生物信息学课程总结交流报告

数据库检索

报告人：　严盈

PPT制作：08F1B队全体队员

- 一、相关概念；
- 二、主要数据库介绍；
- 三、**PubMed**的使用；
- 四、**Entrez**的使用；
- 五、**SRS**的使用；
- 六、总结

一、相关概念

- **数据库检索：**指通过相关的检索系统，对序列、结构以及各种二次数据库中的注释信息进行关键词匹配查找。
- **数据库搜索：**指特定的序列相似性比对算法，找出核酸或蛋白质序列数据库中与检测序列具有一定程度相似性的序列。

二、主要生物信息数据库简介

1、Genbank

2、EBI

3、ExPASy

- 1.Genbank库包含了所有已知的核酸序列和蛋白质序列，以及与它们相关的文献著作和生物学注释。它是由美国国立生物技术信息中心(NCBI)建立和维护的，其网站为：

http://www.ncbi.nlm.nih.gov

National Center for Biotechnology Information

National Library of Medicine | National Institutes of Health

PubMed | All Databases | BLAST | OMIM | Books | TaxBrowser | Structure

Search All Databases for Go

SITE MAP
Alphabetical List
Resource Guide

About NCBI
An introduction to NCBI

GenBank
Sequence submission support and software

Literature databases
PubMed, OMIM, Books, and PubMed Central

Molecular databases
Sequences, structures, and taxonomy

Genomic biology
The human genome, whole genomes, and related resources

Tools
Data mining

Research at NCBI
People, projects, and seminars

Software engineering
Tools, R&D, and databases

Education
Teaching resources and on-line tutorials

FTP site
Download data and software

Contact information
How to reach us

What does NCBI do?

Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. More about NCBI...

Primer-BLAST

The new Primer-BLAST service designs more effective and gene-specific PCR primers. The tool combines Primer 3 primer design with a specificity check provided by a specialized BLAST search. For more information, see BLAST News.

PubMed Central

PubMed Central is an archive of biomedical and life sciences journals.

- Free full text
- Over 1,500,000 articles from over 450 journals
- Linked to PubMed and fully searchable

Use of PubMed Central requires no registration or fee. Access it from any computer with an Internet connection.

NCBI News

NCBI News available online
In this issue

- Featured Resource: Improvements to NCBI Services Promote Discovery
- New Databases and Tools
- GenBank News
- Updates and Enhancements
- Announce Lists and RSS Feeds

Hot Spots

- Clusters of orthologous groups
- Coffee Break, Genes & Disease, NCBI Handbook
- Electronic PCR
- Entrez Home
- Entrez Tools
- Gene expression omnibus (GEO)
- Human genome resources
- Influenza Virus Resource
- Map Viewer
- dbMHC
- Mouse genome resources
- My NCBI
- ORF finder
- Rat genome resources
- Reference sequence project
- Short Read Archive
- SKY/CGH database
- dbSNP
- Trace Archives
- VecScreen
- Viral Genotyping Tool
- Web (RSS) Feeds
- NCI-CGAP

National Center for Biotechnology Information
U.S. National Library of Medicine
8600 Rockville Pike, Bethesda, MD 20894

（1）、NCBI数据库的主要组成部分

NCBI主要部分	中文释义
PubMed	生物医学类文献引文和摘要
PubMed Central	免费的全文期刊文章
Site Search	NCBI网站和FTP站点
Books	在线书库
OMIM	人类孟德尔遗传数据库
OMIA	动物孟德尔遗传数据库

（2）、NCBI链接的数据库（共24个）

Nucleotide核苷酸序列 记录的核心子集	CDD 保守的蛋白质域数据库	PubChem Substance 沉积化学物质记录
EST 标签序列表达记录	3D Domains 来自Entrez结构的结构域	Protein Clusters 相关蛋白序列的集合
GSS 基因组调查序列记录	UniSTS 标记和图谱数据	Genome Project 基因组工程信息
Protein 序列数据库	PopSet 种群研究数据集	Probe 特殊序列试剂
Genome 全基因组序列	GEO DataSets GEO数据的试验系列	GENSAT 小鼠中枢神经系统的基因表达图
Structure 三维宏观分子结构	Cancer Chromosomes 细胞生成数据库	HomoloGene 真核生物的同源种群
Taxonomy GenBank 中的生物体	PubChem BioAssay 化学物质的生物活性筛选	Gene 以基因为中心的信息
dbGaP 基因型和表现型	PubChem Compound 独特的小分子化学结构	SNP 单核苷酸多态性

（3）、NCBI工具功能

1）、核酸序列分析

2）、蛋白质序列分析

3）、蛋白质结构功能分析

4）、基因组分析

5）、基因表达分析

1）、核酸序列分析工具（共10个）

BLAST	基因和蛋白质序列与其它公共数据库中的序列进行比对
Electronic PCR	识别DNA序列中的序列标签（STSs）位点
Entrez Gene	搜索基因的相关信息
Model Maker	从基因组数据库中构建mRNA序列
ORF Finder	寻找开放阅读框
SAGEmap	基因表达的一系列分析
Spidey	mRNA与单一基因组序列的比对
Splign	cDNA 与基因组的比对
VecScreen	核酸序列片段的识别
Viral Genotyping Tool	重组或非重组病毒核酸序列基因型的识别

2）、蛋白质序列分析和蛋白质组学工具

BLAST	搜素与所比对的蛋白质序列相似的蛋白质
CD Search	搜索蛋白质的保守域
CDART	保守区域特点的检索
OMSSA	高效的搜索引擎，通过搜索已知蛋白质序列文库用于鉴定**MS/MS**肽光谱
TaxPlot	完整的真核生物基因组中蛋白质的同源性

3）、蛋白质结构分析工具

(1)Cn3D

蛋白质的3D空间结构

(2)VAST Search

NCBI中的结构对结构相似性搜索

(3)CD Search

搜索保守域

4）、基因组分析工具

Map Viewer

提供了大量的基因组图谱和序列资料

5）、基因表达分析工具

(1)SAGEmap

基因表达的一系列分析

(2)CGAP

癌细胞的分子解剖学的确认

(3)UniGene DDD

在选择的cDNA文库中比较基因表达特性

- 2.EMBL （The European Molecular Biology Laboratory）核酸序列数据库由欧洲生物信息学研究所(EBI)维护的核酸序列数据构成，由于与Genbank和DDBJ的数据合作交换，它也是一个全面的核酸序列数据库。该数据库查询检索可以通过因特网上的序列提取系统(SRS)服务完成。 网址：
- http://www.ebi.ac.uk/embl/

EBI核酸数据库主要包括：

1、 Databases Home

2、TooIs

3、EBI Groups

4、Training

5、Industry

Databases Home

- Database Browsing 数据库搜索
- Biological Ontologies 生物学本体论
- Literature 参考资料
- Microarray 微阵列
- Nucleotide 核酸
- Pathways&Networks 途径和网络研究
- Protein 蛋白质
- Proteomic 蛋白质组学
- Structure 结构

EBI链接的数据库

EMBL Nucleotide Database	欧洲最重要的核酸序列数据库
UniProt Knowledgebase	完整的有注释的蛋白质序列数据库
Protein Databank in Europe Database	关于大分子结构管理和分类数据库的欧洲项目
ArrayExpress	基因表达数据库
Ensembl	该数据库有完整的最新的**metazoic**基因组和最好的可能的自动注释
IntAct	提供免费，开放的资源数据库系统和蛋白质相互作用的数据

EBI数据库的工具箱

一、工具箱作用

1、相似性和同源性

2、蛋白质的功能分析

3、序列分析

4、结构分析

5、网页搜索服务

二、EBI中的工具

1、BLAST 或 FASTA

序列相似性和同源性比对

2、InterProScan

蛋白质功能分析

3、ClustalW2

序列比对工具

4、MSDfold 或DALI

蛋白质结构分析或站点服务

5、基因表达和基因组数据的聚类、分析工具

3.ExPASy是蛋白质数据库，该数据库主要包括蛋白质氨基酸序列、结构、组成、以及蛋白质分子模型的检索。该数据库主要软件工具有蛋白质氨基酸序列、结构、组成分析工具 。网站：

http://www.expasy.ch/

ExPASy链接的数据库

UniProt Knowledgebase (Swiss-Prot and TrEMBL)	**Swiss-Prot蛋白质数据库**
ViralZone	病毒的蛋白质序列数据库
PROSITE	蛋白质家族和蛋白质区域数据库
SWISS-2DPAGE	**2-D聚丙烯酰胺凝胶电泳数据库**
MIAPEGelDB	**MIAPE凝胶电泳数据库**
ENZYME	酶的系统命名数据库
UniPathway	代谢途径数据库
SWISS-MODEL Repository	自动获得的蛋白质模型数据库

