５ ． Entrez の利用法

目次

1 Entrez の起動 1

2 検索語句の指定 4

3 関連項目を辿る：同一データベース内の関連事項 7

4 関連項目を辿る：データベースにまたがる相互参照 9

5 検索条件の変更： OR 結合 11

6 ファイルへの保存 15

7 終了 16

8 HGC 外部からの無記名利用資格によるアクセス 16

9 参考 . 検索モード 17

10 補足 Macintosh からネットワーク版 Entrez を使う 18

はじめに

Entrez （アントレと発音します）はアメリカ National Center for Biotechnology Information (NCBI) で作成された、文献データベースと配列データベースを有機的に関連づけて情報を検索するシステムで、強力なキーワード検索と、情報の linking による関連項目検索に特徴があります。また、ウィンドウシステムを使用しているため、操作が感覚的で優しいという便利さもあります。

講習内容

- 起動方法
- キーワードの指定の仕方：通常 = AND 結合、 OR 結合の指定方法
- 関連項目を辿る：同一データベース内の関連項目、別のデータベースの相互参照
- 出力フォーマットの指定とファイル保存
- HGC 外部からの無記名利用資格によるアクセス

1 Entrez の起動

Entrez は Xwindow というウィンドウシステムを使用しているので、まずこれが動くマシンからでないと使用できない。 HGC 内部のマシンでは、 Sun, IRIS 共に利用可能で、 Sun の場合は通常利用者がつかっている OpenWindows でそのまま使える。 IRIS でも X window が使われているのでそのまま使える。逆に、 PC などでは使えるものが限られ、 Macintosh の場合、 eXodus または MacX といった市販ソフトを、また PC-98 / IBM 互換マシンでもそれぞれ Xwindow 用のソフトを入れないと動かないうえに、能力の小さな PC では使い勝手が悪い。ここではまず HGC 内部の Sun を利用していることを想定して説明する。

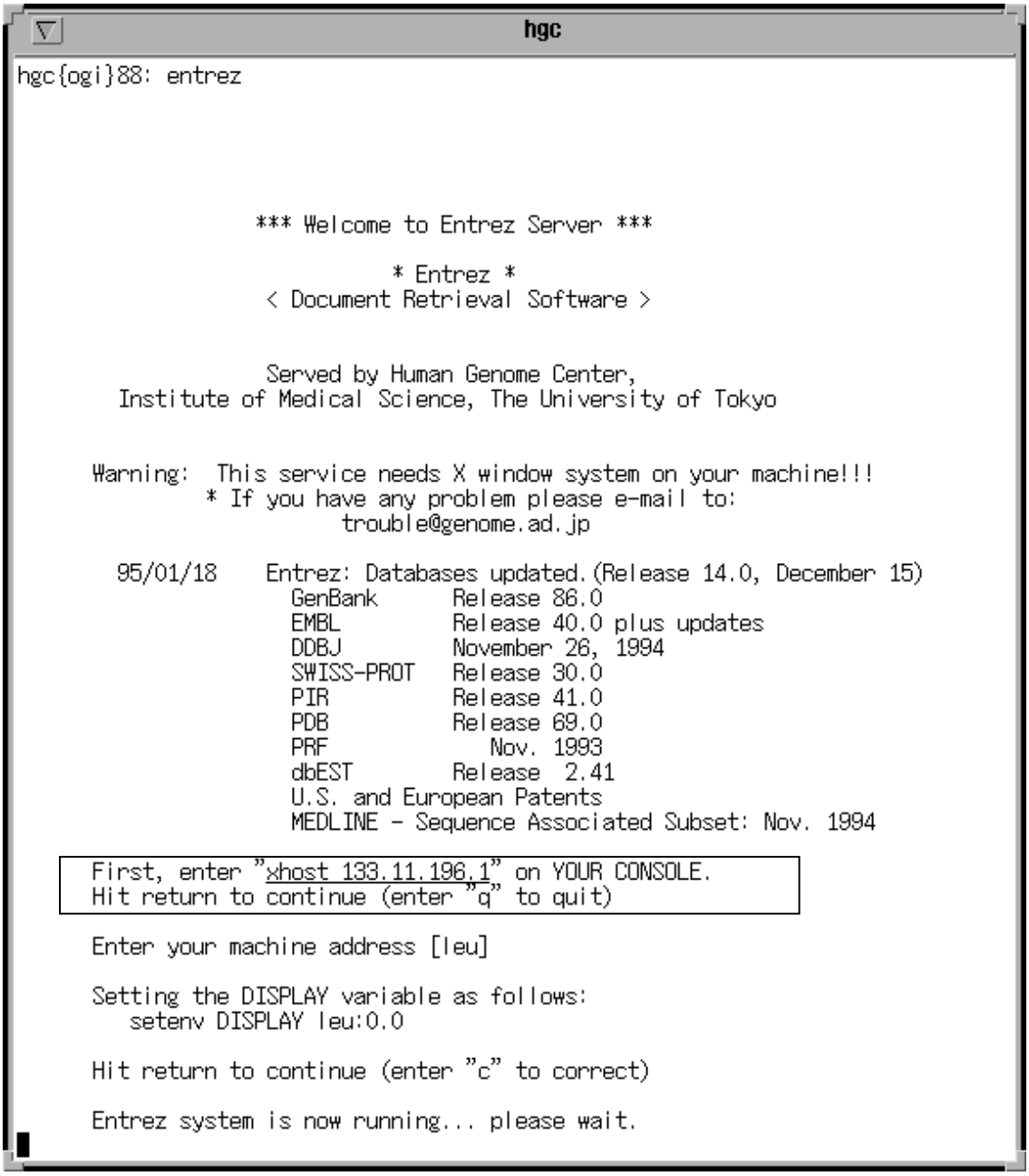
まずそのワークステーションで OpenWindows （または Xwindow）が動いていないといけないが、 HGC 利用資格を持った人がセンター設置のワークステーションにログインすると自動的に立ち上がるようになっているので特に操作はいらない。

次に hgc に rlogin する（最初から hgc のコンソールを使用した場合は不要）。

```
rlogin␣hgc
```

そしてここで Entrez を起動する。

そのためには `entrez` とタイプすればよい。次のようなオープニングメッセージが現われる。



ここで、四角で囲ったメッセージに注意。これは、利用者が向かっているワークステーションに Entrez のウィンドウを表示することを許可するための操作を指示しているものである。これを怠るとそのワークステーションにウィンドウが表示できないことがある。そこで指示されたとおりに入力すればよいのだが、メッセージにも書かれているように、このウィンドウに続けて書くのではなく、コンソールウィンドウから入力すること。（なお、これはウィンドウシステムを再立ち上げしないかぎり効果が持続するので、同一のログインセッション内で何回も Entrez を使用するときは 2 回目以降は省略してよい。）

（コンソールウィンドウで）

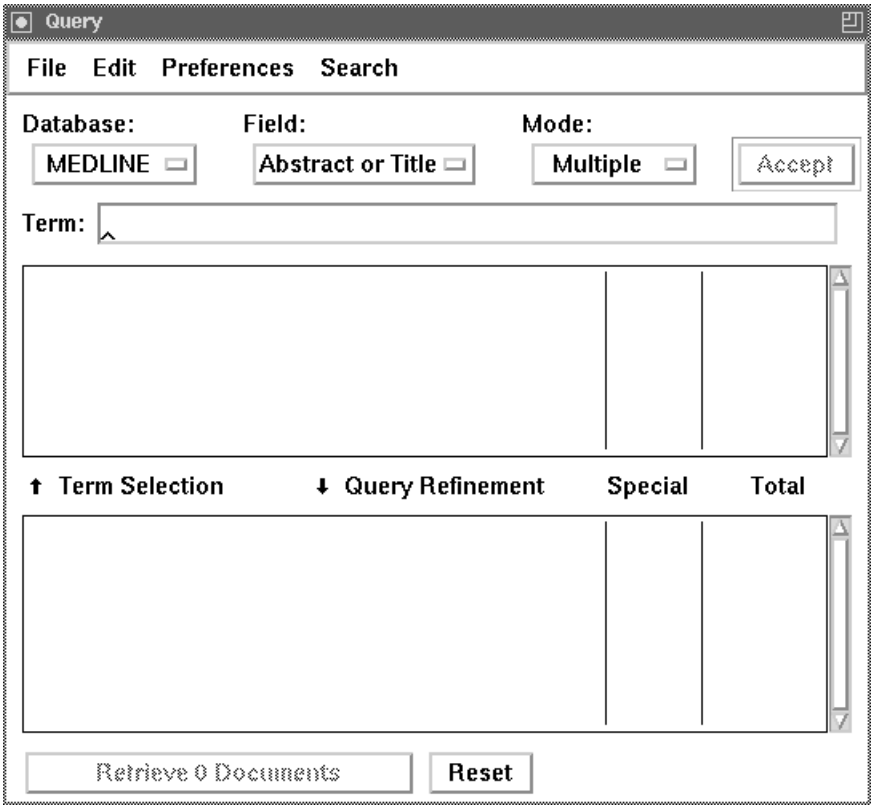
```
xhost␣133.11.196.1
```

このように入力し終わったら、Entrez を立ち上げたウィンドウに戻ってリターンキーを叩く。

その次に尋ねているのは、そのとき利用者が向かっている（ウィンドウを立ち上げている）マシンの名前である。これもシステムが推定して角括弧内に名前を表示しているが、これが正しかったならばリターンキーを打つだけでよい。もし違ったり、途中で切れている場合は初めからきちんといれなくてはならない。もし名前をうまく受け付けないようならば、IP アドレスを試してみるとよい。

