasie pojedynczej sesji szukania. Zapamiętaj również, że numery dokumentów tych trzech pozycji (tzn. dokumentów 1, 3 i 5) są teraz zielone aby wskazać, że są one teraz w Schowku. Właściwość ta jest użyteczna, ponieważ kontynuując szukanie, jeśli te dokumenty zostały wyszukane w innej strategii szukania, numery tych dokumentów będą zielone, wskazując że już znajdują się one w Schowku.

Odświeżając dokumenty ze Schowka – wybierz przycisk Schowek. Pozycje w Schowku są wyświetlane domyślnie w formie streszczenia. Zauważ, że dokumenty mają zmienioną numerację, ale numery są zielone wskazując, że pozycje te znajdują się w Schowku. Zauważ także, że możesz wyświetlać pozycje ze Schowka we wszystkich dostępnych formatach, możesz łączyć je z sąsiednimi dokumentami lub ustalić związek w innych bazach danych. Pozycje są usuwane ze Schowka przez wybranie pozycji używając okienek zaznaczenia i wybranie opcji „Clip Remove” z menu „Wyślij do”.

Zapisywanie na lokalnym dysku – wybierz przycisk Zapisz na górze (lub na dole) w wynikach na ekranie monitora, zaraz obok przycisku „Tekst”. Dokumenty ze Schowka mogą zostać zapisane w ten sam sposób, opisany tutaj. Przed kliknięciem przycisku Zapisz, zdecyduj o dwóch rzeczach: które dokumenty chcesz i w jakim formacie. Po wybraniu dokumentów przez kliknięcie w okienkach zaznaczenia i wybraniu formatu z rozwijanego menu, wybierz przycisk „Wyświetl”. Po wyświetleniu wszystkich wybierz przycisk Zapisz. Będziesz musiał podać nazwę dokumentu, w którym wyniki zostaną zapisane na twoim lokalnym dysku. Jeśli nie wybierzesz określonych dokumentów, wszystkie dokumenty z puli wyników zostaną zapisane. W przykładzie poniżej, dokumenty 2, 3, 4, 6 i 9 zostaną zapisane na dysku w formacie FASTA. Jeśli

dokumenty te nie zostały wybrane, wszystkie 30 dokumentów (tzn. cała wyszukana pula) zostaną zapisane na dysku w formacie FASTA.

Drukowanie – użyj funkcji drukowania w twojej przeglądarce sieci Web. Jeśli życzysz sobie zapisać na dysku lokalnym, przed drukowaniem, zdecyduj o dwóch rzeczach: które dokumenty chcesz wydrukować i w jakim formacie. Ponieważ używasz funkcji drukowania przeglądarki sieci Web, możesz drukować tylko te dokumenty, które są wyświetlane. Dlatego rozważ zwiększenie liczby dokumentów wyświetlanych na jednej stronie, ponieważ całkowita liczba dokumentów, które chcesz wydrukować będzie wyświetlona na jednej stronie. Wskazówka drukowania: aby zaoszczędzić papier, przed drukowaniem rozważ użycie przycisku Tekst lub Zapisz. Zrobienie tego wyeliminuje wszystkie ale obecne dane, które potrzebujesz (tzn. Entrez search interface, pasek menu), jeśli użyjesz przycisku Tekst drukowanie będzie z twojej przeglądarki sieci Web. Jeśli użyjesz przycisku Zapisz drukowanie będzie z innych aplikacji na twojej maszynie.

LinkOut

LinkOut jest narzędziem, które zapewnia połączenie z wyników Entreza do zasobów NCBI, takich jak UniGene i LocusLink, i do zasobów zewnętrznych takich jak artykuły pełno tekstowe, dane biologiczne i centra sekwencji. Te inne zasoby dostarczają URL, nazw zasobów i krótki opis ich stron Web, których PubMed używa do stworzenia linków ich stron. Rejestracja użytkowników, opłata abonamentowa czy inne typy opłat mogą być wymagane do dostępu do pełno tekstowych artykułów w niektórych czasopismach używających tych artykułów. Informacje dla deweloperów są dostępne na LinkOut.

