パスカルデータベースシステム (I)
Pascal Database System (I)

白石 修二
Shuji SHIRAISHI

福岡大学 理学部 応用数学科
Department of Applied Mathematics, Faculty of Science, Fukuoka University

藤村 丞
Shou FUJIMURA

長崎大学 経済学部 総合経済学科
Faculty of Economics, Nagasaki University

あらまし: パスカルデータベースは、メナール編纂パスカル全集既刊本全4巻「ŒUVRES COMPLÈTES DE BLAISE PASCAL」(Desclée de Brouwer) の全テキストをデータベース化したものである。2000年10月からインターネットで試験公開中である。インターネットを介してのいくつかの語彙検索や頻度作成ができるようになっている。検索に関しては、ヒットした文を含む原著でのページ画像も表示できるようになっている。頻度表作成に関しては、ABC順、多寡順の選択ができ、逆引き辞書の作成も可能である。本稿では、パスカルデータベースシステムの概要と使用法について述べる。

Summary: We experimentally release an online database on "ŒUVRES COMPLÈTES DE BLAISE PASCAL"(Desclée de Brouwer publishers) in 4 volumes edited by Jean Mesnard. This database can be accessed through Internet from all over the world, and you can use it for vocabulary searches, making frequency lists , etc. In this paper, we sketch the summary and the useage of Pascal Database System.

キーワード: パスカル, データベース, メナール

Keywords: Pascal, database, J. Mesnard

1 はじめに

1997年、文部省科学研究「人文科学とコンピュータ」の研究資金援助を受けて、当研究グループによるパスカルデータベース構築プロジェクトが本格化した。当初の予想ではテキストデータベース構築はさほどの問題なく進むだろうと考えていたが、そこまで簡単ではなかった。この完成に至るまでにはさまざまな問題が生じ、そのひとつひとつに解答を与えることで着実に構築をめざしてきた。1998年度特定領域研究「人文科学とコンピュータ」報告集 [6] を参照。研究着手から4年を経て、2000年3月にほぼ現在の形にまで完成させることができた。研究成果を多くの研究者に還元することを最終的な目的としているが、その試験的な公開を10月よりインターネットを介して行っている。このインターネット上での試験公開の目的はいくつかある。その一つは、利用者の利便性が最大限に図られているかどうかということの検証である。つまりこのシステムが利用者（研究者）の要求に、でき得る限り答えることができるかどうか、という能力を量ることを意味しているのだが、具体的にはそういう利用者の利便性を追及するために、インターフェースを使いやすくすることを心がけ、また同時に柔軟で強力な検索システムを与えるなど、いろいろな工夫を加えている。インターネット上で公開し、実際に多くの人に利用してもら

うことで、このシステムの利便性に対する配慮が実際上機能するかどうかを確実に把握することができると考える。試験公開の目的のほかの部分は、このような実際の利用を通じて、データベースの記述上のミスがあれば、その発見に大きく貢献するだろうし、ひいてはデータベースの精度を高めることにもつながるものである。しかし、この公開に関して注意を払わねばならないことはこれにとどまらない。つまり、利用者の便宜を最大限に図ることを主眼におかねばならないし、おきたいが、利用者が自由自在にデータを操れるようにすることは、オリジナルデータが大量にコピーされる可能性を生み、データベース作成に関する知的財産所有権の侵害を図らずも許してしまう可能性を示唆している。このことに関して、利用者のためには無制限の公開が望ましいし、それこそがデータベース公開の本領だと思われるが、そのような無制限の公開は、デジタルで作成されたものの運命として簡単にコピーし得るという弱点を持ち合わせていることとあわせると大きなジレンマとなるのである。

そういうジレンマを少しでも解消するべく、利用者にはやさしく、しかし製作者の保護も考えたシステム製作を模索しており、現システムにおいては、情報ではなく、情報を引き出す方法において負荷を与えることで知的財産所有権の侵害への抑止策として考えている。

さらにネットワークの速度あるいはコンピュータの能力が充分かどうか、というハードの部分での効率を確かめるという意味がある。

システムは、現在、パソコン上で維持管理及び、保守がなされているが、利用状況及び、改善すべき点に関しては、データの蓄積に伴い追って報告の予定である。偉大なパスカルの多岐に亘る業績の全てを考えたとき、後世に残したい自然や建造物を指定し保存し維持しようという趣旨のもとで設定されている＜世界遺産＞のような観点をもって、卓抜したその業績と思想に対して、その全ての保存と維持に勤め、後世に残す義務があると考える。そこで、ここまでの完成をみたテキストデータベースを一歩先に発展させ、これを中心に据えた＜パスカル総合資料館 ＞(パスカルデジタルアーカイブ) の構築を行うことがこのデータベースの次なる使命と考えている。更にこれに引き続いて、この＜パスカル総合資料館＞もインターネット上での公開を目指したいと考えている。このパスカル総合資料館の早期公開の実現のために、パスカル関係資料のあらゆるものに関しての大規模で組織的な収集を予定している。