最後に確認のリターンを打つと、次のようなウィンドウがまず生成される。

このウィンドウはしばらく表示された後消えて、次のウィンドウが現われる。



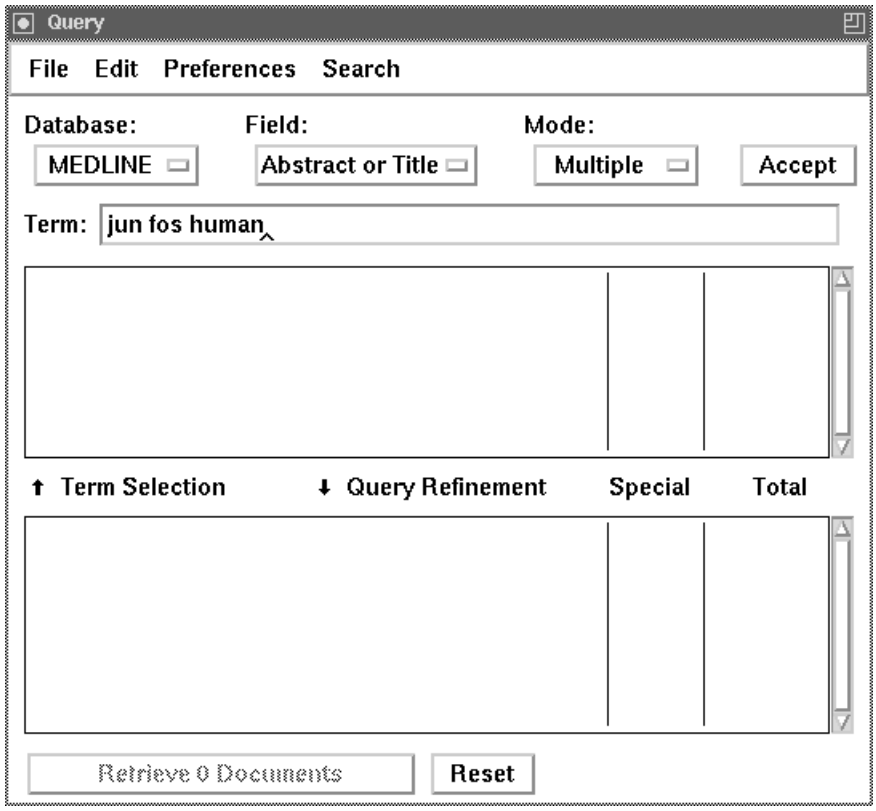
2　検索語句の指定

この Query ウィンドウが Entrez の検索のベースとなるメインウィンドウで、問い合わせキーワードなどはここから入力する。ウィンドウ上部の Database: Field: Mode: の下はそれぞれ選択ボタンになっていて、ここで検索対象データベースや、検索対象フィールド、検索のモードがあらかじめ設定できる。この下の Term: に続く細長い領域が問い合わせの言葉を入れるところで、ここに言葉を入れ終わったらリターンを叩くか、または右上の Accept (Find) ボタンを押せば検索が開始される。

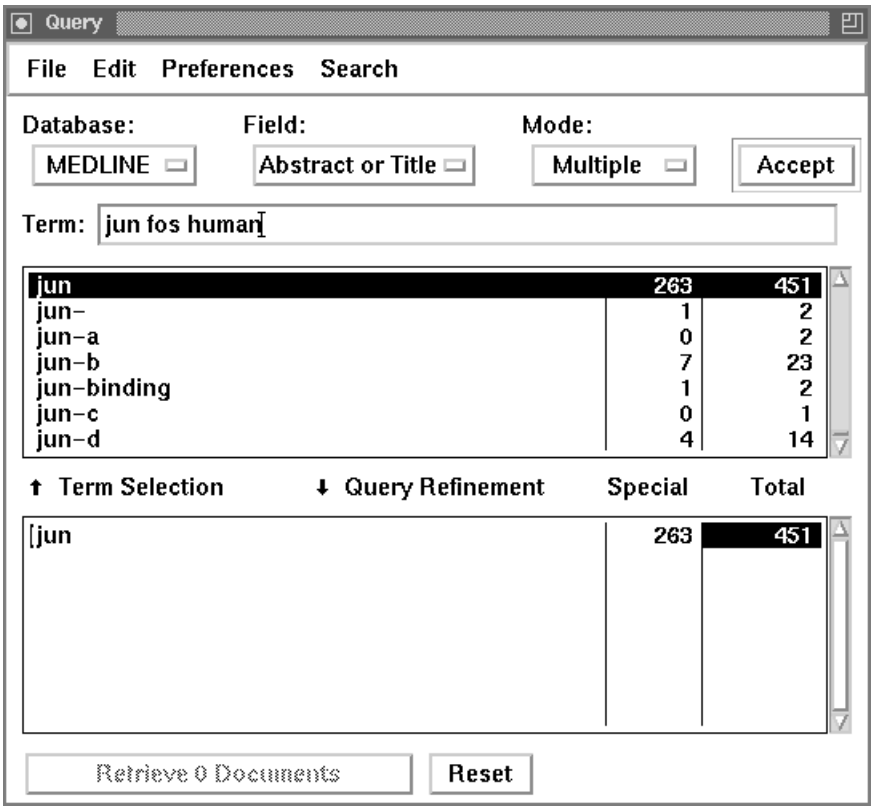
詳しい説明はまた後ほどにして、まず実例で試してみよう。最初の状態では Database: MEDLINE つまり文献データベースに、Field: Abstract or Title に、また Mode: Multiple になっている。ではまずここで Term: フィールドに

```
jun␣fos␣human
```

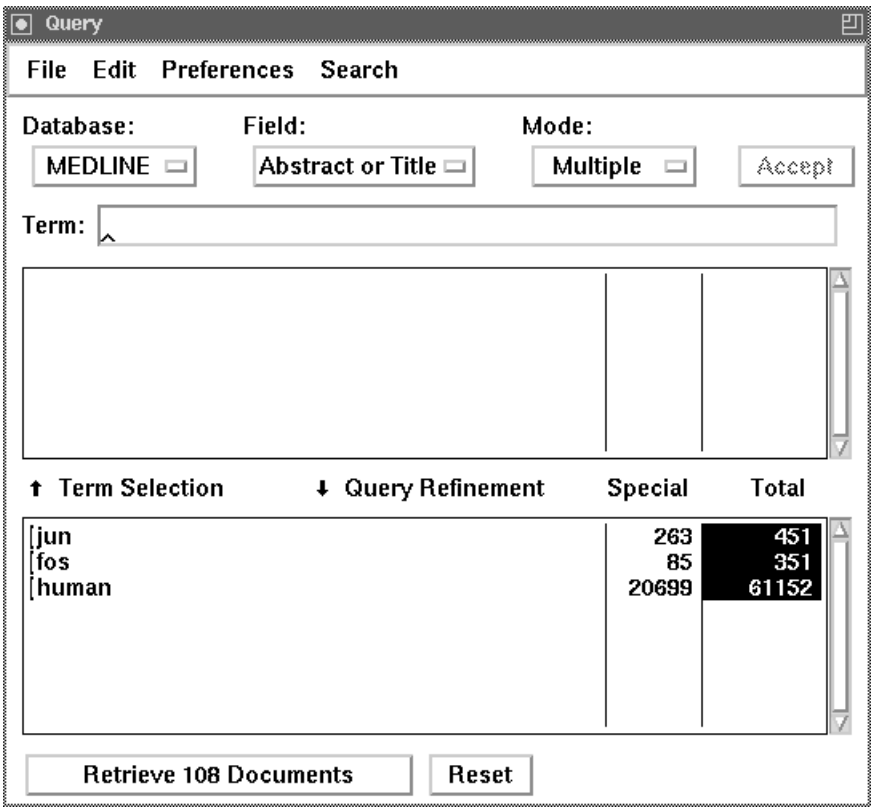
とタイプしてみよう。これは、Medline の Abstract か Title フィールド中で、jun, fos, human の 3 つのキーワードを持ったエントリーを検索しようとしている。



ここでリターンあるいはアクセプトボタンを押すと、次の図のように、まず上側の Term Selection サブウィンドウに、キーワードの単語毎に近いもの順にヒットした件数の表示が一瞬出て、その結果が下の Query Refinement サブウィンドウに蓄積されて行く。