Narzędzia Entrez dla Power User

Narzędzia programowe Entrez - The Entrez Programming Utilities (E-Utils) są zbiorem ośmiu programów serwerowych, które zapewniają stały interfejs z kwerendą Entrez i systemowymi bazami danych. E-Utilities używa stałego syntaksu URL, który tłumaczy standardowy zbiór parametrów wejściowych koniecznych do oszacowania dla różnych elementów oprogramowania NCBI w szukaniu i odświeżaniu danych.

Dlatego E-Utilities jest strukturalnym interfejsem dla baz danych systemu Entrez. W celu dalszej nauki o E-Utilities zobacz Building Customized Data Pipelines Using the Entrez Programming Utilities (eUtils) lub rozważ wzięcie udziału w kursie NCBI PowerScripting.

Tabela 1. Limity udostępniane przez bazę danych

Bazy danych								
Limity	**Nukleotydowa**	**Core Nucleotide**	**EST**	**GSS**	**Proteinowa**	**Genomowa**	**Strukturalna**	**PopSet**
Szuka pliki	Tak	Tak	Tak	Tak	Tak	Tak	Tak	Tak
Wyklucza EST	Tak	Nie	Nie	Nie	Nie	Nie	Nie	Nie
Wyklucza STS	Tak	Tak	Nie	Nie	Nie	Nie	Nie	Nie
Wyklucza GSS	Tak	Nie	Nie	Nie	Nie	Nie	Nie	Nie
Wyklucza szkice robocze	Tak	Tak	Nie	Nie	Nie	Nie	Nie	Nie
Wyklucza patenty	Tak	Tak	Nie	Nie	Tak	Nie	Nie	Nie
Typ molekularny	Tak	Tak	Tak	Tak	Nie	Nie	Nie	Nie
Lokalizacja genu	Tak	Tak	Nie	Nie	Tak	Nie	Nie	Nie
Sekwencje segmentowe	Tak	Tak	Nie	Nie	Tak	Nie	Nie	Nie
Źródłowa baza danych	Tak	Tak	Tak	Tak	Tak	Nie	Nie	Nie
Dane zmodyfikowane	Tak	Tak	Tak	Tak	Tak	Nie	Nie	Nie

Tabela 2. Dostępny zakres przeszukiwań baz danych

	Baza danych

Opis pola „Szukaj” i kwalifikator	Nukleotydowa	Proteinowa	Genomowa	Strukturalna	PopSet
Dostęp	Tak	Tak	Tak	Tak	Tak
Wszystkie pola	Tak	Tak	Tak	Tak	Tak
Nazwisko autora	Tak	Tak	Tak	Tak	Tak
Numer EC/RN	Tak	Tak	Tak	Tak	Tak
Klawisz funkcyjny	Tak	Nie	Tak	Nie	Tak
Filtr	Tak	Tak	Tak	Tak	Tak
Nazwa genu	Tak	Tak	Tak	Nie	Tak
Numer czasopisma	Tak	Tak	Tak	Tak	Tak
Nazwa czasopisma	Tak	Tak	Tak	Tak	Tak
Słowo kluczowe	Tak	Tak	Tak	Nie	Tak
Data modyfikacji	Tak	Tak	Tak	Tak	Tak
Masa molekularna	Nie	Tak	Nie	Nie	Nie
Organizm	Tak	Tak	Tak	Tak	Tak
Numer strony	Tak	Tak	Tak	Tak	Tak
Podstawowy dostęp	Tak	Tak	Tak	Nie	Tak
Właściwości	Tak	Tak	Tak	Nie	Tak
Nazwa białka	Tak	Tak	Tak	Nie	Tak
Data publikacji	Tak	Tak	Tak	Tak	Tak
Identyfikacja sekwencji	Tak	Tak	Tak	Nie	Tak
Długość sekwencji	Tak	Tak	Tak	Nie	Nie
Nazwa	Tak	Tak	Nie	Tak	Nie

substancji					
Słowo w tekście	Tak	Tak	Tak	Tak	Tak
Słowo w tytule	Tak	Tak	Tak	Nie	Nie
Identyfikacja użytkownika	Nie	Nie	Nie	Nie	Nie
Tom	Tak	Tak	Tak	Tak	Tak

Tabela 3. Definicje i kwalifikatory indeksów pola „Szukaj”.