其它分子生物学数据库的链接

（1）蛋白质相关数据库

（2）蛋白质3-D结构相关数据库

（3）蛋白质组学数据库及其链接

（4）核酸及相关数据库

（5）碳水化合物资源

（6）专一物种数据库：人、脊椎动物、线粒体和叶　　绿体、昆虫、无脊椎动物、植物、真菌、细菌、古细菌、病毒和噬菌体、

（7）人的变异数据库

（8）翻译后修饰数据库

（9）基因表达数据库

……

ExPASy数据库工具

1、蛋白质组学和序列分析工具

（1）相似性搜索

（2）翻译后修饰和拓扑结构预测

（3）初级结构分析

（4）二级和三级结构工具

（5）比对和系统发生分析

2、2-D凝胶电泳分析软件

3、质谱图像分析工具

三、PubMed的使用

1.PubMed的概述
2.PubMed的简单检索技巧
3.PubMed的高级检索技巧

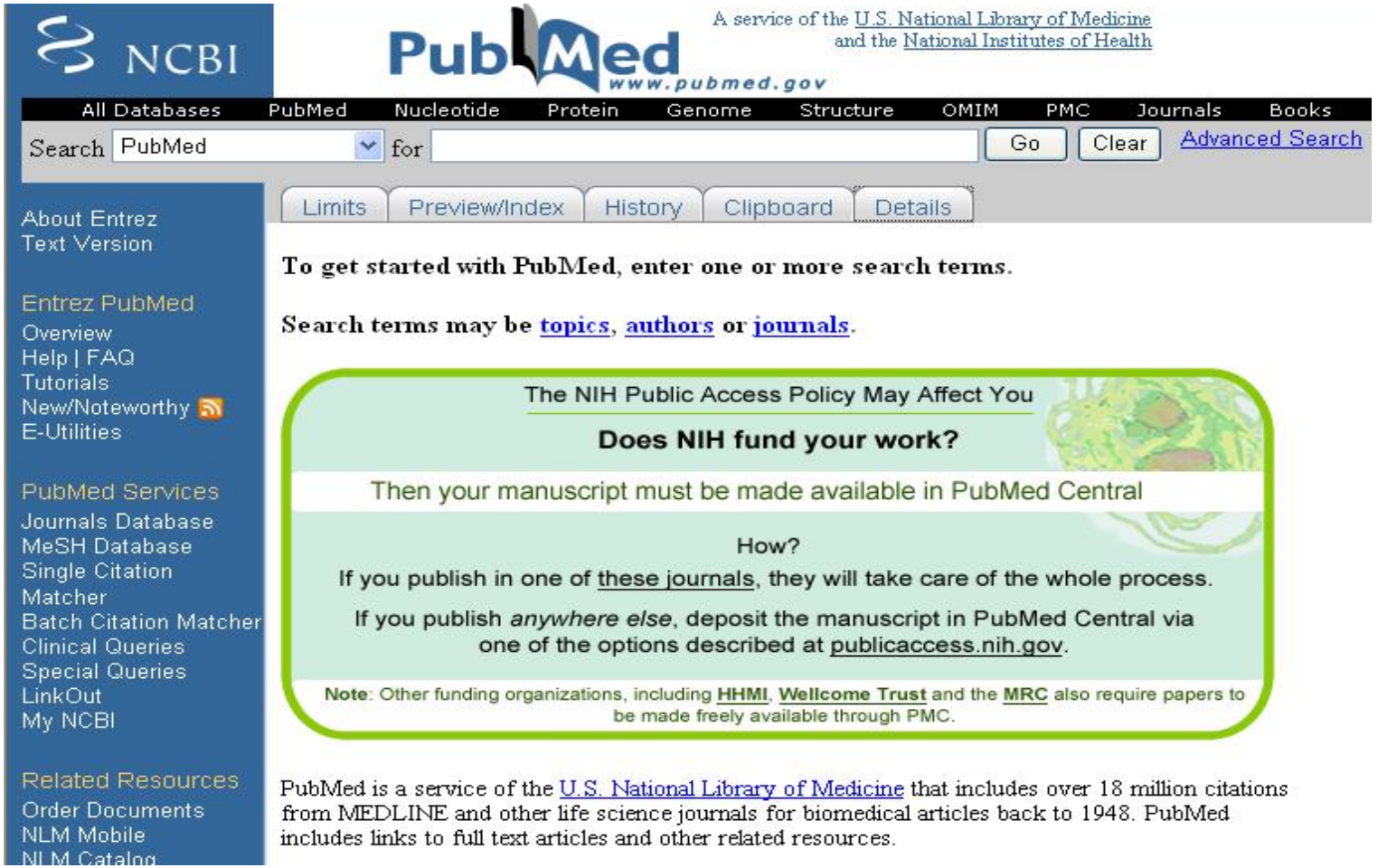
PubMed的概述

PubMed是生物医学信息检索系统，PubMed覆盖了全世界70多个国家4300多种主要生物医学期刊的摘要和部分全文。其覆盖的时间段也非常长，早的可以追溯到20世纪60年代。

PunMed的优点

与PubMed挂钩的出版商会自动向PubMed提供最新的文献摘要。尽管生物医学的文章从被期刊接受到出版往往要好几个月的时间，但借助于PubMed，我们仍旧可以**随时掌握最新的动向。**

PubMed的简单检索技巧



通过添加限定词可以精炼检索的范围，如以下几种方式：

1.作者检索

作者姓+空格+名字首字母缩写+[AU]

2.期刊名检索

在检索框中输入期刊全称或MEDLINE形式的简称、ISSN号，如“molecular biology of the cell[TA]”，或“mol biol cell[ta]”，或1059-1524

3.作者的单位检索

作者单位+[AD]

4.日期或日期范围检索

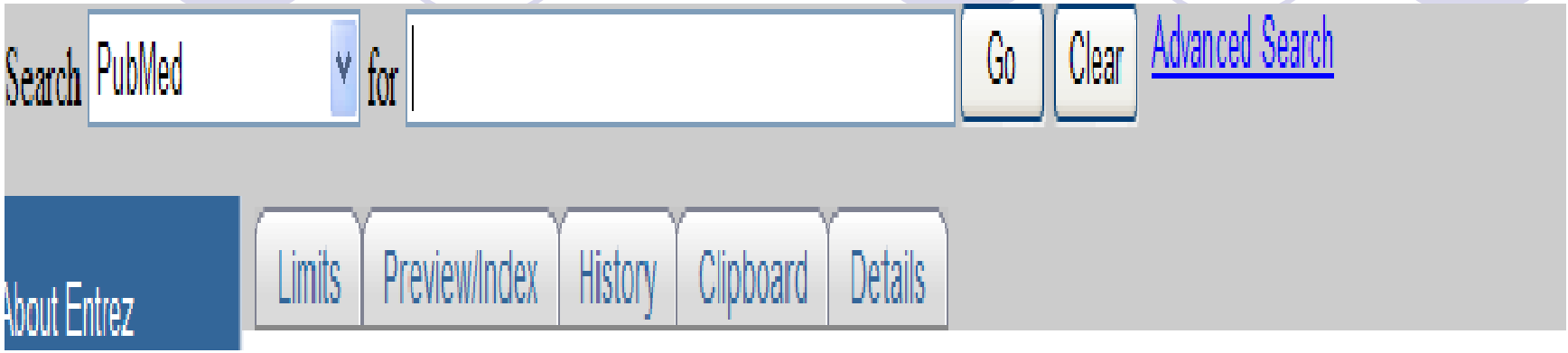
格式为YYYY/MM/DD，如2009/2/27,也可不输后两者

5.检索期刊子集

可供检索的期刊子集有三种：AbridgedIndexMedicus、Dental和Nursing。分别使用jsubsetq, jsubsetd, jsubsetn进行限定，格式为：检索词+AND+jsubsetq

PubMed允许使用布尔逻辑检索，只要在检索词间键入布尔运算符（AND，OR，NOT），可以提高检索的效率。

PubMed的高级检索技巧



在检索框的下方，有PubMed所提供的五个重要的功能按钮，即Limits按钮, Preview/Index按钮, History按钮, Clipboard按钮,及Details按钮。利用好这几个功能按钮，检索则事半功倍。

Limits按钮

可以进行以下设定：

- 将搜索范围设定在一个特定的域。
- 将搜索范围设定在特定的年龄组、性别组、人类或动物学范围。
- 也可以将搜索限定在某一语种出版的或某一特定的文章类型（如综述）。
- 设定只搜索包含摘要，或是有免费全文的文献。
- 设定搜索Entrez数据库或Publication数据库。
- 设定搜索范围为PubMed数据库的某一子数据库（如Abridged Index Medicus 或AIDS相关的条目）。