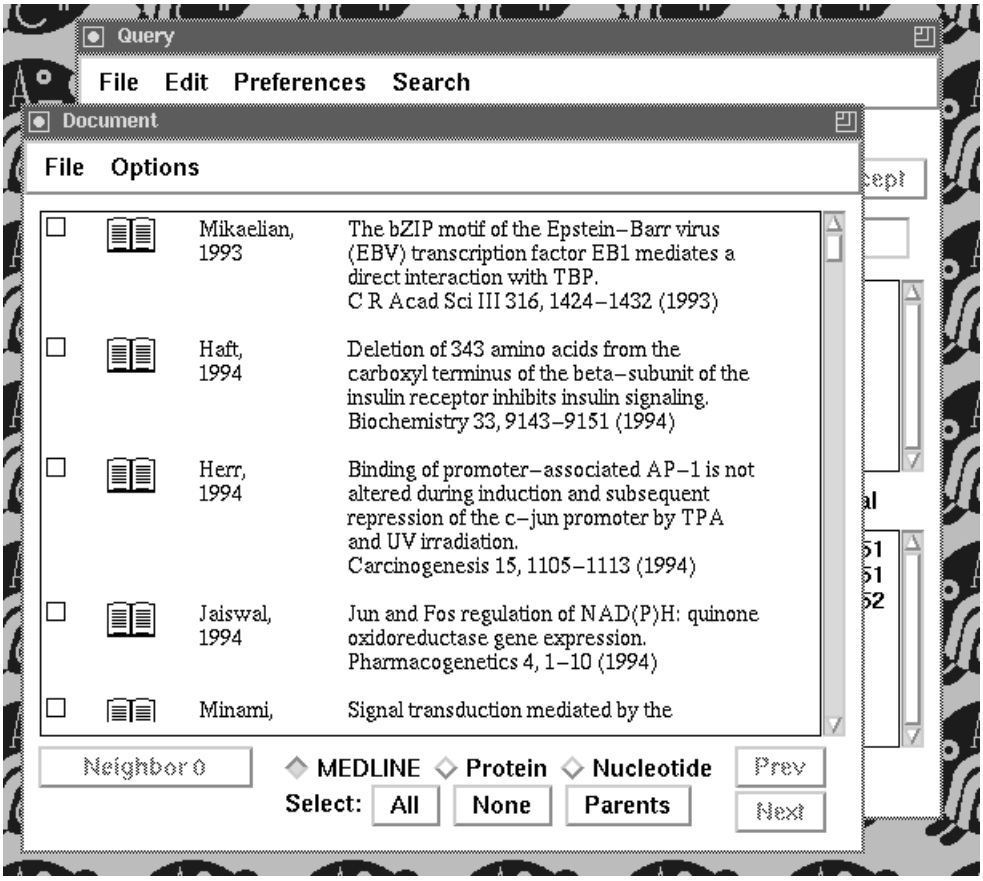
（上の図は検索途中、下の図は検索が終わったところ）



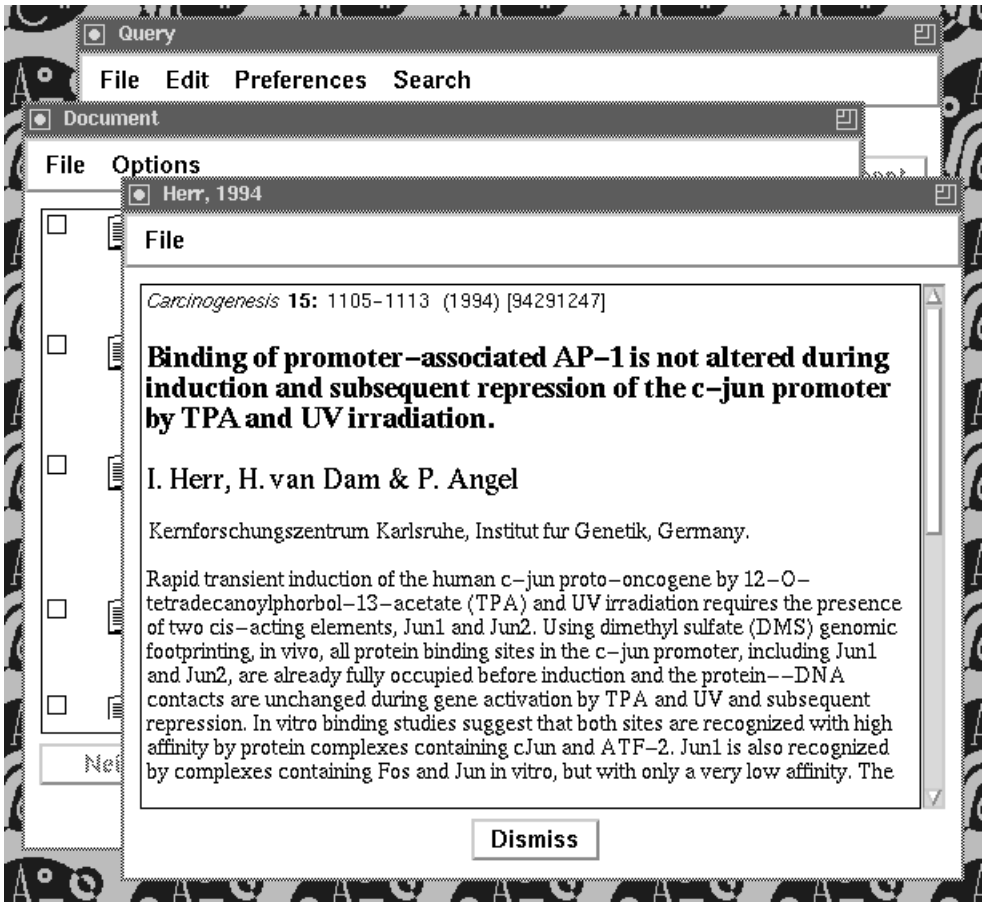
Query Refinement サブウィンドウには、検索結果として各キーワードがそれぞれ何件あったかが示されている。Special というのはこの場合 Title フィールド中にこのキーワードをもつ

件数、 Total は Abstract フィールドでの出現も併せたものである。これらの集計結果が左下の Retrieve ボタンの中に示されている。この件数は、黒く反転したところ、ここでは３つのキーワードが Abstract と Title のどこかで出現したものの AND 結合、つまり３つのキーワード全てを備えたエントリーの件数である。デフォールトでは Entrez は各キーワードの AND 結合を求める。

この結果の詳細を見るには、件数が表示された Retrieve ボタンを押す。すると次のようなウィンドウが新たに現われる。



このリストウィンドウでは Retrieve ボタンに出ていた件数分のリスト全てがのっている。ウィンドウ右側にあるスクロールバーをマウスで上下に操作することで、表示される部分を動かせる。ご覧になって分かるとおり、文献の First Author, 出版年、タイトルが見えている。ここでさらに詳細に見たいときには、見たい文献の上にカーソルを合わせてマウスの選択（左）ボタンをダブルクリックする。例として先頭のエントリーを見てみよう。

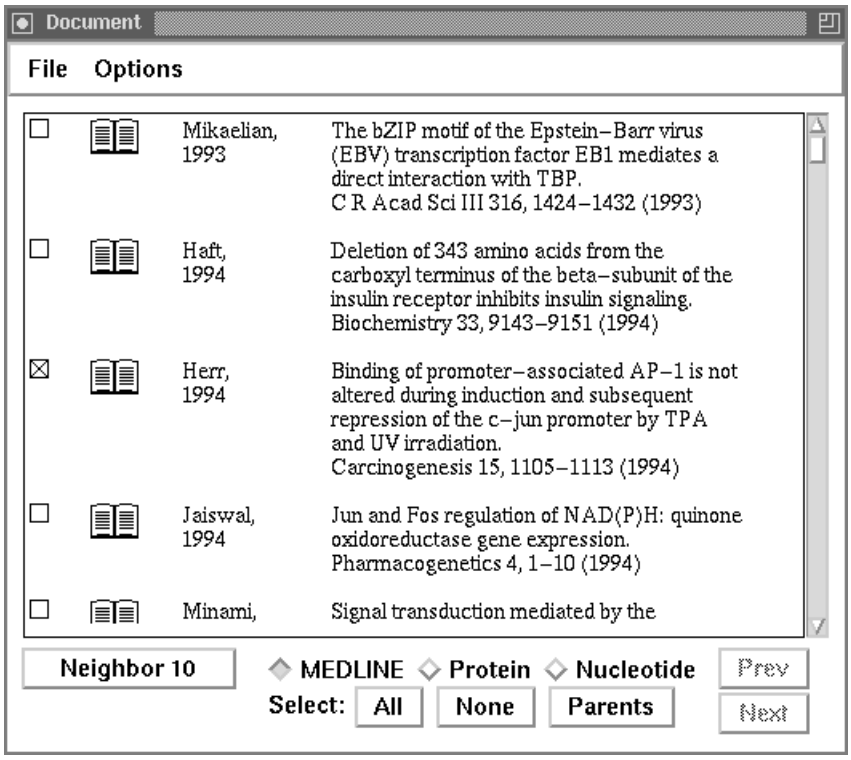


エントリー内容表示のウィンドウでは、あたかもジャーナルの最初の頁を見るように、雑誌名、巻、頁、タイトル、著者名、アブストラクトが表示され、このあとに Medline 独自のキーワードである MeSH Term などが続いている。このウィンドウも左側のスクロールバーを操作して見える範囲を上下に動かして読むことができる。