Indeks pola „Szukaj”	**Definicja**	**Kwalifikator**
Dostęp	Zawiera unikalny numer dostępu sekwencji lub rekordu, przypisany do nukleotydu, białka, struktury lub genomu. Indeks dostępu Strukturalnej bazy danych zawiera PDB ID ale nie MMDB ID, na przykład: AF123456[accn].	[ACCN] lub [ACCESSION]
Wszystkie pola	Obejmuje wszystkie dostępne w bazie danych indeksy przeszukiwania.	[ALL] lub [ALL FIELDS]
Nazwisko autora	Zawiera nazwiska wszystkich autorów ze wszystkich publikacji w bazie danych rekordów. Format to: nazwisko odstęp pierwsza litera imienia (imion), bez znaków przestankowych (np. marley jf)	[AUTH] lub [AUTHOR]
Numer EC/RN	Numer przypisany przez Enzyme Commission lub Chemical Abstract Service (CAS) oznaczający konkretny enzym lub substancję chemiczną.	[ECNO]
Klawisz funkcyjny	Obejmuje biologiczne cechy przypisane sekwencji nukleotydowej i zdefiniowane w DDBJ/EMBL/GenBank Feature Table (http:// www.ncbi.nlm.nih.gov/projects/collab/FT/index.html). Niedostępny w Proteinowej i Strukturalnej bazie danych.	[FKEY]
Filtr	Obejmuje ustalony lub przefiltrowany podzbiór różnych baz danych. Podzbiory te powstają poprzez grupowanie rekordów, które są powszechnie połączone z innymi bazami danych Entrez lub z tą samą bazą danych. Na przykład indeks bazy danych PopSet zawiera PopSet all, PopSet medline, PopSet nucleotide, PopSet protein. Filtr PopSet medline zawiera wszystkie rekordy bazy PopSet z linkami do bazy PubMed; filtr PopSet nuclotide zawieraz wszystkie rekordy bazy PopSet z linkami do bazy Nukleotydowej.	[FILT] lub [SB]

Nazwa genu	Zawiera standardową i pospolitą nazwę genu znalezionego w bazie danych rekordów. Pole jest niedostępne w strukturalnej bazie danych.	[GENE]
Numer czasopisma	Oznacza numer czasopisma, w którym opublikowane dane.	[ISS] lub [ISSUE]
Słowo kluczowe	Specjalny indeks obejmujący słownik powiązany z bazami danych GenBank, EMBL, DDBJ, SWISS-Prot, PIR, PRF, lub PDB. Przeszukuj bazy danych za pomocą tego indeksu, aby zapoznać się ze słownictwem. Indeks niedostępny w Strukturalnej bazie danych.	[KYWD] lub [KEYWORD]
Nazwa czasopisma	Obejmuje nazwę czasopisma, w którym opublikowano dane. Nazwy czasopism są zindeksowane w skróconej formie (np. J Biol Chem). Czasopisma są indeksowane również poprzez ich numery ISSN. Przeglądaj indeks jeśli nie znasz numeru ISSN lub nie jesteś pewien jaki skrót jest przypisany do konkretnego czasopisma.	[JOUR] lub [JOURNAL]
Data modyfikacji	Zawiera datę ostatniej modyfikacji rekordu w bazie Entrez, w formacie RRRR/MM/DD (np. 1999/08/05). Sam rok (np. 1999) obejmuje wszystkie rekordy dla danego roku, rok i miesiąc obejmuje rekordy modyfikowane w danym miesiącu zindeksowane w bazie Entrez.	[MDAT]
Organizm	Obejmuje naukową i popularną nazwę organizmu związanego z sekwencją białkową bądź nukleotydową.	[ORGN] lub [ORGANISM]
Numer strony	Zawiera numer strony czasopisma, na której zostały opublikowane dane.	[PAGE]
Podstawowy dostęp	Obejmuje główny numer dostępu sekwencji lub rekordu, przypisany do nukleotydu, białka lub struktury. Indeks nie jest dostępny w Strukturalnej bazie danych.	[PACC]
Właściwości	Zawiera właściwości sekwencji nukleotydowych lub białkowych. Na przykład indeks Właściwości Nukleotydowej bazy danych obejmuje typ molekuły, status publikacji, lokalizację cząsteczki, i dział bazy GenBank. Indeks niedostępny w Strukturalnej bazie danych.	[PROP]
Nazwa białka	Obejmuje standardową nazwę białka wyszukaną w bazie rekordów. Popularne nazwy mogą nie być indeksowane w tym polu, więc warto rozważyć przeszukiwanie za pomocą indeksu „Wszystkie pola” lub „Słowo w tekście”. Indeks niedostępny w Strukturalnej bazie danych.	[PROT]
Data publikacji	Zawiera datę publikacji danych w bazie Entrez, w formacie RRRR/MM/DD (np. 1999/08/05). Data pierwszego pojawienia się w bazie GenBank, zindeksowanie w bazie Entrez. Sam rok (np. 1999) obejmuje wszystkie rekordy dla danego roku, rok i miesiąc obejmuje rekordy opublikowane w bazie danych GenBank w danym miesiącu.	[PDAT]
Identyfikacja sekwencji	Obejmuje specjalny identyfikator łańcucha, podobny do identyfikacji FASTA. Indeks niedostępny w Strukturalnej bazie danych.	[SQID]