Preview/Index 按钮

可以进行的设定有：

- 在显示条目之前显示所查到的文献数。
- 随时通过增加查询单词来修改查询方案。
- 在特定的搜索域中向方案里加入查询词。
- 从Index中查看并选择词语来修改查询方案
- 在你修改查询时查看方案。

History 按钮

History屏幕将会显示：查询方案、查询时间、查询到的文献数量。

Preview显示的是历史记录中最近的三条记录，而使用History可以看到最近100次的查询结果。

Clipboard 按钮

剪贴板可以帮助保存或查看在一个或多个查询中选择的条目，然后就可以打印、保存、订购剪切板中的内容了。

Details 按钮

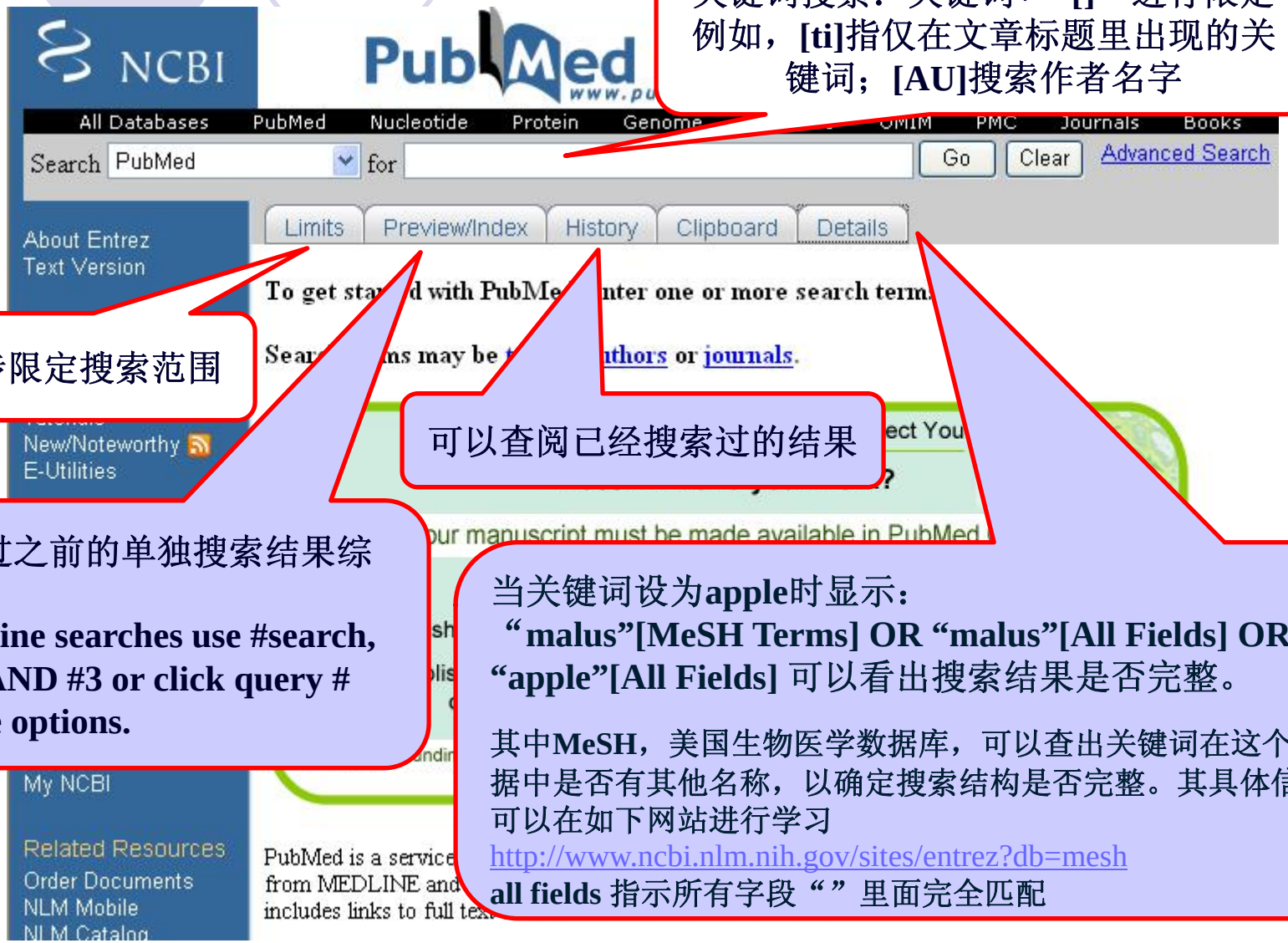
该框下有四个区域：

Result区显示检索结果的记录总数，点击这个数字，可回到检索结果显示屏；

Translation区显示检索词转换的详细情况；

Database区显示检索的数据库；

User Query 区显示用户键入的检索词或检索式。



关键词搜索：关键词+“[]”进行限定
例如，**[ti]**指仅在文章标题里出现的关键词；**[AU]**搜索作者名字

进一步限定搜索范围

可以查阅已经搜索过的结果

可以通过之前的单独搜索结果综合查找
To combine searches use #search, e.g., #2 AND #3 or click query # for more options.

当关键词设为**apple**时显示：
“malus”[MeSH Terms] OR “malus”[All Fields] OR “apple”[All Fields] 可以看出搜索结果是否完整。

其中**MeSH**，美国生物医学数据库，可以查出关键词在这个数据中是否有其他名称，以确定搜索结构是否完整。其具体信息可以在如下网站进行学习
http://www.ncbi.nlm.nih.gov/sites/entrez?db=mesh
all fields 指示所有字段“”里面完全匹配

四、Entrez的使用

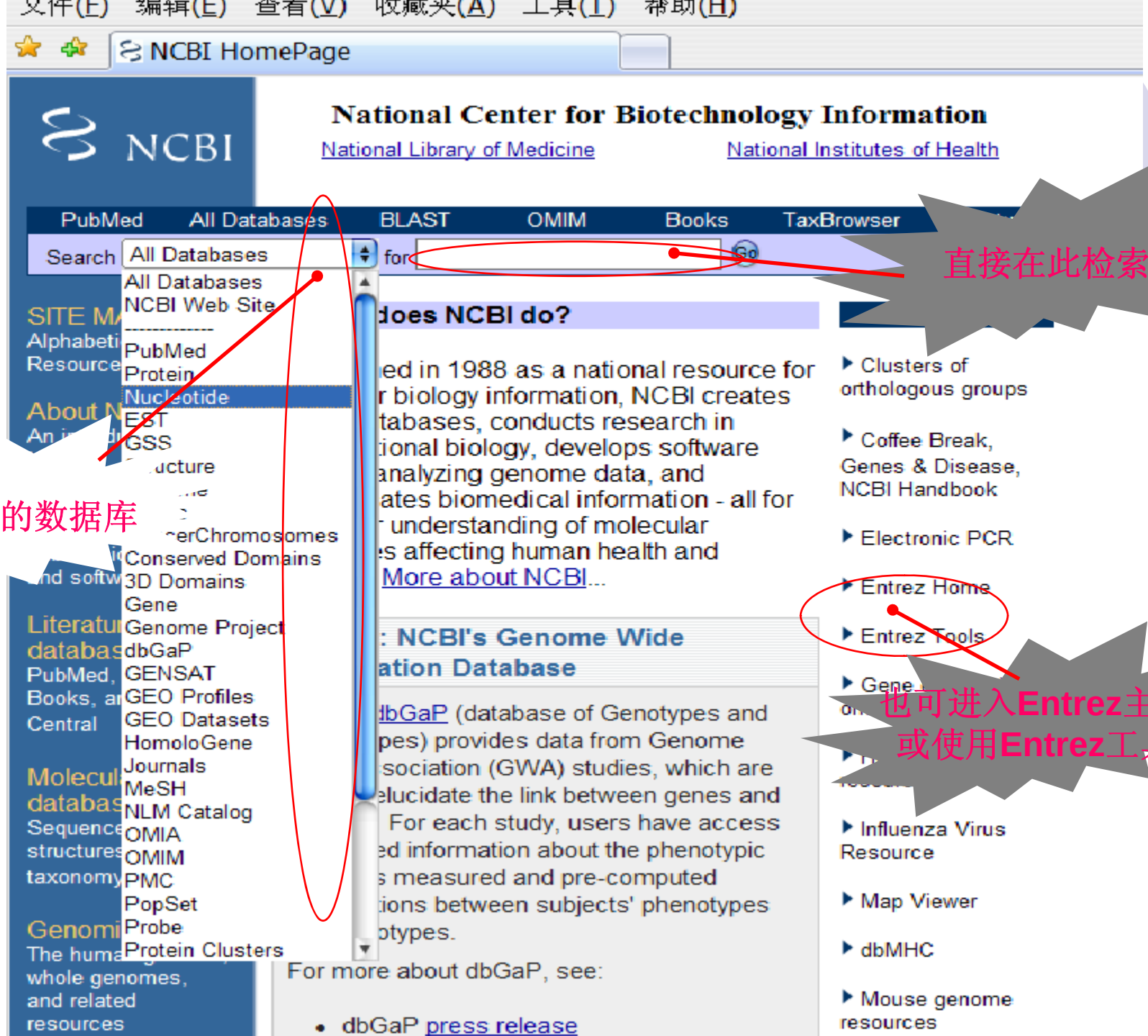
- Entrez系统简介
- Entrez系统的使用方法
- Entrez系统的特点
- Entrez系统的小技巧
- Entrez高级工具