3　関連項目を辿る：同一データベース内の関連事項

Entrez の 1 つの大きな特徴に、1 つの検索項目から関連事項を芋蔓式に辿る機能がある。先ずは同一データベース内、この場合は Medline 内で関連項目を探してみよう。文献の場合、関連項目とは、あらかじめ調べられている語句の出現統計に基づくキーワード同士の類似度から、近いものを辿って行くという操作になる。

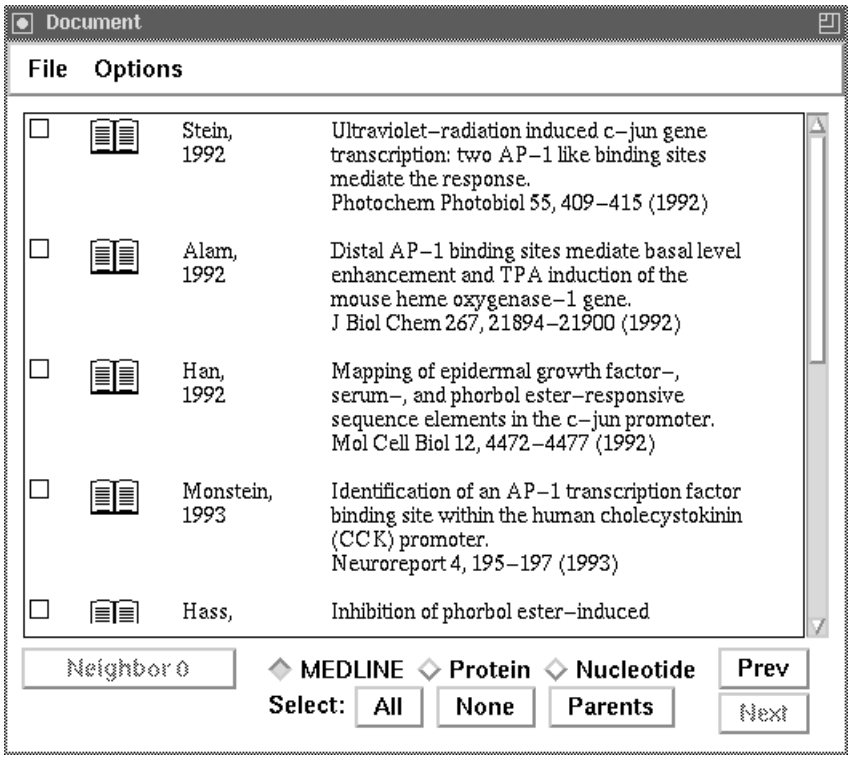
先ほどのリストウィンドウに戻って、今度は 2 番目の文献に関連するものを引いてみる。それには、各文献の先頭にあるチェックボックス (□の部分) をマウスでクリックする。するとこの中に×マークがつくと同時に、右下に Neighbor いくつ といったボタンが現われる。



このボタンを押すと、次のようにリストウィンドウが変わり、選んだ文献に関連するものが表示されている。この内容を見たい場合は先ほどと全く同様に目的の文献の上でダブルクリックすればよい。また、１つ前のリストに戻りたいときは、右下にある Prev ボタンを押せばよい。

この関連項目検索は１つだけでなく、１回に複数の文献を同時に対象として探すことも可能で、その場合は必要な数だけチェックボックスをつけて行けばよい。チェックボックスは１回クリックする毎にマークがついたり消えたりする。また一挙に全てをマークしたければ下にある Select: All ボタンを押せばいいし、逆に全て消したいなら None を押せばよい。

なお、今の例は文献データベースであったが、配列データベースでも同様な関連項目検索ができ、この場合、関連項目はあらかじめ計算された配列のホモロジーによっている。



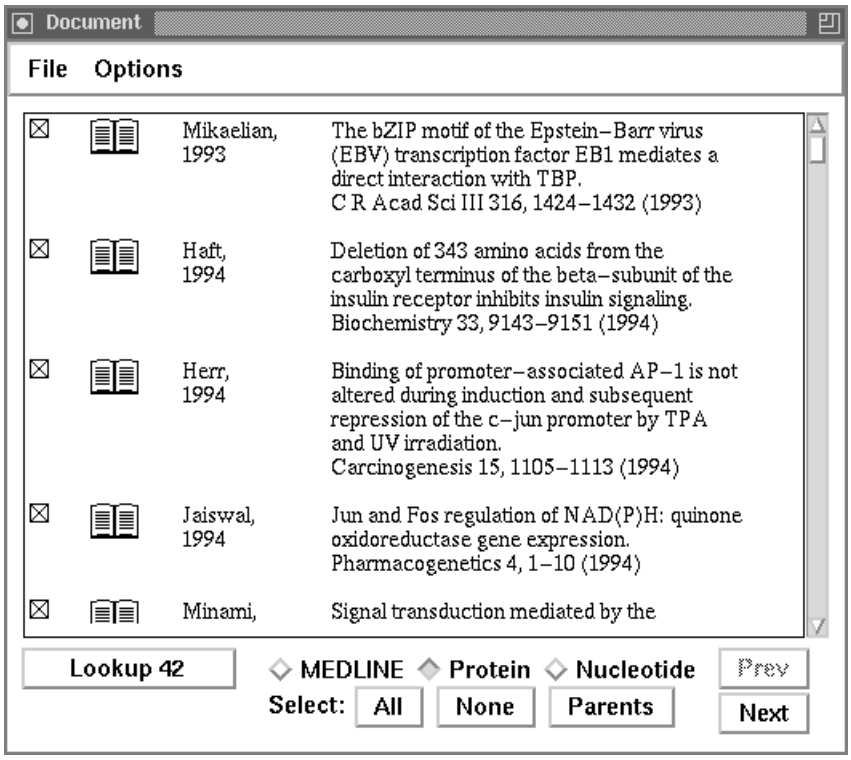
4　関連項目を辿る：データベースにまたがる相互参照

同一データベース内だけでなく、文献から配列データベースへ、またその逆へといったデータベースにまたがる関連項目検索も可能である。Entrez のデータベースは文献は Medline をベースにし、また配列データベースとしては核酸配列と蛋白質配列を備え、こちらは GenBank, EMBL, DDBJ, PIR, SWISS-PROT, PDB, PRF といったメジャーなデータベースのほとんどが網羅されている。これらは文献情報フィールドや FEATURES テーブルなどに書かれているような情報をもとに相互に参照づけがなされ、あらかじめリンク情報として蓄えられている。このリンクを辿ることで相互のデータベースに行き来できるようになっている。

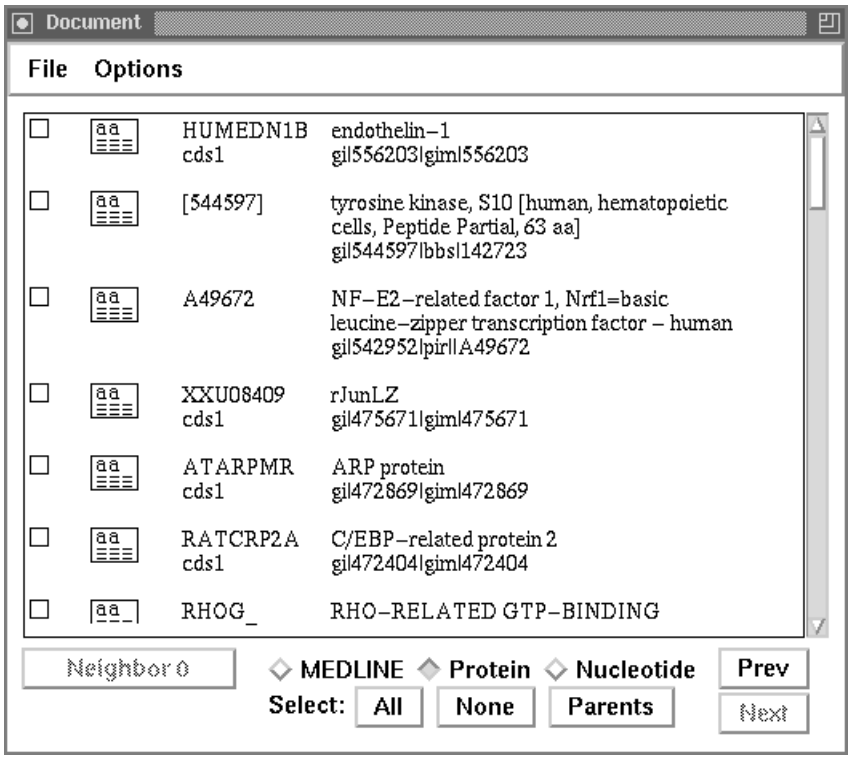
例として、先ほどの最初のリストウィンドウであげられている全ての文献に関連する蛋白質配列情報を探してみよう。まず、ウィンドウ下の中ほどにある