関係資料収集以外にこのデジタルアーカイブ構築のための準備を平行して行うこととなるが、テキストデータベースと手書きの原稿、写本等の画像データベースを統合する仕組みの研究に着手するとともに、柔軟でかつ強力な検索システムも考案中である。

2　システム概要

2000 年 3 月にメナール版パスカルデータベースシステムが完成。2000 年 10 月から データベースをインターネットで試験公開中である。アドレスは、
http://pascal.sci.fukuoka-u.ac.jp:8080/
である。

システムは、パソコン上に Java と Oracle を用いて構築している。データベースメンテナンスに関しては、アクセス (Microsoft Access) でリモート処理できるようになっている。

＜システム構成＞

コンピュータ本体:
DELL PowerEdge 1300
PentiumIII 500MHz
メインメモリ 256MB

サーバ OS:
Turbo Linux 4.2 ftp 版

Web サーバ:
Apache.1.3.6.tar.gz

データベース:
Oracle8 Workgroup Server for Linux R8.0.5

Java に関しては、次の 2 つのパッケージを使用した。
Java 開発キット:
jdk_1.1.7-v3-glibc-x86.tar.gz
(Java Development Kit)
Java サブレット開発キット:
jsdk2.0-solaris2-sparc.tar.z
(Java Servlet Development Kit)

3　データベース

データベースは、既刊本全 4 巻の全文テキストデータベースと見開きの 2 ページを単位とした画像デー

タベースからなる。テキストデータベースの容量は凡そ 14.6MB で、画像データベースは 176MB である。

3.1 テキストデータベース

次は、パスカルデータベースの電子テキストサンプルである。電子テキストには文頭に、文の区切り語 # と同時に原著での位置が分かるようにボリューム番号 (巻数) とページ数、行数が記されている。多段組のページがあるので、列数も入れている。例えば、#2-0232-11-00-P は第 2 巻 232 頁の 11 行目を意味している。P はパスカル本人が書いた文を表す。00 は列数が 1 段組であることを表す。全集には、姉妹やパスカル周辺の人々の書簡集及び編者メナールによる註も含まれているので、ここで区別できるようになっている。したがって、検索時にパスカル本人が書いた文だけを対象にしたり、あるいは姉だけの書簡集に限定することができる。また、パラグラフの区切りとして、記号 @ を入れている。文中での \e'\ 等は、フランス語の è を表している。詳しくは、4.5 変換対応表を参照。

電子テキスト例

```
@#2-0231-38-00-D 4. C'est-\a'\-dire,
d'apr\e'\s la d\e'\finition prcmi\e'\rf,
concourantes. En d'autres termes.
le point d'intersection des droites
NO et PQ se trouve sur la droite MS:
les c\o^\t\e'\s
@#2-0232-01-00-P LEMME \!R2\
@#2-0232-02-00-P Si par la m\e^\me droite
passent plusieurs plans, qui soient
coup\e'\s par un autre plan, toutes les
lignes des sections de ces plans sont de
m\e^\me ordre avec la droite par
laquelle passent lesdits plans.
@#2-0232-06-00-P \L*\ Ces deux
lemmes pos\e'\s, et quelques faciles
cons\e'\quences d'iceux, nous
d\e'\montrerons que, les m\e^\mes choses
\e'\tant pos\e'\es qu'au premier lemme,
si par les points K, \!R5\, passe une
quelconque section de c\o^\ne qui coupe
les droites MK, MV, SK, SV,
es points O, N, Q, les droites
MS, NO, PQ, seront de m\e^\me ordre.
#2-0232-11-00-P Cela sera un
troisi\e'\me lemme.
@#2-0232-12-00-P En suite de ces trois
lemmes et de quelques
cons\e'\quences d'iceux, nous donnerons
des \E'\l\e'\ments coniques complets,
\a'\ savoir toutes les propri\e'\t\e'\s
des diam\e'\tres et c\o^\t\e'\s droits\L2\,
des tangentes, etc., la restitution du
c\o^\ne presque sur toutes les donn\e'\es,
la description des sections de
c\o^\ne par points, etc.
```

3.2 画像データベース

既刊本全 4 巻を見開き 2 ページをひとつの単位として、全て画像データベース化した。検索された文章を原著で確認できるようになっている。原著に含まれる数学や物理関係の図等は、テキスト化していないので、当面このような処置をとっている。図 1 は、画像データの例である。閲覧時、コピー防止のため PDF 形式とした。PDF は Adobe Acrobat のファイル形式で、Portable Document Format の略である。