Entrez系统简介

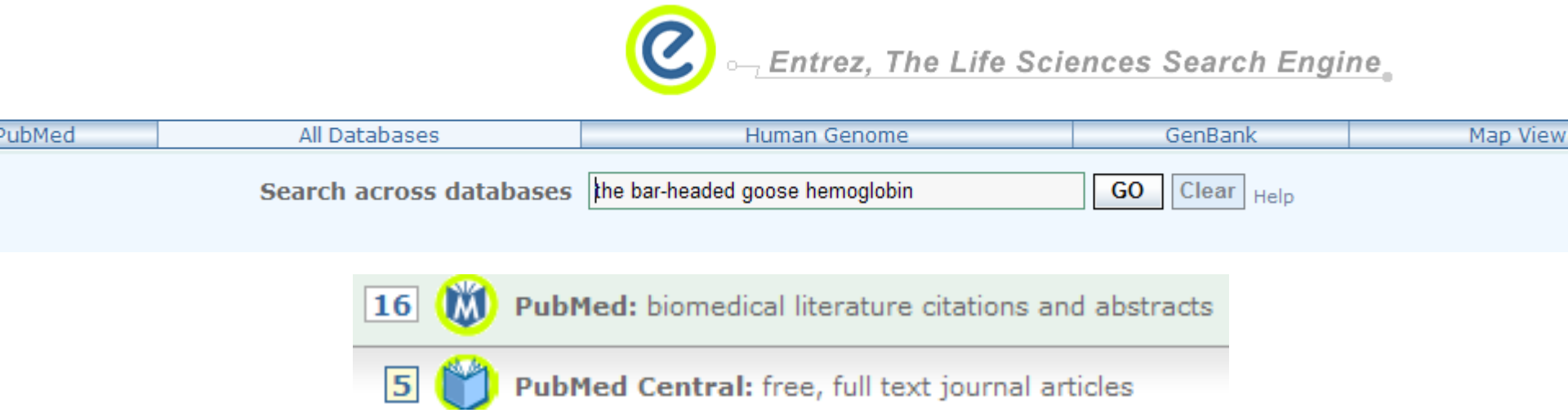
- Entrez整合了科学文献，DNA和蛋白序列数据库，3D蛋白结构和蛋白结构域数据库，种群研究数据库，表达数据库，全基因数据库，分类信息库等，成为一个联系紧密的检索系统（retrieval system）。

Entrez系统的使用方法

- 进入NCBI主页(www.ncbi.nlm.nih)，即可看到位于页面上部的数据库检索栏，其缺省检索选项为All Databases 。可以在检索栏中直接输入需要查询的内容。此外，也可先链接到Entrez主页或使用Entrez高级工具检索。


文件(F) 编辑(E) 查看(V) 收藏夹(A) 工具(T) 帮助(H)
NCBI HomePage
NCBI
National Center for Biotechnology Information
National Library of Medicine
National Institutes of Health
PubMed
All Databases
BLAST
OMIM
Books
TaxBrowser
Search
All Databases
for
Go
All Databases
NCBI Web Site
PubMed
Protein
Nucleotide
EST
GSS
Conserved Domains
3D Domains
Gene
Genome Project
dbGaP
GENSAT
GEO Profiles
GEO Datasets
HomoloGene
Journals
MeSH
NLM Catalog
OMIA
OMIM
PMC
PopSet
Probe
Protein Clusters
does NCBI do?
ed in 1988 as a national resource for
r biology information, NCBI creates
tabases, conducts research in
tional biology, develops software
analyzing genome data, and
ates biomedical information - all for
understanding of molecular
s affecting human health and
More about NCBI...
: NCBI's Genome Wide
ation Database
dbGaP (database of Genotypes and
pes) provides data from Genome
sociation (GWA) studies, which are
elucidate the link between genes and
For each study, users have access
ed information about the phenotypic
s measured and pre-computed
tions between subjects' phenotypes
otypes.
For more about dbGaP, see:
dbGaP press release
Clusters of orthologous groups
Coffee Break, Genes & Disease, NCBI Handbook
Electronic PCR
Entrez Home
Entrez Tools
Influenza Virus Resource
Map Viewer
dbMHC
Mouse genome resources
直接在此检索
选择需要的数据库
也可进入Entrez主页
或使用Entrez工具

以“the bar-headed goose hemoglobin”为例Entrez检索



- 检索结果：PubMed(16), PubMed Central(5), Protein(23), Structure(4), 3D Domains(16), PubChem Substance(3), 其它库中为（0）

- Notice：GenBank和EMBL等核酸序列数据库中的大部分数据，是由生物学家通过计算机网络直接提交，或通过计算机程序直接从大规模序列测定所得结果送入数据库中，没有严格的标准。在数据库查询时，经常会遇到“想找的找不到，找到的却不是”这样的问题。

- 例如，检索“spider toxin”查询所得到的225个核酸序列条目，有很大一部分是重复的；而我国特有蜘蛛“虎纹捕鸟蛛”的毒素(Huwentoxin)却没有检索到。这是因为作者在提交该序列时，使用了“Huwentoxin”，而没有使用“spider toxin”。因此，必须输入“Huwentoxin”，才能找到该序列条目。

Entrez系统的特点

- 查询系统，使用方便
  - 它把序列、结构、文献、基因组、系统分类等不同类型的数据库有机地结合在一起，通过超文本链接，用户可以从一个数据库直接转入另一个数据库。
- 整合了数据库和应用程序
  - 通过“Related sequence”工具，可以直接找到与查询所得蛋白质序列同源的其它蛋白质。
- 实际匹配+相近匹配
  - 通过点击“Related Articles”继续查找相关文献。
- 伴随数据实时更新，每次检索结果也会不同