◇ MEDLINE　◇ Protein　◇ Nucleotide

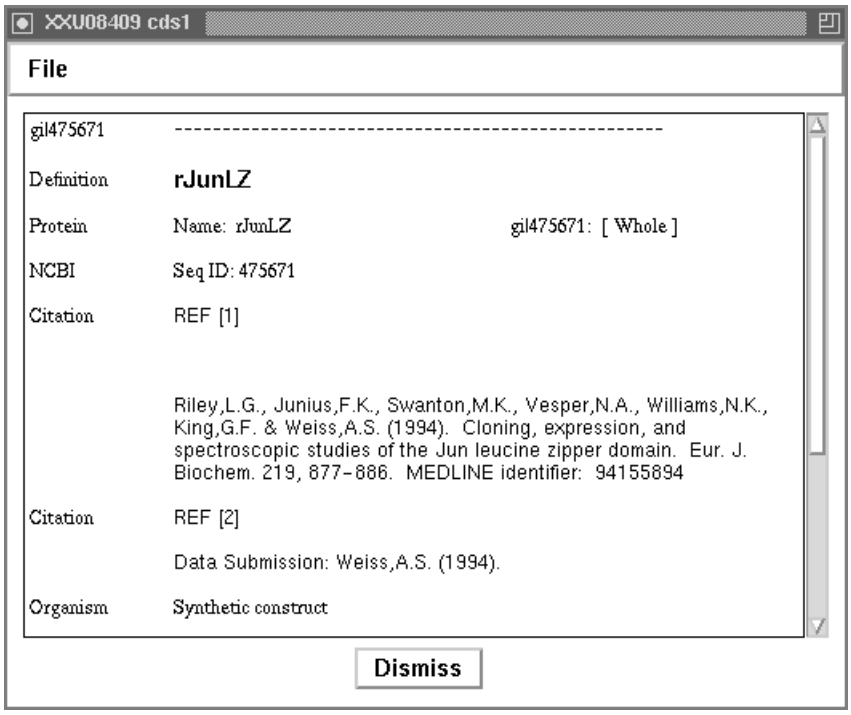
と書かれたボタンで (今は MELINE がマークされているが)、Protein と書かれたところをクリックすると、今度は Protein にマークが移る。そこで、次に Select: all ボタンを押して全項目にマークを付けよう。すると今度は左下に Lookup 何件 というボタンが現われる。



ここで Lookup ボタンを押すと次のようなリストウィンドウに変わる。



文献の時と若干見かけは異なるが操作上は変わることがなく、見たいエントリーがあったらそこでダブルクリックすればよい。

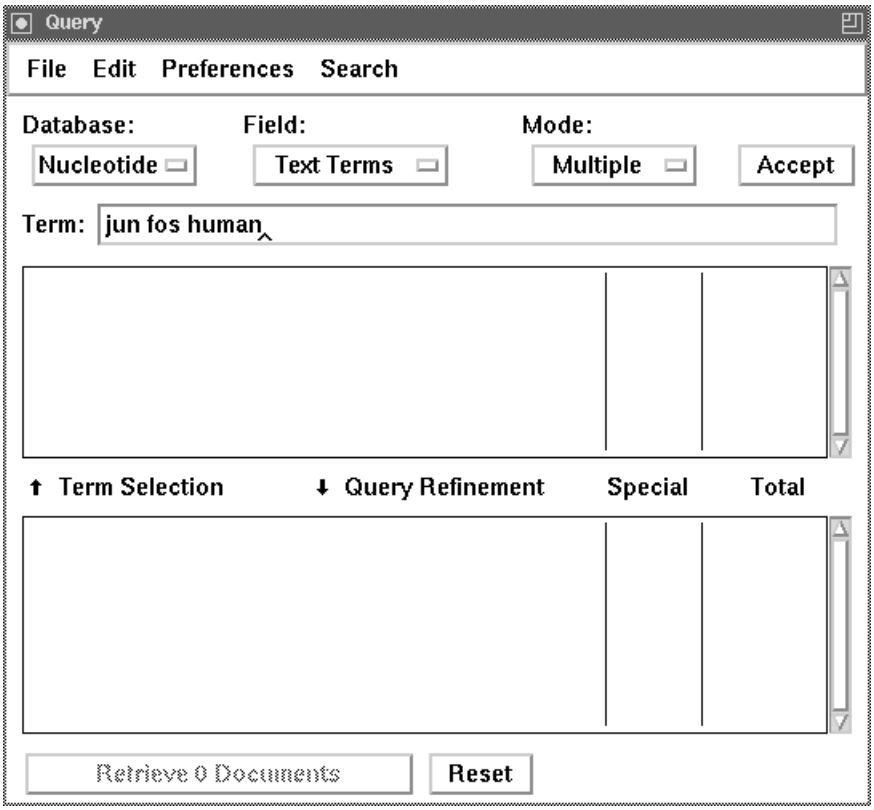


5 検索条件の変更：OR 結合

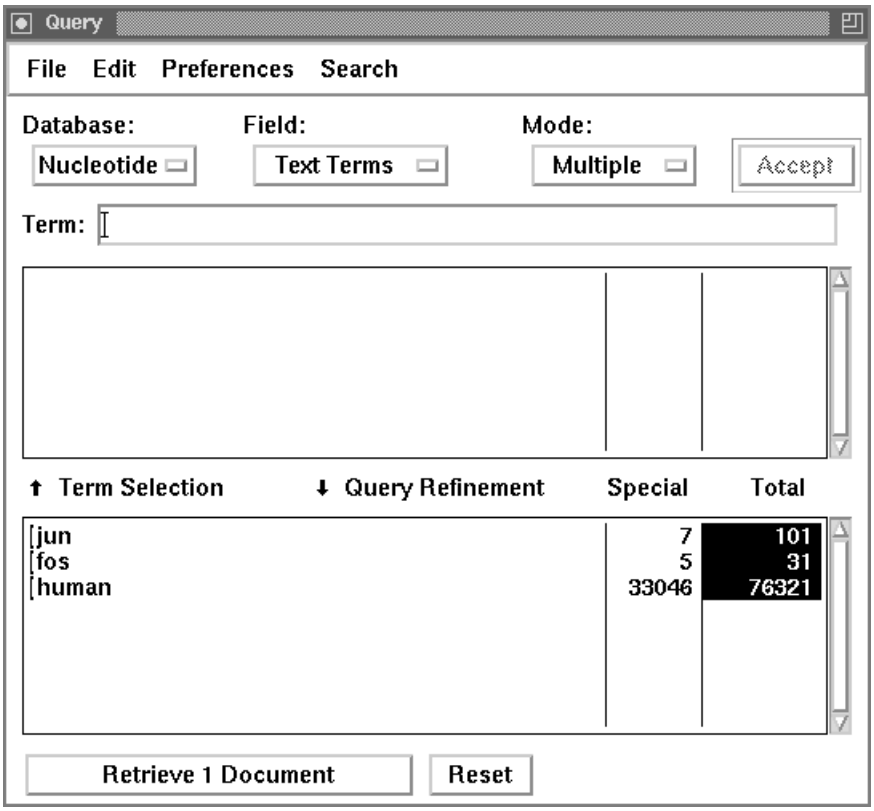
一通り Entrez の中をナビゲートする方法が分かったので、次に検索条件の修正について述べる。その前に、不要なウィンドウを全てクローズして最初の Query ウィンドウに戻り、そこで下の真ん中にある Reset ボタンを押して先ほど入れた問い合わせをクリアーしよう。

ついでにデータベースも変えてみる（図は略）。Database: の下のボタン（今は MEDLINE と表示されている）の上にカーソルを合わせ、マウスの選択（左）ボタンを押し続ける。するとプルダウンメニューが現われ、MEDLINE, Protein, Nucleotide と書かれている。マウスボタンを押しながらカーソルを下げて行き、Nucleotide についたところでマウスボタンを離すとデータベースが切り替わり、ボタンにも Nucleotide と表示される。それと共に Field: のところも変わり、Text Terms となる（Field 選択もデータベースに応じていくつかの選択肢が用意されている。ここの変更もデータベースの変更の仕方と同じ）。