Długość sekwencji	Zawiera całkowitą długość sekwencji. Indeks „Długość sekwencji” nie jest dostępny w Genomowej bazie danych i w bazie danych PopSet.	[SLEN]
Nazwa substancji	Zwiera nazwy substancji chemicznych związanych z rekordem, z rejestru CAS i MEDLINE. Indeks „Nazwa substancji” nie jest dostępny w Genomowej bazie danych i w bazie danych PopSet.	[SUBS]
Słowo w tekście	Zawiera „wolny tekst” związany z rekordem.	[WORD]
Tytuł	Zawiera tylko te słowa zdefiniowane jako rekord. Definicja podsumowuje biologię sekwencji i jest dokładnie konstruowana przez personel tworzący bazę danych. Standardowa definicja będzie opisywała organizm, nazwę produktu, symbol genu oraz typ molekuły. Indeks ten nie jest dostępny w Strukturalnej bazie danych i w bazie danych PopSet.	[TITL]
Tom	Zawiera numer tomu czasopisma, w którym zostały opublikowane dane.	[VOL]

Tabela 4 Definicje i kwalifikatory indeksów pola „szukaj” proteinowych baz danych

Proteinowa baza danych		
Indeks pola „szukaj”	**Definicja**	**Kwalifikator**
Dostęp	Zawiera unikalny numer dostępu sekwencji lub rekordu, przypisany do nukleotydu, białka, struktury lub genomu. Indeks dostępu Strukturalnej bazy danych zawiera PDB ID ale nie MMDB ID, na przykład: AF123456[accn].	[ACCN] lub [ACCESSION]
Wszystkie pola	Obejmuje wszystkie dostępne w bazie danych indeksy przeszukiwania.	[ALL] lub [ALL FIELDS]
Nazwisko autora	Zawiera nazwiska wszystkich autorów ze wszystkich publikacji w bazie danych rekordów. Format to: nazwisko odstęp pierwsza litera imienia (imion), bez znaków przestankowych (np. marley jf).	[AUTH] lub [AUTHOR]
Numer EC/RN	Numer przypisany przez Enzyme Commission lub Chemical Abstract Service (CAS) oznaczający konkretny enzym lub substancję chemiczną.	[ECNO]
Filtr	Obejmuje ustalony lub przefiltrowany podzbiór różnych baz danych. Podzbiory te powstają poprzez grupowanie rekordów, które są powszechnie połączone z innymi bazami danych Entrez lub z tą samą bazą danych. Na przykład indeks bazy danych PopSet zawiera PopSet all, PopSet medline, PopSet nucleotide, PopSet protein. Filtr PopSet medline zawiera wszystkie rekordy bazy PopSet z linkami do bazy PubMed; filtr PopSet nuclotide	[FILT] lub [SB]

