

パスカルデータベースシステム (II)

白石 修二

福岡大学 理学部応用数学科

藤村 丞

長崎大学 経済学部 総合経済学科

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 COMPLETES DE BLAISE PASCAL」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作品検索や語彙検索そして頻度作成ができるようになっている。語彙検索に関しては、And 検索、Or 検索、Not 検索に加え、ワイルドカードを用いた順序検索ができる。また、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。作品検索に関しては、各作品の写本データや、参考文献、メナールの注解が得られ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。

Pascal Database System (II)

SHIRAISHI Shuji

Department of Applied Mathematics, Faculty of Science, Fukuoka University

FUJIMURA Shou

Department of Economics, Faculty of Economics, Nagasaki University

This paper presents an overview of the revised Pascal Database System. The Pascal Database includes all the text from the 'Œuvres Complètes de Blaise Pascal' in 4 volumes. The online database was released experimentally in October 2000. It is possible to display material, perform a vocabulary search, and to make frequency lists of material in the database via the Internet. The content display can access each volume, plus manuscript data, edition, references, annotations of J. Mesnard, and other documents, which is a great advantage when studying the material. The vocabulary search can perform Boolean searches with 'And', 'Or', and 'Not' and can also use the wild card '*'. Frequency lists can be made using alphabetical or frequency order, and it is even possible to create a list based on the alphabetical order of the reversed words.

1 はじめに

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 COMPLETES DE BLAISE PASCAL」(Desclee 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。インターネット上で公開し、実際に多くの人にご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に機能するかどうかを確実に把握することができると思う。

試験公開から今のところ1年半経過した。パスワード認証を経ての利用を促しているの、利用を希望

する人はほとんど研究者にとどまっているが、国内外を問わず着実にユーザが増えている。これらパスカルの専門家に実際に運転してもらい、その結果をフィードバックしてもらっている。

本稿では、改訂版パスカルデータベースシステムの概要についてそのユーザインターフェース部分と検索エンジンについて述べる。

語彙検索に関しては、ブール検索である And 検索、Or 検索、Not 検索に加え、ワイルドカードを用いた順序検索(正規表現検索)ができる。また、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。作品検索に関しては、各作品の写本デー

タや、参考文献、メナールの注解が得られ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。

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.rsp.fukuoka-u.ac.jp:8080/> である。使用にあたっては、ブラウザに Java 1.3 Plug-in が必要である。

システムは、ワークステーション上に 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、サプレットを用いてのリモート処理が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:

Sun Microsystems Ultra 60

Ultra SPARC ii 450MHz

メインメモリ 512MB

サーバ OS:

Solaris 8

Web サーバ:

Tomcat 4.0.1

データベース:

Oracle8i Workgroup Server for SPARC

Java に関しては、次のパッケージを使用した。

Java 開発キット:

j2sdk version 1.3.1.02

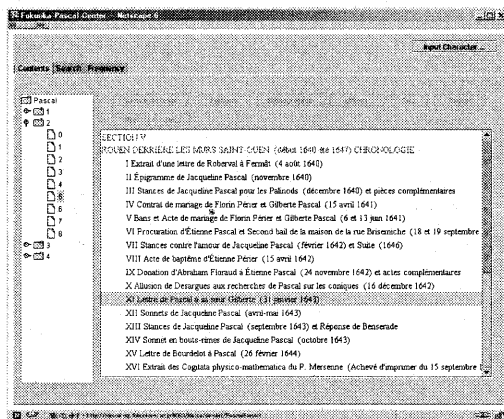
(Java Development Kit)

3 作品検索

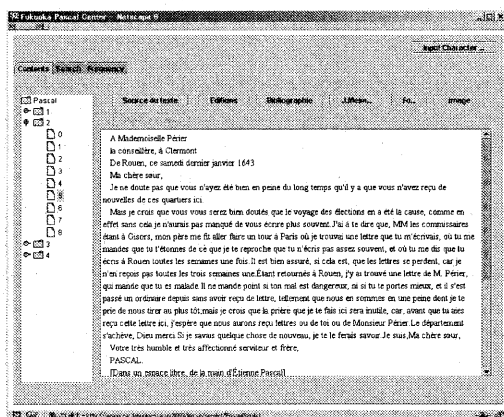
目次検索を行うには、Main 画面で、Contents タブをクリックする。そうすると、Contents 画面になり、目次を通して、各作品にアクセスできるようになっている。また、各作品に対して、写本データボタン、参考文献ボタン、メナールの注解ボタン等がついているので、それぞれのボタンを押すことにより、内