用意ができたところで先ほどと同様のキーワード、jun␣fos␣human と入れてみる。



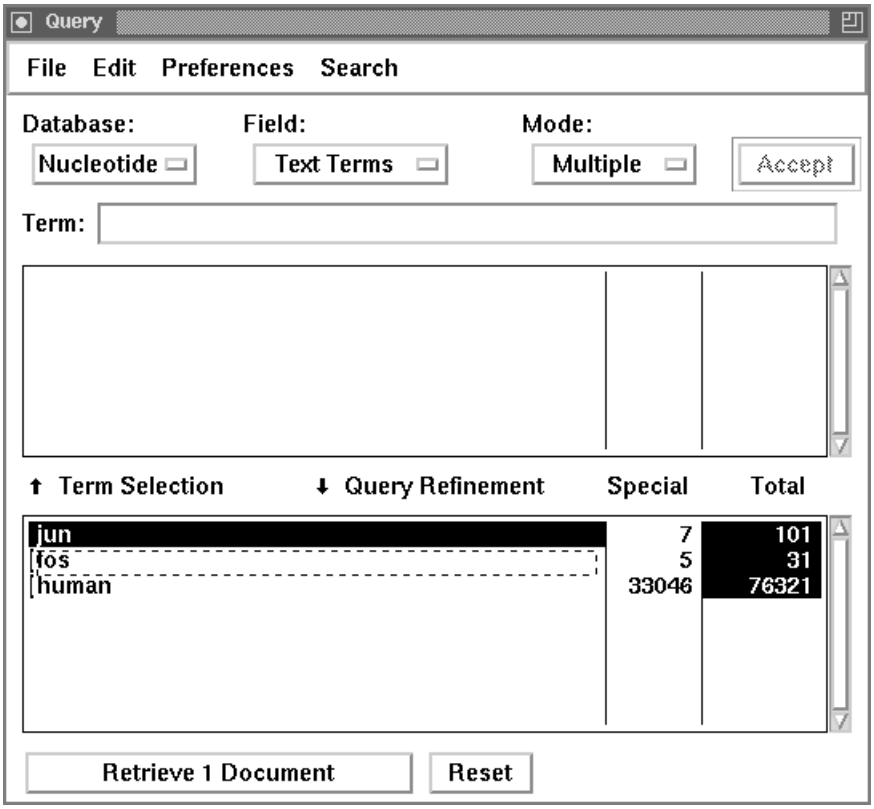
ここで Accept ボタンを押すと先ほどと同じように検索が進むが、今度は最終的には Retrieve 件数が 1 しかなかった。



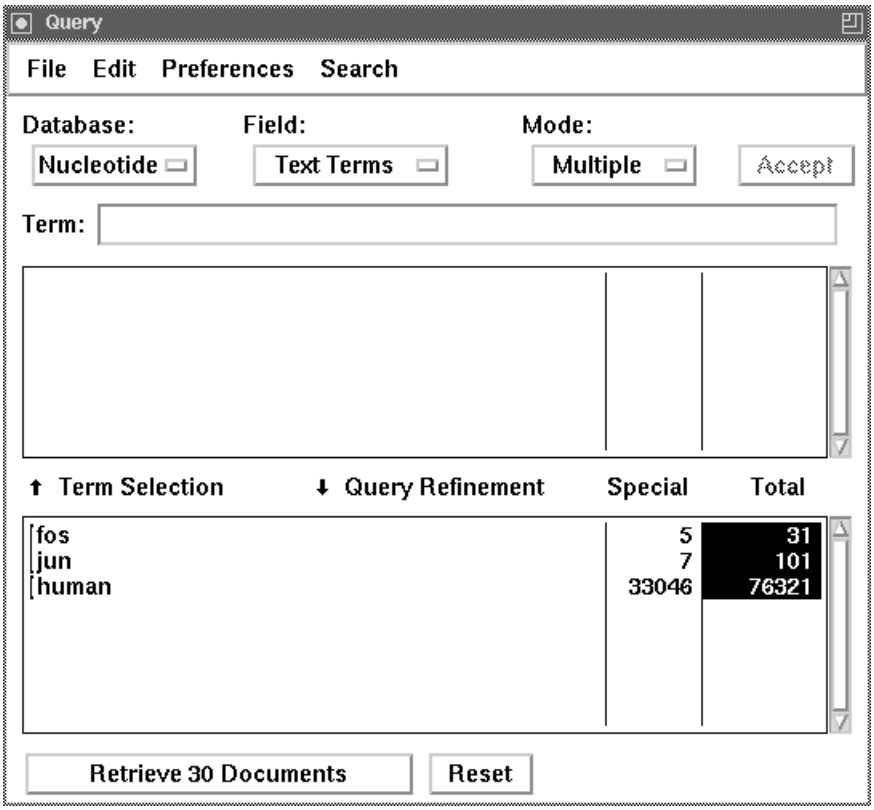
つまり、 jun かつ fos として登録された human の塩基配列エントリーはないわけで (ここで

見つかった例は引用されている文献タイトルに Jun Fos があった）、そこでこの jun かつ fos のところを jun 又は fos に変えてみよう。

まず、下半分の Query Refinement ウィンドウの jun の上にカーソルを合わせ、マウス選択（左）ボタンを押す。次に押し続けたまま静かに下にずらして行き、点線で描かれた箱が fos と重なったところでマウスボタンを離す。

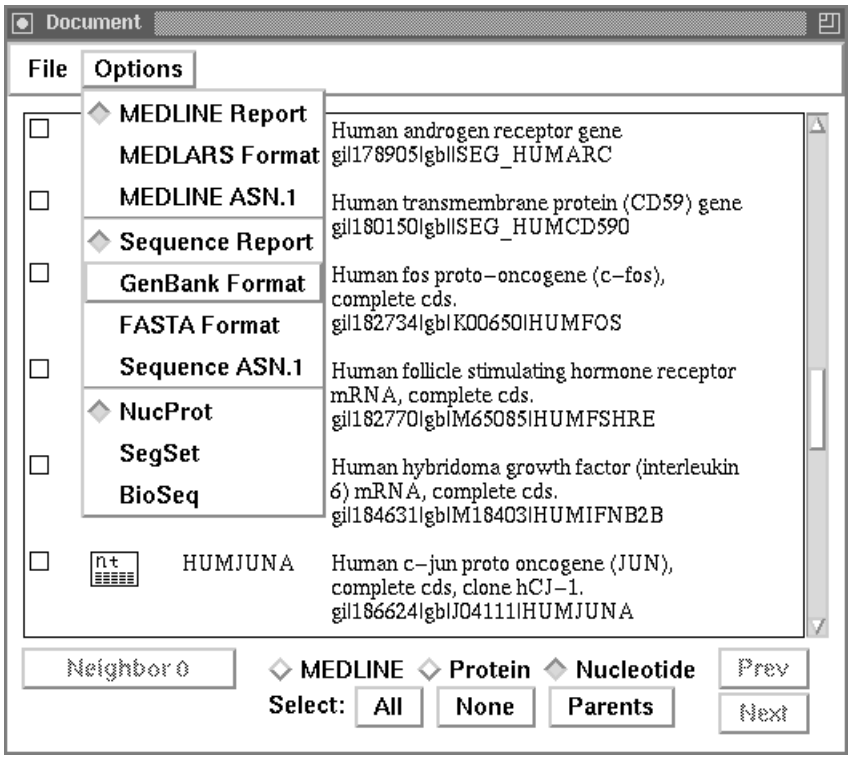


Query Refinement ウィンドウ内の各キーワードの先頭についている角括弧に注意。重ねる前は全てのキーワードに独立についていたが、うまく jun と fos を重ねると、 jun fos の前の角括弧は 1 つながりになる（次の図）。これと同時に、（ 'jun' or 'fos') and 'human' の条件にマッチするエントリーがあったために Retrieve 件数が表示される。前と同様に Retrieve ボタンを押せばリストウィンドウが現われる。



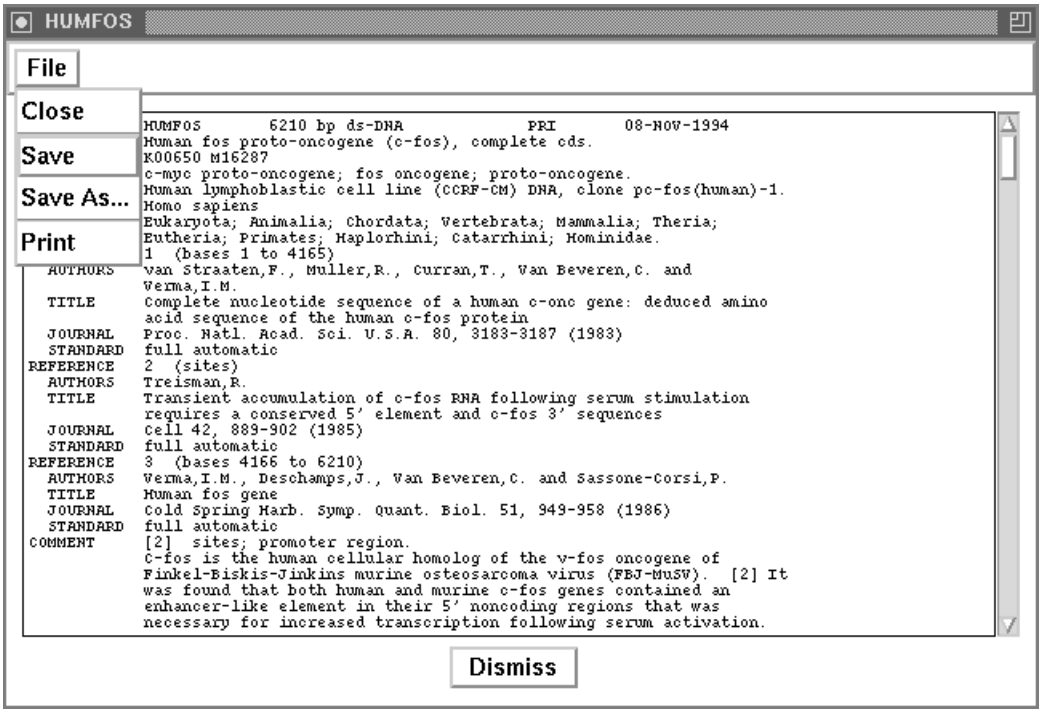
このように、 Entrez ではキーワードの結合条件は Query Refinement ウィンドウのキーワード先頭の角括弧の範囲で示されるが、そのうちどれを選択するかは Special または Total のところで選択されている (反転表示されている) 部分を変えることで選べる。これは単に目的のフィールドをクリックすれば選択・非選択が交代する。今の例では (‘jun’ or ‘fos’) and ‘human’ であるが、たとえば human の選択部分をクリックして非選択状態にすると、検索条件は ‘jun’ or ‘fos’ だけになる。