	zawieraz wszystkie rekordy bazy PopSet z linkami do bazy Nukleotydowej.	
Nazwa genu	Zawiera standardową i pospolitą nazwę genu znalezionego w bazie danych rekordów. Pole jest niedostępne w strukturalnej bazie danych.	[GENE]
Numer czasopisma	Oznacza numer czasopisma w którym opublikowane dane.	[ISS] lub [ISSUE]
Słowo kluczowe	Specjalny indeks obejmujący słownik powiązany z bazami danych GenBank, EMBL, DDBJ, SWISS-Prot, PIR, PRF, lub PDB. Przeszukuj bazy danych za pomocą tego indeksu, aby zapoznać się ze słownictwem. Indeks niedostępny w Strukturalnej bazie danych.	[KYWD] lub [KEYWORD]
Czasopismo	Obejmuje nazwę czasopisma, w którym opublikowano dane. Nazwy czasopism są zindeksowane w skróconej formie (np. J Biol Chem). Czasopisma są indeksowane również poprzez ich numery ISSN. Przeglądaj indeks jeśli nie znasz numeru ISSN lub nie jesteś pewien jaki skrót jest przypisany do konkretnego czasopisma.	[JOUR] lub [JOURNAL]
Data modyfikacji	Zawiera datę ostatniej modyfikacji rekordu w bazie Entrez, w formacie RRRR/MM/DD (np. 1999/08/05). Sam rok (np. 1999) obejmuje wszystkie rekordy dla danego roku, rok i miesiąc obejmuje rekordy modyfikowane w danym miesiącu zindeksowane w bazie Entrez.	[MDAT]
Masa molekularna	Masa molekularna białka, wyrażona w daltonach (Da), obliczona w sposób opisany w dokumencie Entrez pomoc, w części dotyczącej masy molekularnej. Zauważ, że masa molekularna musi być wrażona jako stałe 6 cyfr, zaczynające się od zera (nie litery „O”), np., 002002 [MOLWT] .	[MOLWT]
Organizm	Obejmuje naukową i popularną nazwę organizmu związanego z sekwencją białkową bądź nukleotydową.	[ORGN] lub [ORGANISM]
Numer strony	Zawiera numer strony czasopisma, na której zostały opublikowane dane.	[PAGE]
Podstawowy dostęp	Obejmuje główny numer dostępu sekwencji lub rekordu, przypisany do nukleotydu, białka lub struktury. Indeks nie jest dostępny w Strukturalnej bazie danych.	[PACC]
Właściwości	Zawiera właściwości sekwencji nukleotydowych lub białkowych. Na przykład indeks Właściwości Nukleotydowej bazy danych obejmuje typ molekuły, status publikacji, lokalizację cząsteczki, i dział bazy GenBank. Indeks niedostępny w Strukturalnej bazie danych.	[PROP]
Nazwa białka	Obejmuje standardową nazwę białka wyszukaną w bazie rekordów. Popularne nazwy mogą nie być indeksowane w tym polu, więc warto rozważyć przeszukiwanie za pomocą indeksu „Wszystkie pola” lub „Słowo w tekście”. Indeks niedostępny w	[PROT]