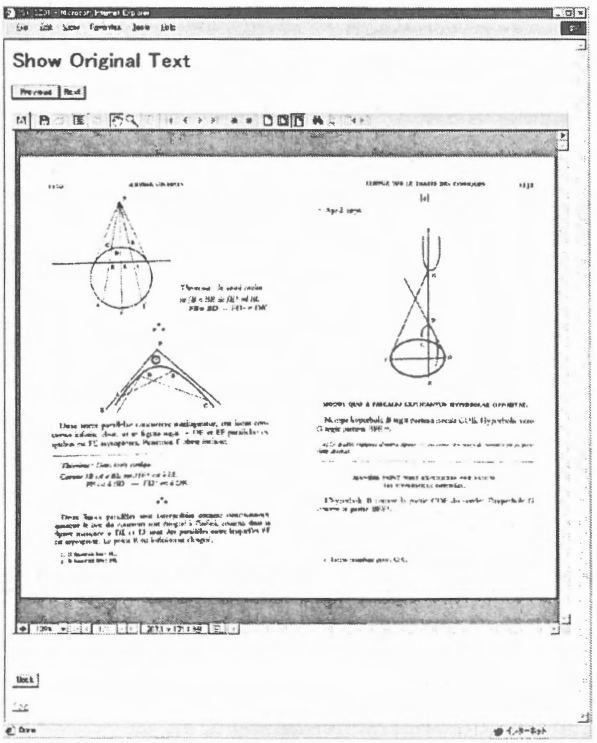

図 1: 画像データベース

4　データベース使用方法

4.1　試験公開要項とパスワードの取得

パスカルデータベースを使用するには、試験公開要項をよく読み、パスワードを入手する必要がある。試験公開要項は、現在、日本語版、韓国版、英語版、ドイツ語版が用意されている。順次他の言語での要項も準備中である。試験公開要項及びパスワードの入手方法に関しては、
http://pascal.sci.fukuoka-u.ac.jp:8080/
にアクセスする。例として、英語とドイツ語の試験公開要項を上げている。図 2,3 を参照。

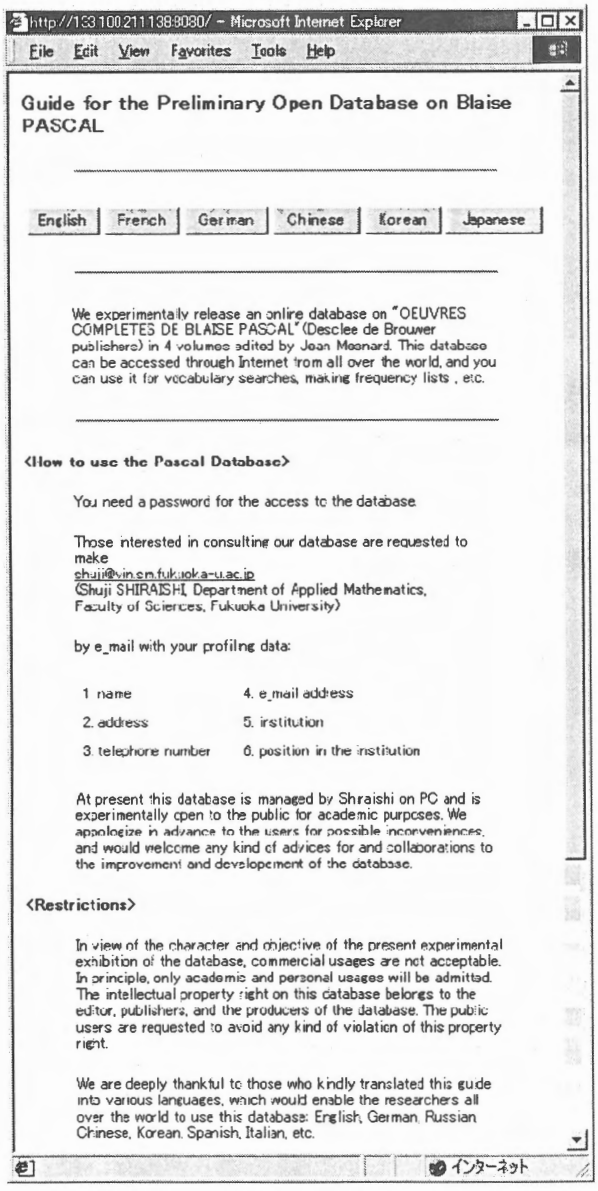

図 2: 試験公開要項 (英語)