再び (jun or fos) and human に戻り、先ほどと同様に Retrieve ボタンを押してリストを表示させた後、さらにここからエントリーを呼び出そうと思うが、このままの設定で表示される Entrez の配列レポート形式はやや馴染みがないので、 GenBank 形式に表示を変えよう。そのためにはあらかじめリストウィンドウで出力形式を選んでおかなくてはならない。次の図のように Options のところでマウス選択ボタンを押すとプルダウンメニューが出るので、そこで GenBank Format を選ぶ。



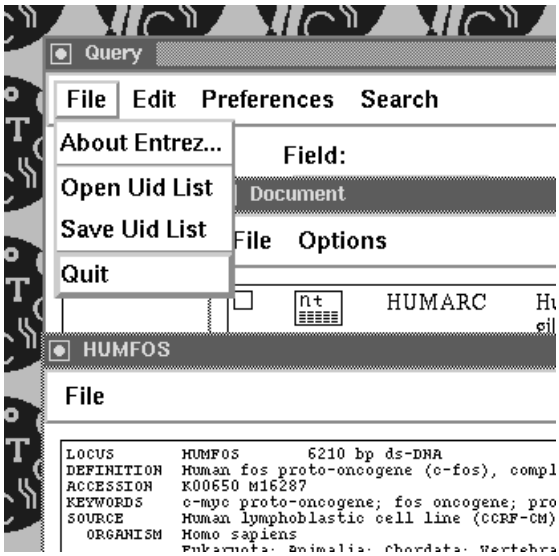
6　ファイルへの保存

ファイルへの保存は極めて簡単で、エントリー表示ウィンドウの File メニューの下の Save を選ぶだけでよい。 Entrez を立ちあげたディレクトリに、そのエントリー名 (この例では HUMFOS) を名前としたファイルが作られる。



7　終了

最後に終了するには、メインの Query ウィンドウにある File メニューで Quit を選ぶだけである。 Entrez で作られた全てのウィンドウが消えて、立ちあげ前の状態に戻る。



8　HGC 外部からの無記名利用資格によるアクセス

Entrez もまた HGC 利用資格をもたない人のために無記名での利用手段を提供している。ただし条件として、 Internet に十分な回線速度で接続されていることと、 X ウィンドウが使えることである。

利用の仕方は telnet でアドレス entrez.genome.ad.jp に接続し、ログイン名 entrez で入るとパスワードチェックフリーで Entrez のみが利用できる。

```
local{user}61: telnet␣entrez.genome.ad.jp
Trying 133.11.196.1 ...
Connected to hgc.tokyo-center.genome.ad.jp.
Escape character is '^]'.
 SunOS UNIX (hgc)

 login: entrez
Last login: Tue Jan 24 20:27:24 from hgc
SunOS Release 4.1.2-JLE1.1.2 (DISKSUITE) #2: Wed Aug 24 16:57:12 JST 1994
                  *** Welcome to Entrez Server ***


                             * Entrez *

                    < Document Retrieval Software >
                     Served by Human Genome Center,

        Institute of Medical Science, The University of Tokyo
       Warning:  This service needs X window system on your machine!!!

              * If you have any problem please e-mail to:
                          trouble@genome.ad.jp
         95/01/18    Entrez: Databases updated.(Release 14.0, December 15)

                       GenBank      Release 86.0
                       EMBL         Release 40.0 plus updates
                       DDBJ         November 26, 1994
                       SWISS-PROT   Release 30.0
                       PIR          Release 41.0
                       PDB          Release 69.0
                       PRF             Nov. 1993
                       dbEST        Release  2.41
                       U.S. and European Patents
                       MEDLINE - Sequence Associated Subset: Nov. 1994
       First, enter "xhost 133.11.196.1" on YOUR CONSOLE.

      Hit return to continue (enter "q" to quit)
       Enter your machine address [local]

       Setting the DISPLAY variable as follows:

        setenv DISPLAY local:0.0
       Hit return to continue (enter "c" to correct)


      Entrez system is now running... please wait.
```

後の利用法は HGC 内部の場合と同じである。ただし、ファイルに保存を行った場合はその結果を電子メールで送り返すかどうか最後に尋ねられるので、自分の電子メールアドレスを入力する。

```
You have saved files. Enter your e-mail address to receive them.
(Hit return if you don't need them): user@ims.u-tokyo.ac.jp
Sending mail program start...
Sending File Size: 2287 bytes.
  Hit return to confirm(enter "s" to skip):
  sent.
Completed.
Connection closed by foreign host.
local{user}62:
```

9　参考．検索モード

検索モードはどのフィールドが対象として選択されているかによって使えるものが異なるが、Multiple, Selection, Truncation, Lookup の４種類ある。

Multiple は例で説明したように、各キーワードはスペースで区切られ、それらは独立に扱われて各々についての検索が自動的に行われ、その後で AND（または OR）結合される。

Selection では与えられた問い合わせに対して Multiple のように自動的に最適のキーワードを選び出さずに、Term Selection サブウィンドウに類似のものを全て表示し、利用者がその中から選択するのを待つ。

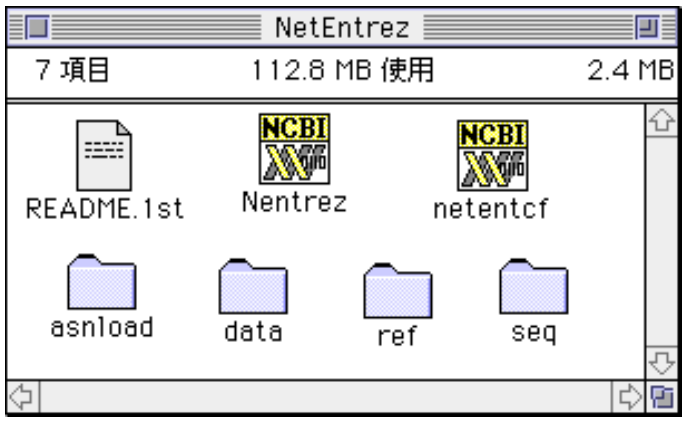
Truncation は選択されているフィールドによって若干意味が異なるが、基本的には与えられた問い合わせ語句で始まるような部分文字列をキーワードとしてもつものが探される。さらに Author Name のようなフィールドが選ばれているときは、スペースは family name と first name の区切れのように扱われ、キーワードの区切れという意味ではなくなる。

Lookup モードは MEDLINE ID が検索対象の時だけ有効で、問い合わせ ID をもつものが探される。

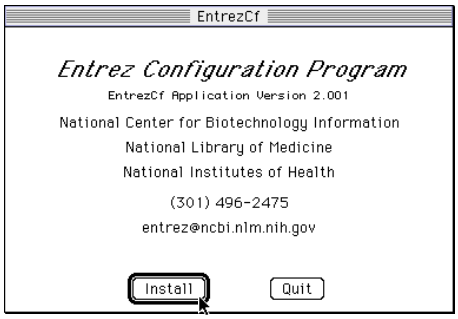
10　補足 Macintosh からネットワーク版 Entrez を使う

今まで解説した Entrez は、ゲノムセンターのサーバーコンピューター上に Entrez プログラムとデータベースの両方が置いてあったが、Macintosh などから直接利用できるネットワーク版 Entrez (NetEntrez, NEntrez) もあり、これは各利用者が直接向かっている計算機にフロントエンド部分のみを置き、データベースへのアクセスはネットワークを経由して直接 Entrez を作成しているアメリカ NCBI のマスターサーバーを参照するようになっている。現在のところ、NCBI の NetEntrez サーバーに接続するには、利用するネットワーク単位で事前に登録が必要であるが、医科研の場合すでに一括して登録してあるのでマッキントッシュ用のフロントエンドプログラムを入手して付属の設定プログラムで初期設定を行うだけでよい。このプログラムはゲノムセンターからも配布している（１０章：配布ディスクについてを参照）。また、ネットワークを利用するためにインストールするマックは UTnet に接続され、MacTCP がインストールされていないといけない（１０章参照）。