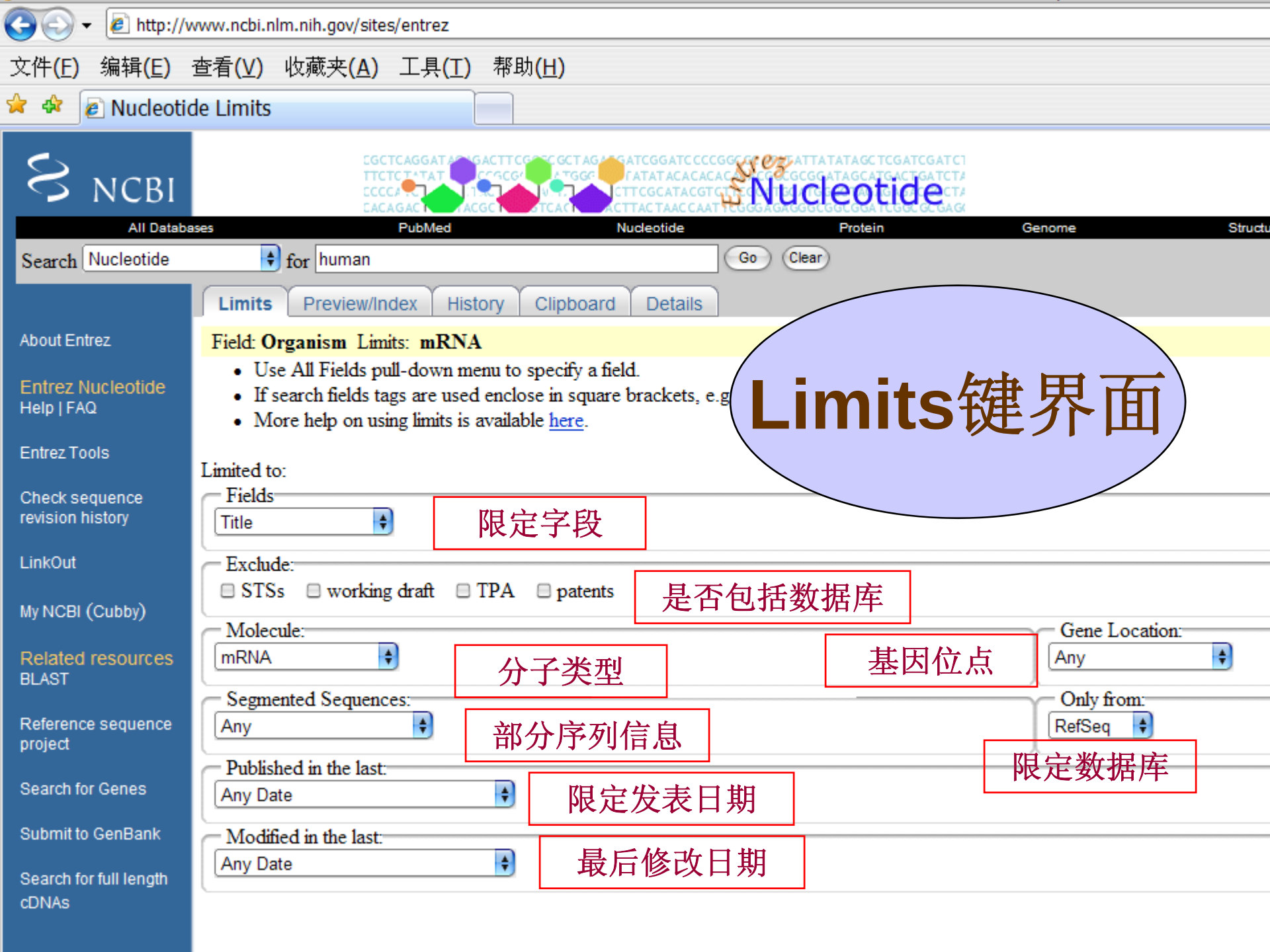
Entrez系统的小技巧

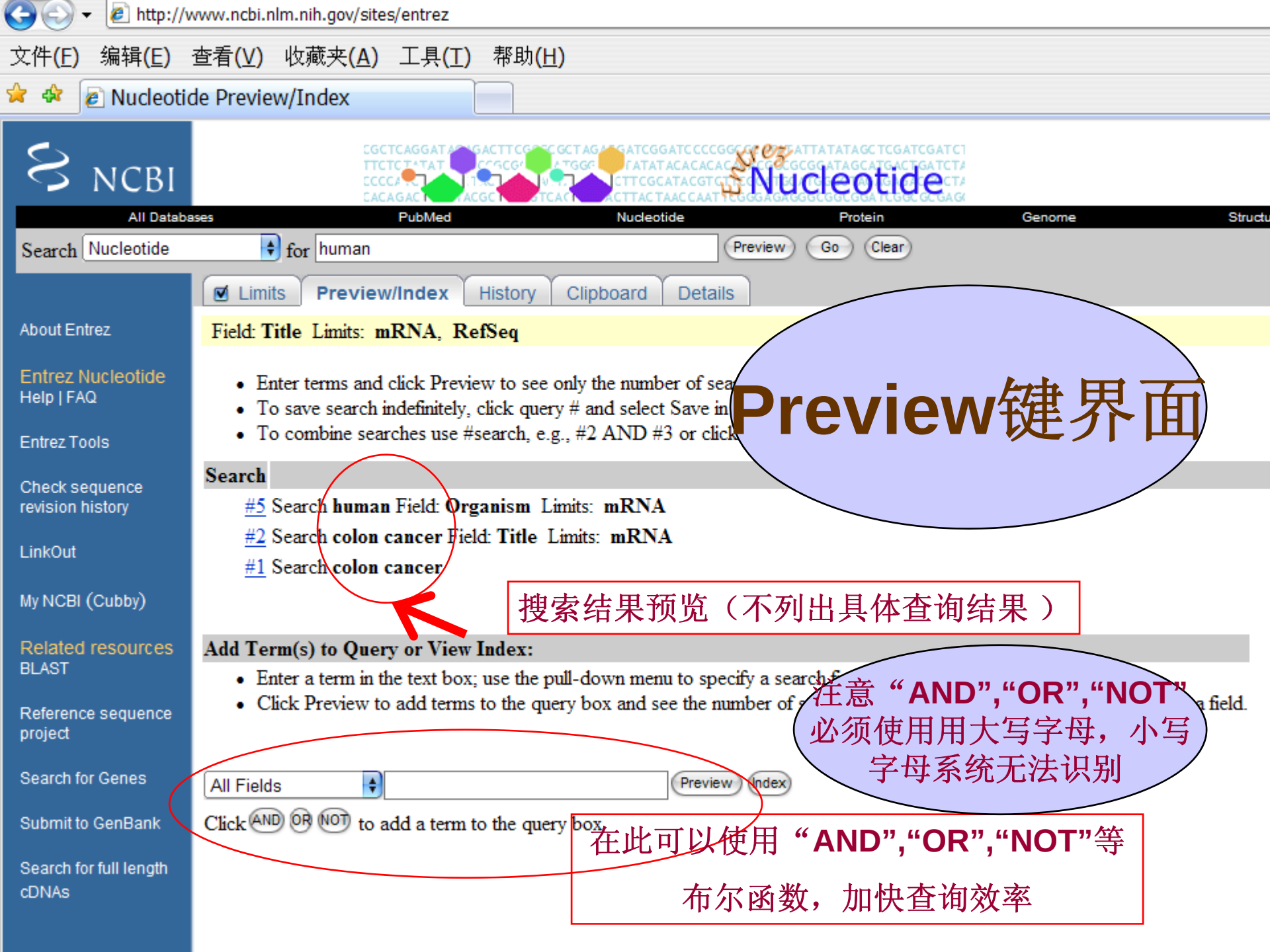
- Entrez系统提供各种辅助功能，包括限定查询范围(Limits)、预览查询结果(Preview/Index)、查看查询记录(History)和操作剪贴板(Clipboard)。

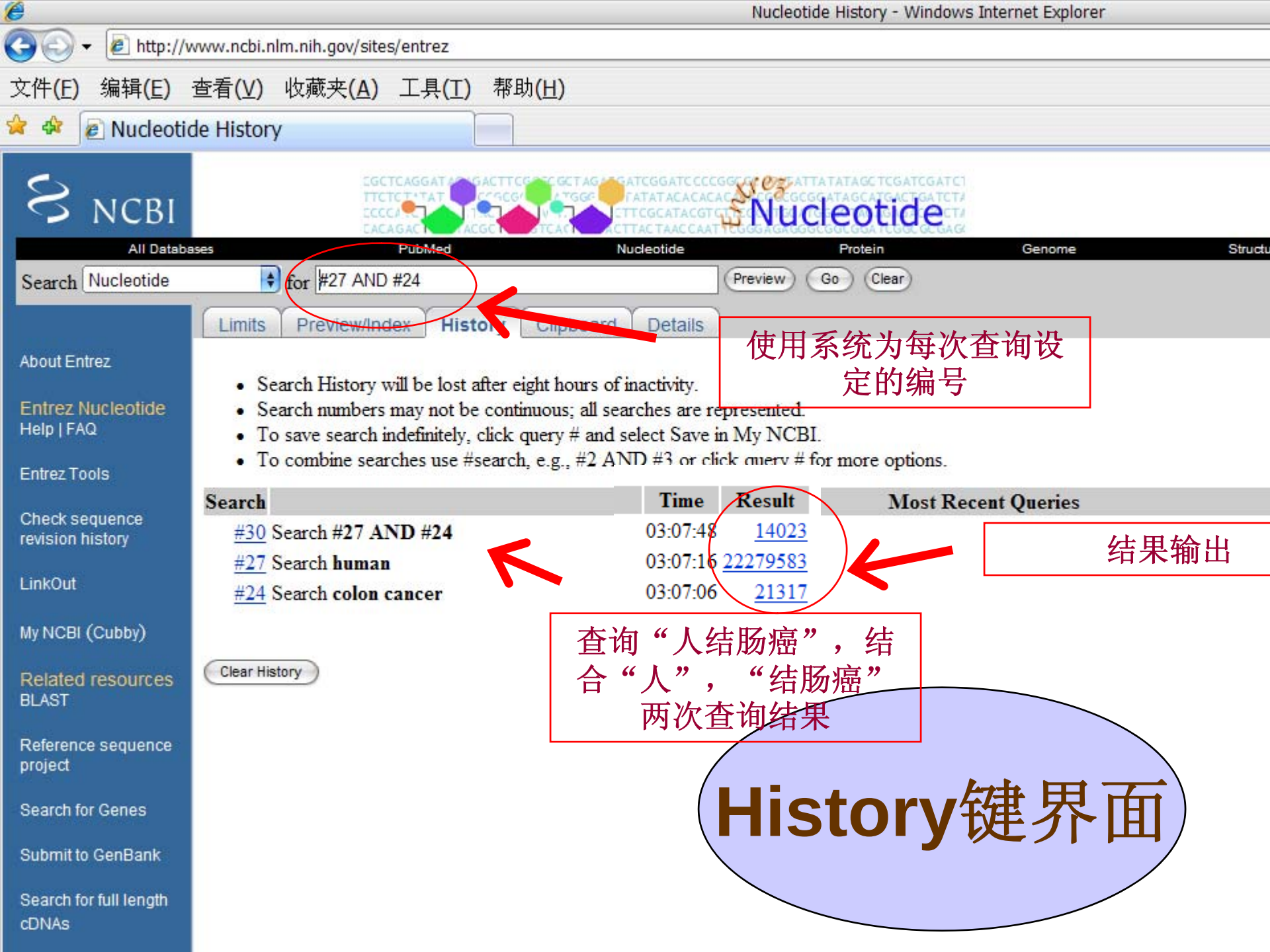
- Geer, R.C. and Sayers, E.W. Entrez: Making use of its power. *Briefings in Bioinformatics*. 2003 June;4(2):1779-184.