試験公開要項には、利用要項、必要連絡事項、制限事項が簡潔に記されている。このデータベースの使用については、実験公開の目的からいっても、商業的な利用の不可、そして原則として個人の使用のみとしている。また、個人の資格としても真に学術的な目的をもつ利用者に限定している。

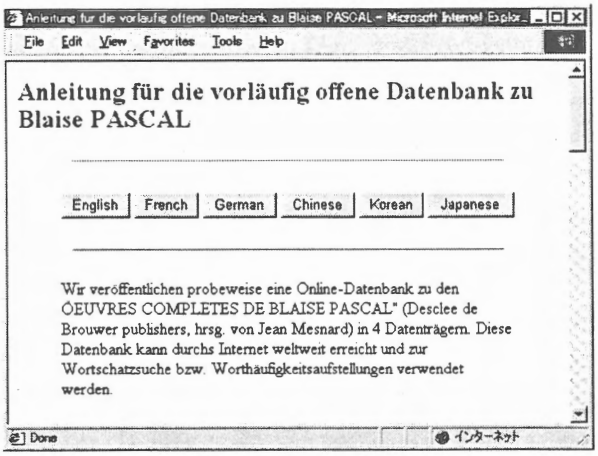

図 3: 試験公開要項 (独語)

4.2　データベース接続

パスワードを取得したら、それを用いてデータベースに接続する。
ブラウザを起動して
http://pascal.sci.fukuoka-u.ac.jp:8080/
にアクセスする。そのページにある Go ボタンをクリックすると、パスワード入力画面 (図 4) になる。

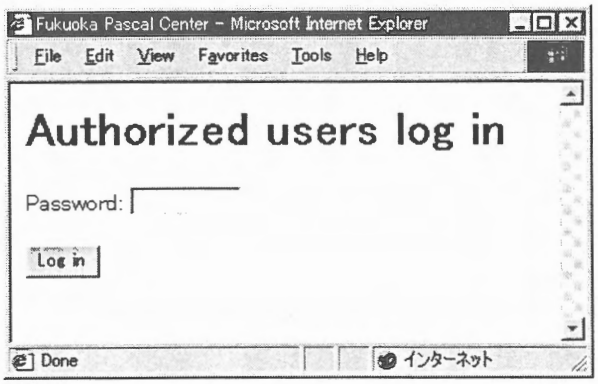

図 4: パスワード入力画面

"Password:" の欄に入手したパスワードを入力して、"Login" をクリックする。認証に成功すると、Search & Frequency 画面になる。(図 5) この画面

には 3 つのボタンがある。search と frequency, input_character_table である。search は語彙検索用ボタン、frequency は頻度作成のボタンである。input_character_table は変換対応表である。

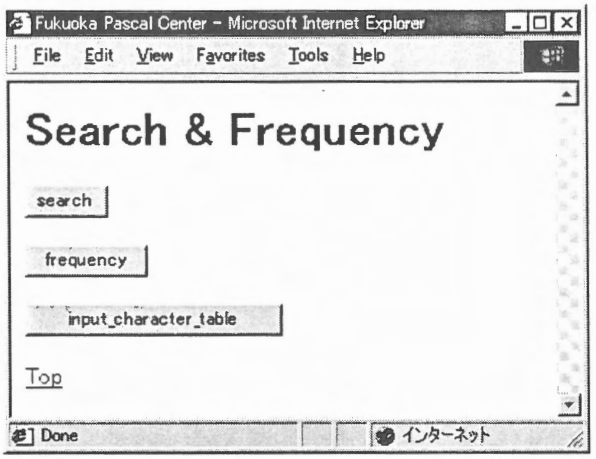

図 5: Search & Frequency 画面

4.3 語彙検索

検索には、語彙検索と目次検索の 2 種類を用意している。目次検索に関しては、作品ごとにアクセスすることができるようになっている。アクセスすると、作品の先頭の頁が表示され、頁の前後はボタンで捲れるようになっている。語彙検索を行うには、Search & Frequency 画面で “search” をクリックする。そうすると、Search 画面になる。

ここで、検索項目を指定する。識別 (P:パスカルが書いた部分等)、言語区分 (F:フランス語部分、L:ラテン語部分)、巻、頁、行、列検索文字列を入力して “Submit Query” をクリックする。

例えば図 6 では、パスカルがフランス語で書いた部分で、第 2 巻の最初から 1400 頁までの中で、4 つの単語または文字列 je、ne、 pas、et を含む文が何処にあるかを検索する (And 検索)。je の前後には空白を入れているので完全な単語として指定している。et の後には空白を入れているので、単語の接尾語になりうる。ne と pas に関しては、前後に空白を入れていないので単語の部分列の可能性が充分にある。