	Strukturalnej bazie danych.	
Data publikacji	Zawiera datę publikacji danych w bazie Entrez, w formacie RRRR/MM/DD (np. 1999/08/05). Data pierwszego pojawienia się w bazie GenBank, zindeksowanie w bazie Entrez. Sam rok (np. 1999) obejmuje wszystkie rekordy dla danego roku, rok i miesiąc obejmuje rekordy opublikowane w bazie danych GenBank w danym miesiącu.	[PDAT]
Identyfikacja sekwencji	Obejmuje specjalny identyfikator łańcucha, podobny do identyfikacji FASTA. Indeks niedostępny w Strukturalnej bazie danych.	[SQID]
Długość sekwencji	Zawiera całkowitą długość sekwencji. Indeks „Długość sekwencji” nie jest dostępny w Genomowej bazie danych i w bazie danych PopSet.	[SLEN]
Nazwa substancji	Zwiera nazwy substancji chemicznych związanych z rekordem, z rejestru CAS i MEDLINE. Indeks „Nazwa substancji” nie jest dostępny w Genomowej bazie danych i w bazie danych PopSet.	[SUBS]
Słowo w tekście	Zawiera „wolny tekst” związany z rekordem.	[WORD]
Tytuł	Zawiera tylko te słowa zdefiniowane jako rekord. Definicja podsumowuje biologię sekwencji i jest dokładnie konstruowana przez personel tworzący bazę danych. Standardowa definicja będzie opisywała organizm, nazwę produktu, symbol genu oraz typ molekuły. Indeks ten nie jest dostępny w Strukturalnej bazie danych i w bazie danych PopSet.	[TITL]
Tom	Zawiera numer tomu czasopisma, w którym zostały opublikowane dane.	[VOL]

Tabela 5 Format wyświetlania dla Nukleotydowych, CoreNucleotydowch, EST, GSS, Genomowych , Projektów Genomowych i Genomowych baz danych.

Format wyświetlania	Opis	Dostępność w Baz danych
Streszczenie/Informacje podstawowe	Widok domyślny, numer dostępu w postaci aktywnego linka i krotki opis.	Nukleotydowa, Proteinowa, CoreNukleotydowa, EST, GSS, PopSet, Genomowa, Projekty Genomowe.
Pobieżne streszczenie	Numer dostępu w postaci aktywnego linka, skrócony opis, numer projektu w postaci aktywnego linka w przypadku Projektu genomu .	Nukleotydowa, Proteinowa, CoreNukleotydowa, EST, GSS, PopSet, Genomowa, Projekty Genomowe.
GenBank	Pełny raport .	Nukleotydowa, CoreNukleotydowa, EST, GSS,

		Genomowa.
GenPept	Pełny raport.	Proteinowa.
GenBank(Pełny)	Kompletny rekord GenBanku ze wszystkim funkcjami i sekwencjami. Ten format jest użyteczny w przypadku bardzo dużych rekordów GenBanku.	Nukleotydowa, CoreNucleotydowa, EST, GSS, Genomowa.
GenPept(Pełny)	Kompletny rekord GenPept ze wszystkimi funkcjami białek i wszystkimi sekwencjami. Ten format jest użyteczny w przypadku bardzo dużych rekordów GenBanku.	Proteinowa.
INSDSeq XML	XML DTD dla rekordów sekwencji.	Nukleotydy, Proteinowa.
Lista GI	Lista GenInfo - - Identyfikatorów GI.	Nukleotydowa, CoreNukleotydowa, EST, GSS, Proteinowa.
ASN.1	Abstrakt Syntax Notation One, używane do przechowywania i pobierania danych oraz pomocy w osiąganiu interoperacyjności pomiędzy platformami.	Nukleotydowa, Proteinowa, CoreNukleotydowa, EST, GSS, PopSet, Genomowa.
EST	Podstawowy format wyświetlania dla rekordów Exspressed Sequence Tag.	EST
FASTA	Wiersz definicji i sekwencja znaków.	Wszystkie Nukleotydowe i Proteinowe bazy danych.
Grafika lub Graph	Dostęp do graficznego obrazu sekwencji po wyborze numeru dostępu w postaci aktywnego linku.	Nukleotydowa, Proteinowa i Genomowa.
GSS	Podstawowy format wyświetlania rekordów dla badanych sekwencji genomowych.	GSS.
TinySeq XML	Uproszczony XML do analiz.	Nukleotydowa, CoreNukleotydowa, EST, GSS, Genomowa, Proteinowa.
Streszczenie PopSet	Kompletny zestaw numerów dostępu obejmujący dostęp do PopSetu po wyborze numeru dostępu PopSetu w postaci aktywnego linka.	PopSet.

Lista UI	Lista numerów identyfikacyjnych poszczególnych baz danych PopSet.	PopSet
XML	Format Script-parseable	Nukleotydowa, Proteinowa, Genowa.