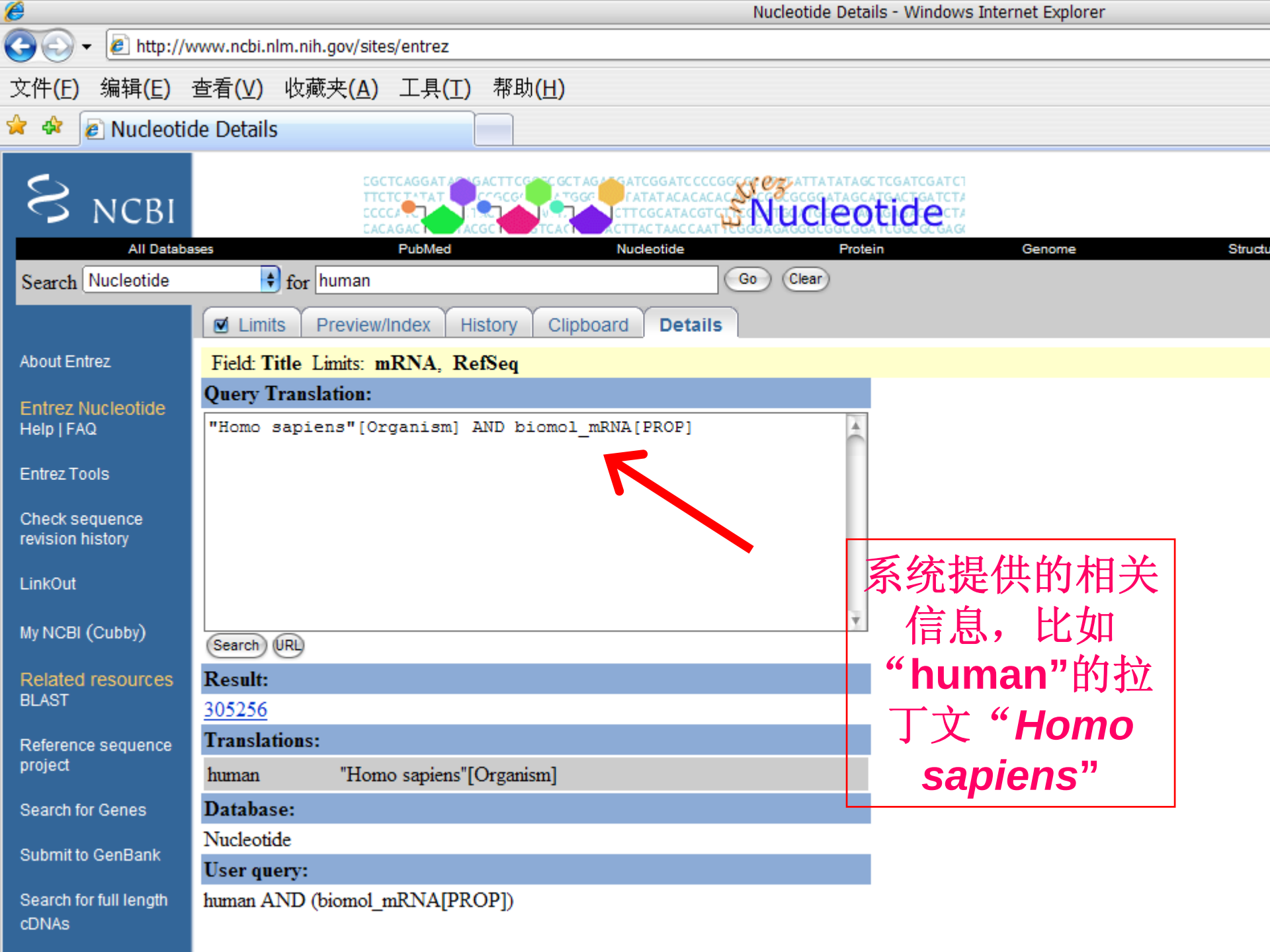
- 击Limits按钮，即可进入限定查询范围页面，可以根据该数据库结构，将输入的关键词的查询范围限制在某个范围内，如编号、代码、提交日期等。而不同的数据库，其限定范围不同，如序列数据库可以限定序列长度，文献数据库则可以限定作者、题目、杂志名称等。
- 点击预览查询按钮(Preview/Index)，检索栏中会增加一个“Preview”按钮，输入关键词后，若点击“Preview”按钮，则不列出具体查询结果，而只列出查询到的数据条目数。利用这一辅助功能，可以提高查询速度，并对查询结果有个初步了解，以便对查询结果作进一步处理，缩小查询范围。
- 点击“History”按钮，则可以查看查询过程的记录，对每次查询结果进行分析，并作进一步处理。


文件(F) 编辑(E) 查看(V) 收藏夹(A) 工具(T) 帮助(H)
colon cancer - Nucleotide Results
以查找“结肠癌”基因为例
选择数据库
NCBI
Entrez Nucleotide
All Databases
PubMed
Nucleotide
Protein
Genome
Structu
Search Nucleotide for colon cancer
Go
Clear
Save Search
Limits
Preview/Index
History
Clipboard
Details
输入关键词
“结肠癌”
History has expired.
Found 22493 nucleotide sequences.
辅助功能按钮
Display Summary
Send to
All: 21317
Bacteria: 10
RefSeq: 587
mRNA: 861
Items 1 - 20 of 21317
Page 1 of 1066
检索结果
This search in Gene shows 614 results, including:
PTPRJ (Homo sapiens): protein tyrosine phosp
MSH2 (Homo sapiens): mutS homolog 2, colon
MLH1 (Homo sapiens): mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)
1: BT059123 Reports
Salmo salar clone ssal-rgf-526-250 Serologically defined colon cancer antigen 1 homolog putative mRNA, complete cds
gi|223648155|gb|BT059123.1|[223648155]
2: FM992691 Reports
Candida dubliniensis CD36 chromosome 4, complete sequence
gi|223643022|emb|FM992691.1|[223643022]


http://www.ncbi.nlm.nih.gov/sites/entrez
文件(F) 编辑(E) 查看(V) 收藏夹(A) 工具(T) 帮助(H)
Nucleotide Limits
NCBI
Entrez Nucleotide
All Databases PubMed Nucleotide Protein Genome
Search Nucleotide for human Go Clear
Limits Preview/Index History Clipboard Details
About Entrez
Entrez Nucleotide Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI (Cubby)
Related resources BLAST
Reference sequence project
Search for Genes
Submit to GenBank
Search for full length cDNAs
Field: Organism Limits: mRNA
Use All Fields pull-down menu to specify a field.
If search fields tags are used enclose in square brackets, e.g
More help on using limits is available here.
Limits键界面
Limited to:
Fields
Title
限定字段
Exclude:
STSs working draft TPA patents
是否包括数据库
Molecule:
mRNA
分子类型
Gene Location:
Any
基因位点
Segmented Sequences:
Any
部分序列信息
Only from:
RefSeq
限定数据库
Published in the last:
Any Date
限定发表日期
Modified in the last:
Any Date
最后修改日期