文字列を含む文にヒットしたら、その位置情報 (巻・頁・行など) がまず表示される。その位置データに関しても 10 個ずつ表示するようになっている。図 7 参照。

Next ボタンをいくつか押した後、位置情報 2-0523-16 を選択すると実際のフランス語の文が表示される。図 8 参照。

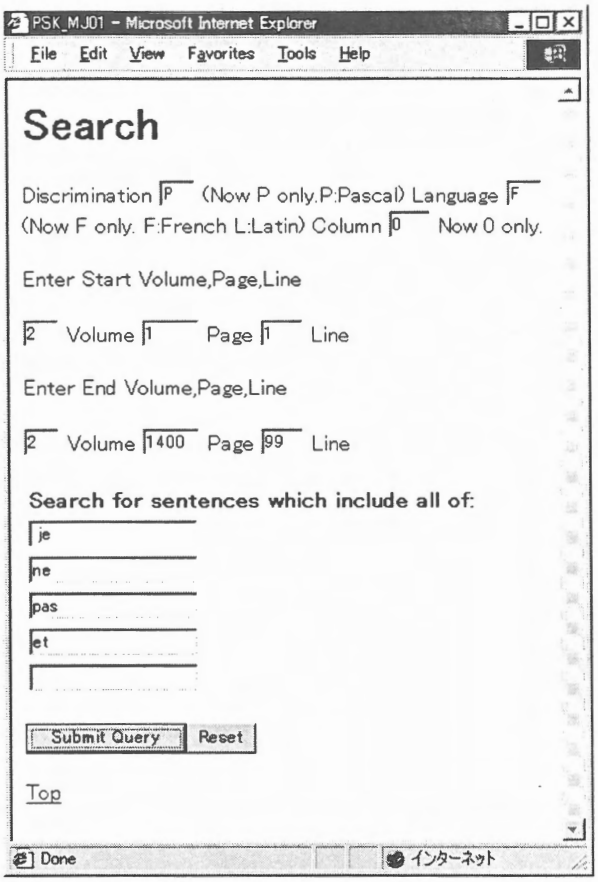

図 6: 語彙検索画面

ここで、さらに Show_Original_Text ボタンを押すと、その頁の画像データが表示される。3.2 画像データベース参照。

4.4 頻度表

データベース作成における頻度表作成は、誤入力の発見に貢献し、データベースをより完璧なものに近づけるものである。と同時に希少語の発見などにみられるように、1 次的な資料としての役割も担うので、大いに意義のあることとして、細心の注意をはらって取り組んできたものである。

その成果の一つであるパスカル頻度表第 2 巻の 1[5] は、パスカルが 1640 年初めから 1654 年の 15 年間に書いた論文、手紙など全ての作品を網羅したもので、もっとも華々しい活躍をした時期であると同時に宗教思想に目覚めた時期のものであり、内容は多岐にわたり、作品ごとの頻度には著しい変化がみら