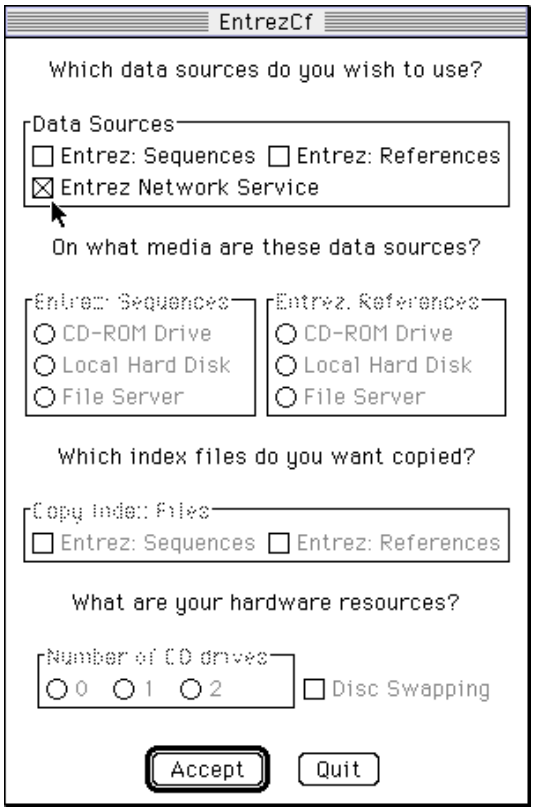
マック用 NetEntrez を入手して展開すると右のようなファイルが得られる。**Nentrez** が Entrez プログラム本体、**netentcf** は初期設定プログラムで、初めにかならず１回、またネットワークなどの構成が変わったならばその都度実行しなくてはならない。Entrez の実行自体はすでに述べたところと同じなので、以下に初期設定の方法を示す。



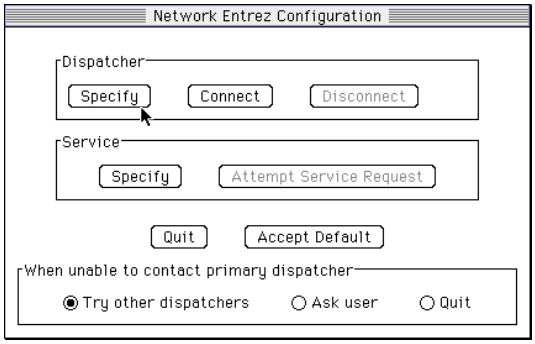
netentcf をダブルクリックすると次のようなウィンドウが現われるので、Install ボタンを押す。



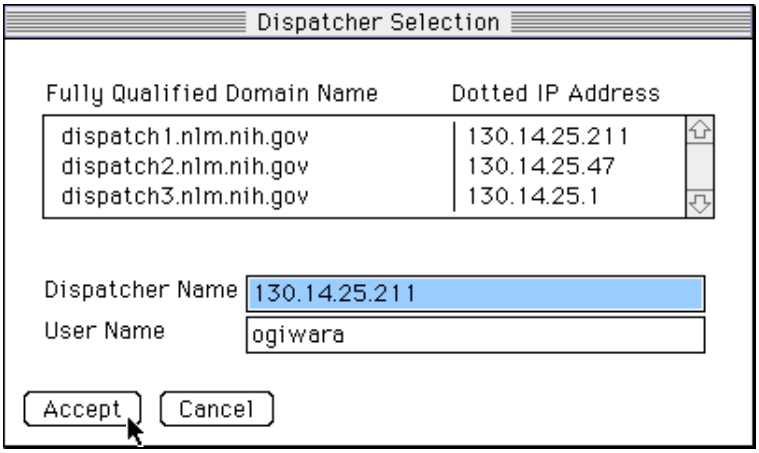
右のようなウィンドウが現われるので、Data Sources の Entrez Network Service だけにチェックマークを付けて Accept を押すと、下のようなウィンドウが表示される。



Dispatcher の Specify を押して、次のようにアドレスが設定されていることを確認する (左下図)。

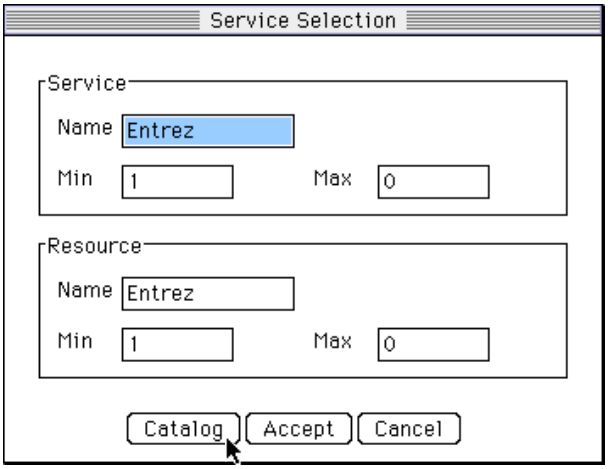


アドレスを確認し、User Name に自分の名前を入れたら Accept ボタンを押す (右下図)。



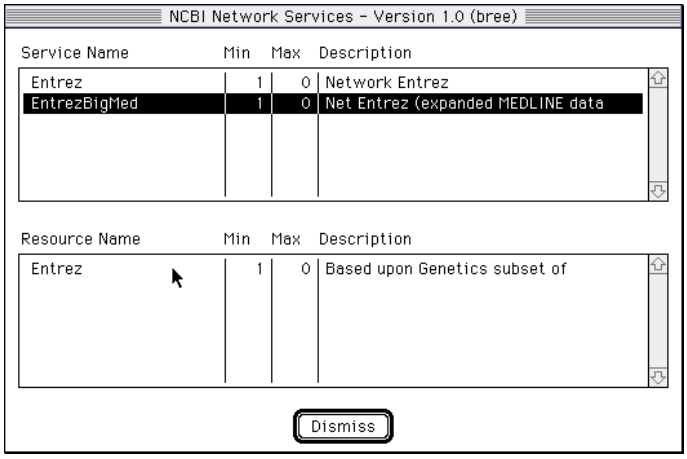
再び 1 つ前のウィンドウに戻るので、Dispatcher の中の Connect ボタンを押す。しばらくすると Connect が消えて Disconnect がつくので、次に Service の中の Specify を押す。

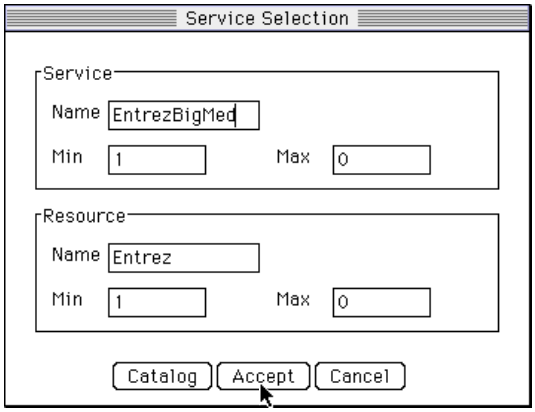
そこで、もし右のように Service がただ Entrez となっていた場合は、Catalog ボタンを押す。Service が EntrezBigMed になっていればそのまま Accept でよい。



EntrezBigMed は通常の CD 配布版よりもずっと多量の Medline エントリーを検索することができる。これはネットワーク版だけの利点である。

Catalog ボタンを押すと左下のようになるので、 Service Name から EntrezBigMed を選び、次いで Resource Name で Entrez を選ぶ。再び右下のように 1 つ前のウィンドウに戻るので、そこで Accept を押す。

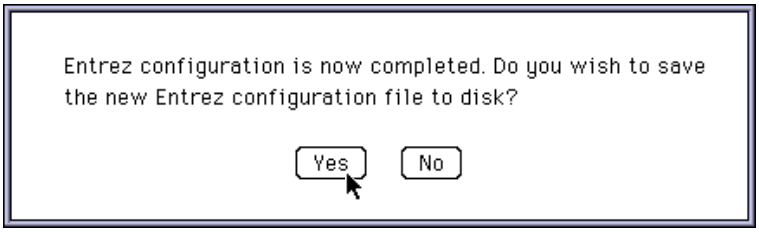




もう 1 つ前の画面に戻るので、そこで Service の中の Attempt Service Request ボタンを押す。

しばらく待って右のようなダイアログボックスが表示されたら接続は成功しているので Yes ボタンを押す。

ここでも Yes を押して設定ファイルを保存すれば設定プログラムは完了する。



なお、ここでうまく設定ができなかった場合、MacTCP Ping などで ncbi.nlm.nih.gov. と通信できるかを試してみるとよい。これができない場合はネットワークの途中経路で障害が発生して通信不能の可能性があるので、しばらく待ってから再度設定してみていただきたい。