http://www.ncbi.nlm.nih.gov/sites/entrez
文件(F) 编辑(E) 查看(V) 收藏夹(A) 工具(T) 帮助(H)
Nucleotide Preview/Index
NCBI
Entrez Nucleotide
All Databases PubMed Nucleotide Protein Genome
Search Nucleotide for human Preview Go Clear
Limits Preview/Index History Clipboard Details
About Entrez
Entrez Nucleotide
Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI (Cubby)
Related resources
BLAST
Reference sequence project
Search for Genes
Submit to GenBank
Search for full length cDNAs
Field: Title Limits: mRNA, RefSeq
Enter terms and click Preview to see only the number of sea
To save search indefinitely, click query # and select Save in
To combine searches use #search, e.g., #2 AND #3 or click
Search
#5 Search human Field: Organism Limits: mRNA
#2 Search colon cancer Field: Title Limits: mRNA
#1 Search colon cancer
Add Term(s) to Query or View Index:
Enter a term in the text box; use the pull-down menu to specify a search f
Click Preview to add terms to the query box and see the number of
field.
All Fields Preview Index
Click AND OR NOT to add a term to the query box
Preview键界面
搜索结果预览（不列出具体查询结果 ）
注意“AND”,“OR”,“NOT”必须使用用大写字母，小写字母系统无法识别
在此可以使用“AND”,“OR”,“NOT”等布尔函数，加快查询效率


Nucleotide History - Windows Internet Explorer
http://www.ncbi.nlm.nih.gov/sites/entrez
文件(F) 编辑(E) 查看(V) 收藏夹(A) 工具(T) 帮助(H)
Nucleotide History
NCBI
Nucleotide
All Databases PubMed Nucleotide Protein Genome
Search Nucleotide for #27 AND #24 Preview Go Clear
Limits Preview/Index History Clipboard Details
About Entrez
Entrez Nucleotide
Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI (Cubby)
Related resources
BLAST
Reference sequence project
Search for Genes
Submit to GenBank
Search for full length cDNAs
Search History will be lost after eight hours of inactivity.
Search numbers may not be continuous; all searches are represented.
To save search indefinitely, click query # and select Save in My NCBI.
To combine searches use #search, e.g., #2 AND #3 or click query # for more options.
Search
Time
Result
Most Recent Queries
#30 Search #27 AND #24 03:07:48 14023
#27 Search human 03:07:16 22279583
#24 Search colon cancer 03:07:06 21317
Clear History
使用系统为每次查询设定的编号
结果输出
查询“人结肠癌”，结合“人”，“结肠癌”两次查询结果
History键界面


Nucleotide Details - Windows Internet Explorer
http://www.ncbi.nlm.nih.gov/sites/entrez
文件(F) 编辑(E) 查看(V) 收藏夹(A) 工具(T) 帮助(H)
Nucleotide Details
NCBI
Entrez Nucleotide
All Databases PubMed Nucleotide Protein Genome
Search Nucleotide for human Go Clear
Limits Preview/Index History Clipboard Details
About Entrez
Entrez Nucleotide
Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI (Cubby)
Related resources
BLAST
Reference sequence project
Search for Genes
Submit to GenBank
Search for full length cDNAs
Field: Title Limits: mRNA, RefSeq
Query Translation:
"Homo sapiens"[Organism] AND biomol_mRNA[PROP]
Search URL
Result:
305256
Translations:
human "Homo sapiens"[Organism]
Database:
Nucleotide
User query:
human AND (biomol_mRNA[PROP])
系统提供的相关信息，比如“human”的拉丁文“Homo sapiens”

Entrez高级工具

- **Web Tools:**
  - **Batch Entrez - Upload a file of GI or accession numbers to retrieve sequences.**
  - **PubMed Batch Citation Matcher - Send citation information to Entrez and retrieve PubMed IDs for linking, citation display, or other applications.**
  - **Advanced Entrez Searching - Advanced searching techniques for Web Entrez.**
  - **My NCBI - includes automatic e-mailing of search updates and filters for search results.**

Programming Tools:

- **E-Utilities** - Run Entrez queries and download data from your own scripts over the Web.
- **Linking to Entrez** - Link to specific Entrez pages from your own web pages or applications.
- **Entrez Client/Server** - C language library for embedding Entrez calls into your programs.

五、SRS的使用

- SRS是Sequence Retrieval System的缩写，由欧洲分子生物学实验室开发，最初是为核酸序列数据库EMBL和蛋白质序列数据库SwissProt的查询开发的。随着分子生物信息数据库应用和开发的需求不断增长，SRS已经成为欧洲各国主要生物信息中心必备的数据库查询系统。

SRS是一个开放的数据库查询系统，即不同的SRS查询系统可以根据需要安装不同的数据库，目前共有300多个数据库安装在世界各地的SRS服务器上。

单 位	网 址
欧洲生物信息研究所	http://srs6.ebi.ac.uk/srs6/
英国基因组资源中心	http://iron.hgmp.mrc.ac.uk/srs6/
英国基因组测序中心	http://www.sanger.ac.uk/srs6/
法国生物信息中心	http://www.infobiogen.fr/srs6/
荷兰生物信息中心	http://www.cmbi.kun.nl/srs6/
澳大利亚医学研究所	http://srs.wehi.edu.au/srs6/
德国癌症研究所	http://genius.embnet.dkfz-heidelberg.de/menu/srs/
加拿大生物信息资源中心	http://www.cbr.nrc.ca/srs6.1/