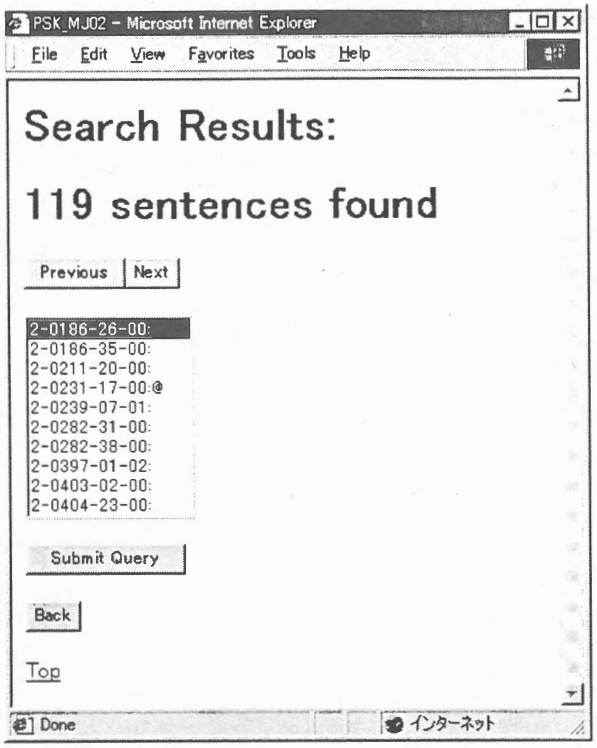

図 7: 検索結果一覧表示画面

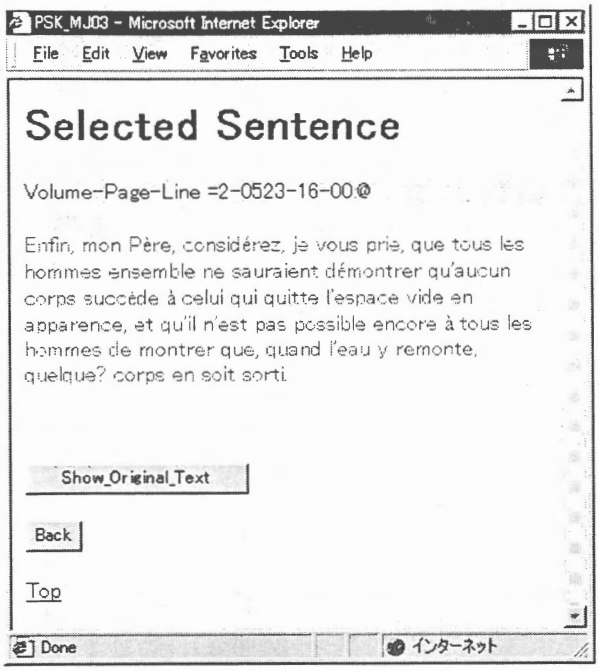

図 8: 検索文表示画面

れる。頻度表は、まず 2 巻中の収録作ごとに分けてアルファベット順、頻度順、逆引き順と作成し、また同時に、メナール版の特徴として、時系列的に作品を並べていることから、全体を併せて一つの作品と考え、この頻度表も作成した。これについても同じようにアルファベット順、頻度順、逆引き順と作成した。

頻度表を作成するには、Search & Frequency 画面 (図 5) で "frequency" をクリックする。頻度表条件入力画面 (図 9) が現れる。

ここで、次の頻度作成項目を指定する。

識別 (P:パスカルが書いた部分等)、言語区分 (F:フランス語部分、L:ラテン語部分)、検索範囲 (巻、頁、行、列)、頻度表を作成したい単語の範囲、検索結果の表示方法 (ABC 順、多寡順、逆引き順)

例えば図 9 は、パスカルがフランス語で書いた部分、検索範囲は、第 2 巻の 518 頁の 28 行から 527 頁までとしている。また、単語の範囲は、"p" から "r" までの間にある単語を対象にしている。

ABC 順 をチェックするとアルファベット順に頻度表が作成される。図 10 を参照。

また、図 9 で "Freq.Order" をチェックすると、図 11 のように多寡順に頻度表が、"Reverse dic" をチェックすると図 12 のように逆引き辞書が作られる。

4.5 変換対応表

パスカルデータベースにはアルファベット以外の記号は、変換対応表に基づいてデータベース化している。その変換対応表を見るには、Search & Frequency 画面 (図 5) で "input_character_table" をクリックする。図 13 参照。

例

é → e'　à → a‘　è → e‘
ù → u‘　â → aˆ　ê → eˆ
î → iˆ　ô → oˆ　û → uˆ
œ→ oe　ç → c/

5 今後の課題 – パスカルデジタルアーカイブの構築に向けて

人類共通の財産として位置付けられた建造物・自然環境を保存するために、各国が協力し連携しようというのが＜世界遺産＞の主旨だが、人類共通の遺産は建造物や環境自然だけではない。同じく人類が誕生して以来、築き積み上げてきた＜人類の叡智＞も、われわれの大切な遺産である。具体的には各学問の知識、芸術、技術なども人類の歴史文化資産として残さねばならない＜世界遺産＞である。文字、記号、映像などによって表現されてきたこのような分野は、