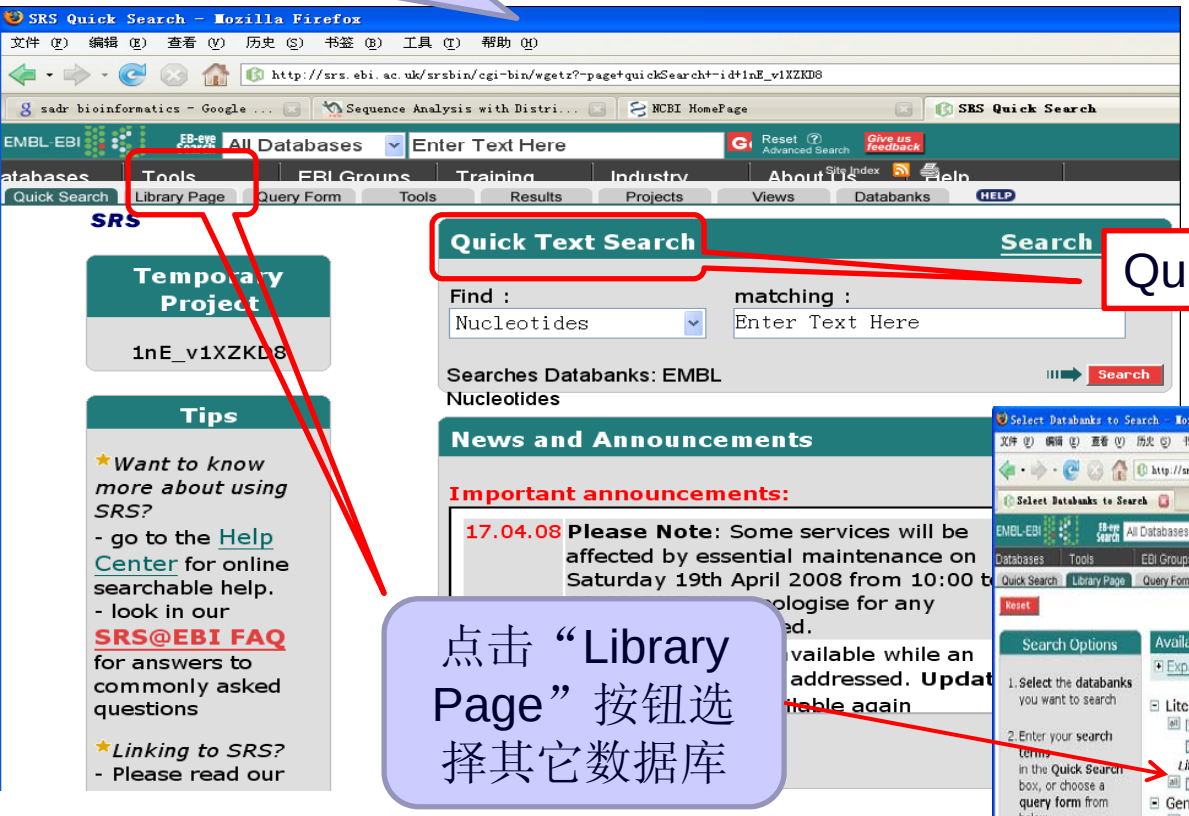
SRS检索方法

- 快速检索
- 标准检索（**Standard Query**）
- 扩展检索（**Extendard Query**）

快速检索

键入该服务器网址（srs.ebi.ac.uk）
进入SRS系统的快速检索界面。

欧洲生物信息学研究所主页（www.ebi.ac.uk）上方导航栏Databases（数据库）按钮



Quick Text Search（快速文本搜索）

点击“Library Page”按钮选择其它数据库

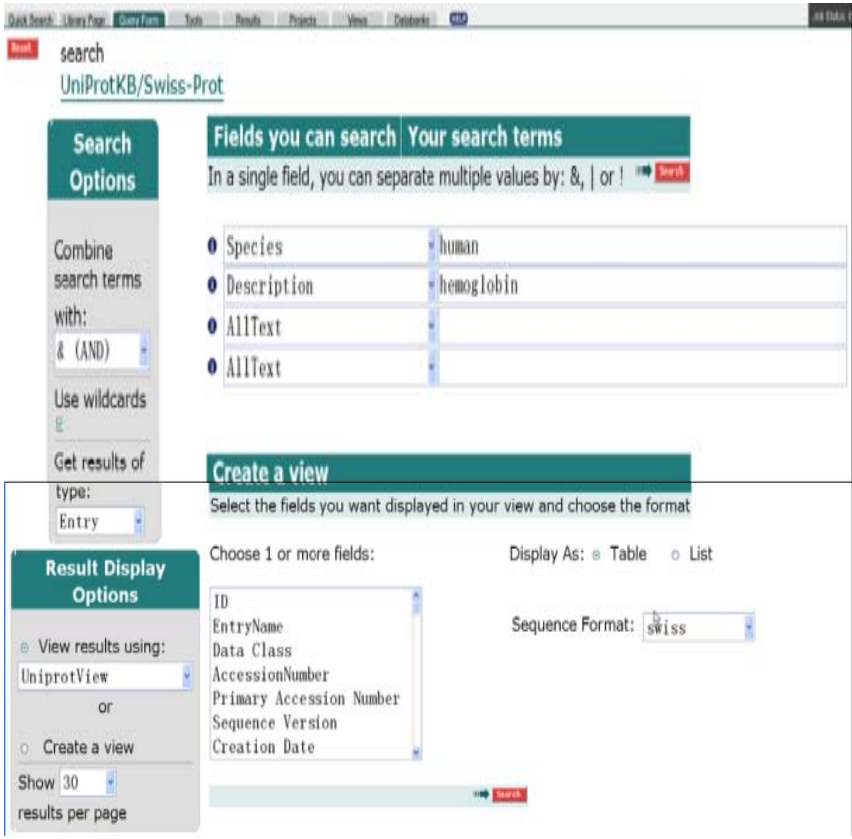
数据库页面

策　略

- 快速检索界面简洁、使用比较方便，便于初学者使用。
- 其缺点是无法限定检索范围，只能进行全文检索
- SRS 检索词不分大小写。通配符“*”表示可匹配任意字符，检索时由系统自动添加，另外两个逻辑运算符号为逻辑与“&”和逻辑或“|”。

SRS 系统的标准检索界面（Standard Query）

- 提供了可供选择的检索范围，用户可以通过下拉式菜单指定所输入的检索词出现在数据库的某个字段
- 将Fields You can Search（可检索字段）提供的第一个下拉菜单中All Text（全文）改为ID（识别符）
- 利用SRS 系统提供的逻辑非运算符“!”（but not），可以进一步减少假阳性，提高检索效率。运算符“!”，如hemoglobin !receptor，可过滤掉Description 字段中出现的Hemoglobin scavenger receptor
- 同时检索两个不同词条的结果时，运用“|”，如在对应的Your Search Term（检索词条）输入框中输入HBA_ANSIN| HBB_ANSIN



扩展检索（Extendard Query）

检索策略：

- 逻辑运算符有“&”、“ |”和“ !”，分别表示与（and）、或（or）和非（but not）
- 默认逻辑运算为与默认检索由左至右，用圆括号“（）”可改变检索优先次序
- 系统在检索词条后自动添加通配符“*”，如输入peptide，结果中包括peptide 和 peptides
- 用双引号可取消通配符功能，如输入“peptide”，则不返回含peptides 的条目
- 用双引号可限定只返回多单词短语，如“disulfide bridge”、“protein kinase”
- 用正则表达式可限定检索范围，如输入/^phos/返回含以“phos”起始的条目，输入/ase$/返回含以“ase”结尾的条目。
- 用冒号“：”可限定数值型字段检索范围，包括序列长度、日期、登陆号等
- 用中括号“[ ]”可限定检索字段，如检索作者Rice，可用Rice[AU]

结果显示和保存

- 1) 数据分析功能

  分析软件包，如EMBOSS、Phylip、HMMER 和BLAST 等、并按功能分门别类。

  点击Tool（工具）按钮，可以查看和选择所整合的分析程序，输出结果界面的Results Options（结果选项）的Luanch Analysis Tool（启动分析工具）下拉菜单中选择需要使用的分析工具，点击Launch（启动）按钮，即可启动所选分析工具，进入分析界面。

- 2) 检索结果整合功能

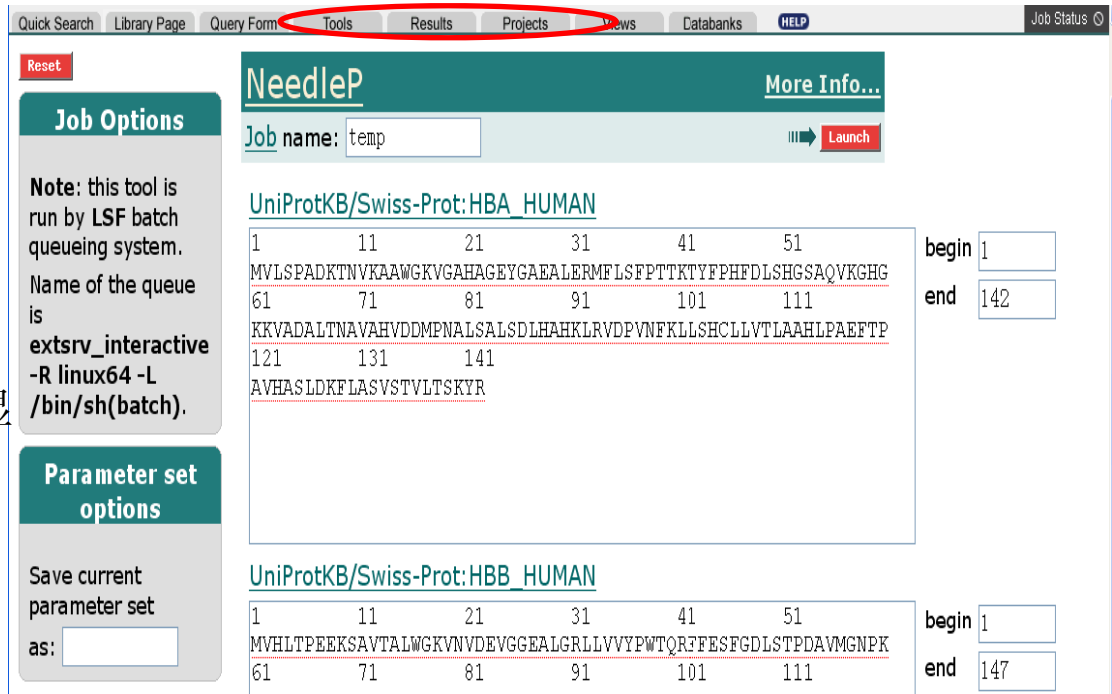
  点击Results 按钮，进入SRS 系统的检索结果整合界面，查看检索过程，并对不同检索步骤进行组合、链接等处理，得到新的检索结果。

- 3) 检索策略保存功能

  点击Project（课题）按钮，进入检索过程存储页面，可以将某次检索过程以文件形式下载到用户本地计算机上保存起来；也可把本地计算机上的存放的检索过程文件上载到服务器上。

- 4) 输出界面定制功能

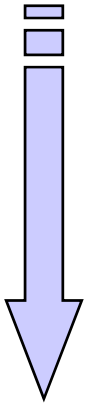
  点击Views（显示）按钮，进入输出显示管理页面，用户可以对不同数据库的检索结果定制功能;



六、总结

- 不同检索系统之间既有相同的规则，也根据检索对象不同而各有特点；
- 熟练掌握检索系统的策略和技巧，能极大地提高检索效率；
- 无论看多少资料，听多少报告，只有亲自进行检索，体会，总结之后，才能真正掌握。

生物信息学



后基因组时代的钥匙

谢　谢！