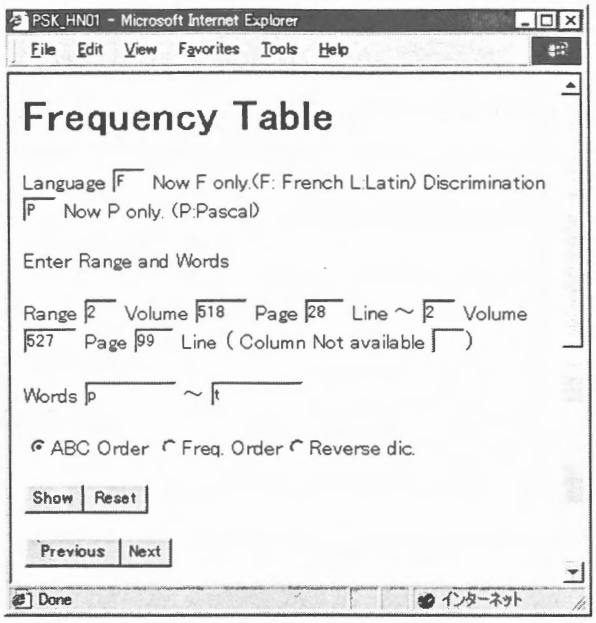

図 9: 頻度表条件入力画面

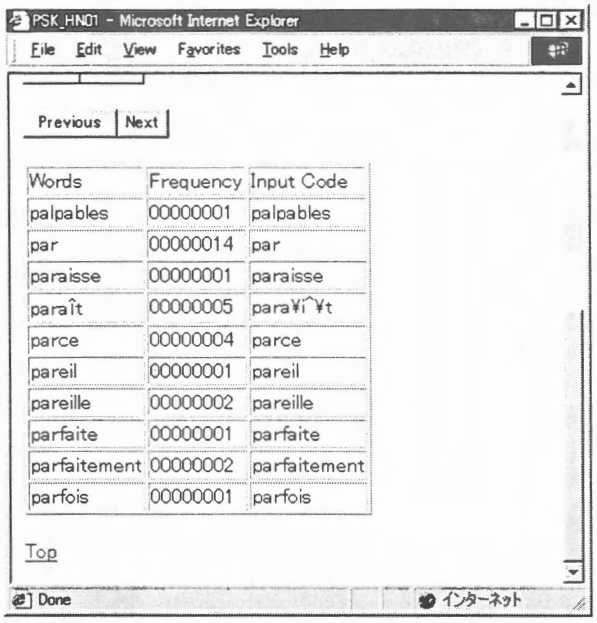

図 10: 頻度表表示画面 (ABC 順)

なまじそのような方法で保存されているために、保存方法の恒久性に付いての検討が後回しにされがちであったが、ここにきて従来の方法とは全く違う形でのディジタル技術を使用した保存方法によってこれら人類の叡智を後世へ継承しようという試みが生まれている。これが<デジタルアーカイブ>である。

デジタルアーカイブの必要性についての認識は急速に高まってはいるが、例えばインターネット公開に向けることを前提にしても、一般利用者、資料利用関係者などからの利便性という視点が欠除しているのが現状である。一般的には過去のマイクロフィルムによる保存のようにただ単に写し取るだけの作業に多かれ少なかれ終始していて、デジタルアーカイブの縦横無尽な可能性についての研究はまったくもって不十分な上、そのために解決しなければならない問題も山積みの状況である。しかし当研究グループは、数年前よりメナール版パスカル全集の編者本人であるメナール氏の協力を得、国内外で唯一というパスカルデータベースのデジタルアーカイブの研究を先行しており、資料を実際に引き出す側の使い勝手にも配慮し構築を試みる研究をし続けている。また、インターネット上での公開に向けて、諸問題も想定し、これに解決を与えてきている。それに加えての第２段階として、このデータベースの充実にヴィジュアルの側面も盛り込む方法を試みている。具体的にはパスカルの直筆の文章、彼に関するあらゆる物品及び映像 (多数のものが主にフランスに存在する) などの画像処理をしなければならないもので、これを収録して初めてこのデータベースが完成に近づいたといえるもので今後はこのデータの収集を目的とするものである。今後の研究では、パスカルに関する資料を収集し、デジタル化し、組織化することにより、多角的、微細に分析、整理、研究することができる。

哲学者、数学者、科学者、宗教家と多彩な顔を持ち、そのいずれの分野においても超一級の仕事を成し遂げた天才パスカルの全生涯、全思想を体系的に網羅したデータベースは世界にも類を見ない規模のもので、これを明快な形で等しく全世界共通の遺産として位置付け、インターネット上の公開を考えている。

6 底本

Œuvres complètes de Blaise Pascal, tomes.1–4, texte établi, présenté et annoté par Jean Mesnard, DDB.

7 謝辞

本研究は、1997 年度 重点領域研究 「人文科学とコンピュータ」研究課題名 パスカル全データベース作成と言語解析、1998 年度 特定領域研究 「人文科学とコンピュータ」研究課題名 パスカル全データベース作成と言語解析、1998 年度 研究公開促進費 学術

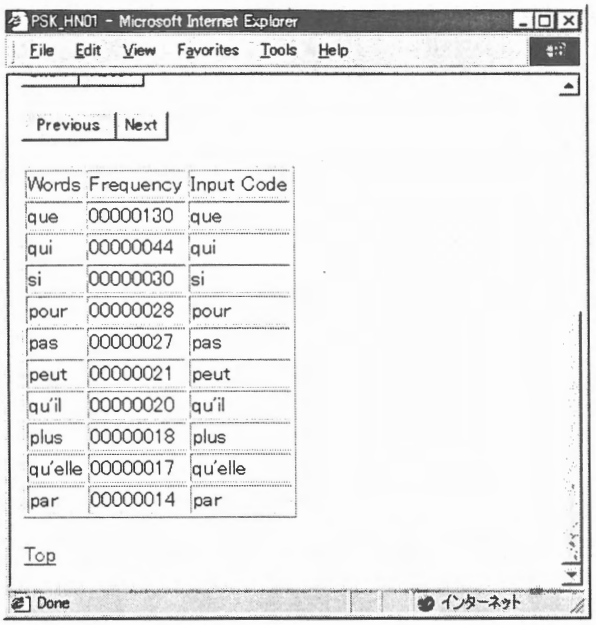

図 11: 頻度表表示画面 (多寡順)

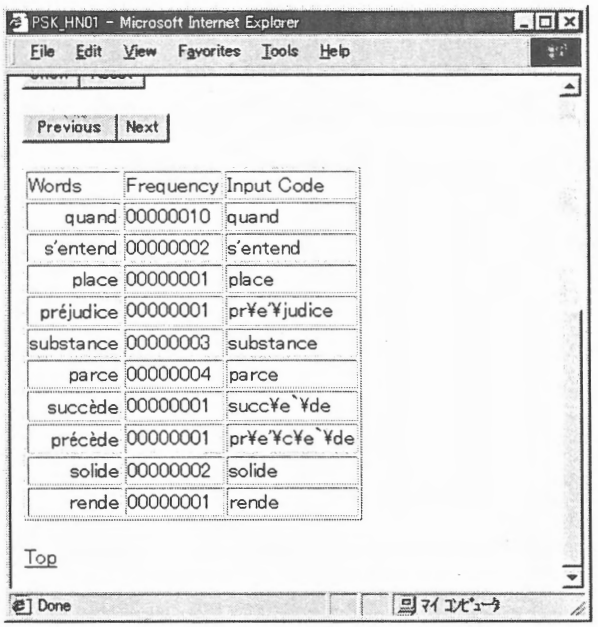

図 12: 頻度表表示画面 (逆引き)

図書図書名 パスカル頻度表第２巻の１、の成果に基づくものであることを付記する。

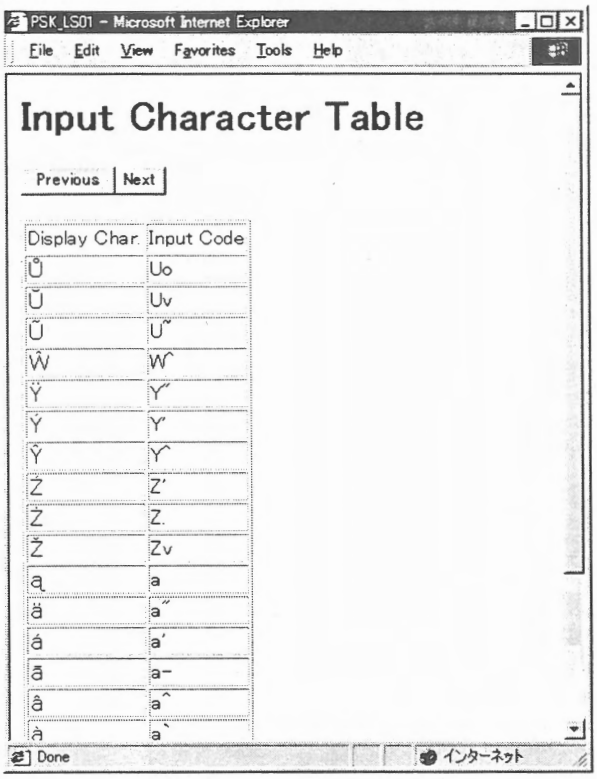

図 13: 変換対応表

参考文献

[1] パスカル全データベース作成と言語解析 (I), 情報処理学会, 情報研報, **36**, 6 (1997) 31–36. (共著: 輪田裕, 藤村丞)

[2] パスカルデータベース作成と言語解析 (II),1998 年 3 月. (情報処理学会「人文科学における数量的分析」にて共同発表)

[3] パスカル全データベース作成と言語解析 (III), 情報処理学会, 特定領域「人文科学とコンピュータ」研究情報誌, **6** (1998) 49–55. (共著:輪田裕, 藤村丞)

[4] パスカル全データベース作成と言語解析 (IV) 1999 年 8 月. (第 4 回「言語・認識・表現」研究会 印刷中)

[5] パスカル頻度表 第 2 巻の 1, 多賀出版, 1999 年 3 月. (共著:輪田裕, 柴田勝征) 533 頁

[6] 特定領域研究「人文科学とコンピュータ」1998 年度研究成果報告書 1999 年.