容が別窓にそれぞれ出るようになっている。

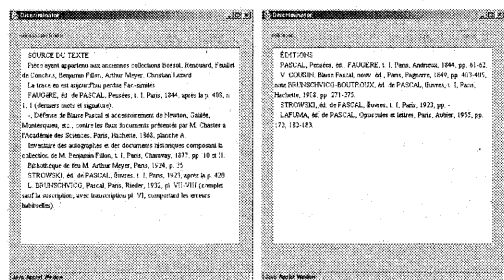
例えば、2 巻 5 章 11 節の作品を見るには、画面左にある tree で管理されている目次を順次図のように選択していった、11 節の見出しをクリックすると作品内容が現れる。ここで、写本データや Edition データが欲しい場合は、ボタンをクリックすると別窓に現れる。各作品に対して、写本データ、Editions、参考文献、編者メナールの解説、フットノート、原著の画像データが得られるようになっている。



2 巻 5 章 11 節



2 巻 5 章 11 節作品内容



写本データ

エディション

4 語彙檢索

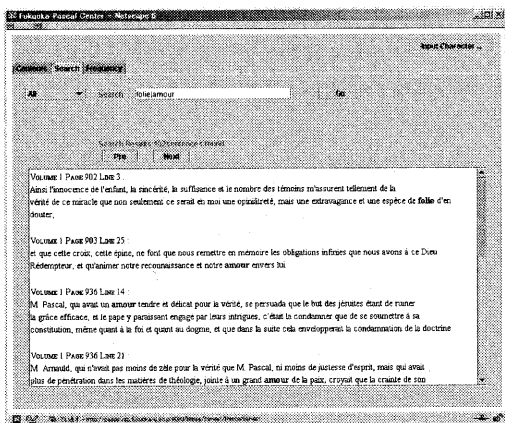
語彙検索を行うには、Main 画面で“search”タブをクリックする。そうすると、Search 画面になる。

デフォルトで、すべての巻がサーチの対象になっている。コンボボックスを選ぶことにより、絞り込んで検索できるようになっている。例えば、3を選ぶと、3巻だけに絞ってサーチできる。現在、各巻ごと、頁範囲指定、ラテン語の6種類を用意しているが、識別子を増やすだけで、いくらでも拡張できるようになっている。

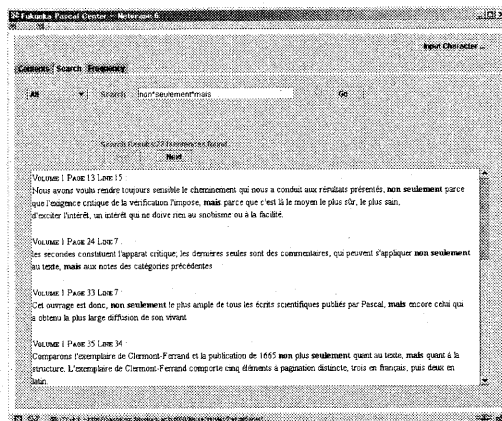
検索に関しては、ブール検索すなわち、And 検索、Or 検索、Not 検索に加え、ワイルドカード * を用いた一部正規表現の検索もできる。

And 検索 : a b c のように単語の間に空白を入れる。
 Or 検索 : a|b|c のように単語の間に | を入れる。Not
 検索 : a~b のようにチルド記号を用いる。ワイルド
 カードを用いた検索。次の形の検索ができる。a* 接
 頭辞が、a の単語を表す。*b 接尾辞が b で終わる
 単語を表す。a*b 部分列 a の後に 部分列 b が出現。
 a*b*c

次は、単語 *amour* と *folie* の Or 検索例と熟語 *non ~ seulement ~ mais* の検索例である。*non ~ seulement ~ mais* は *non*seulement*mais* と指定する。



Or 検索

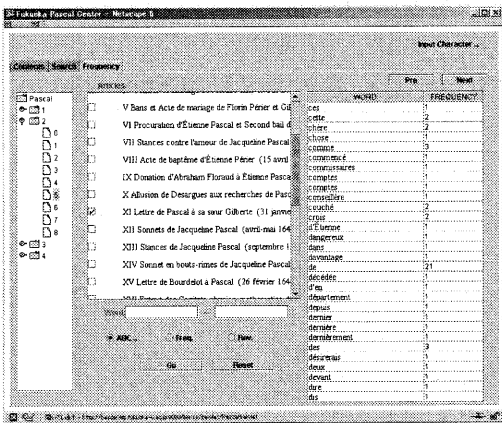


正規表現検索

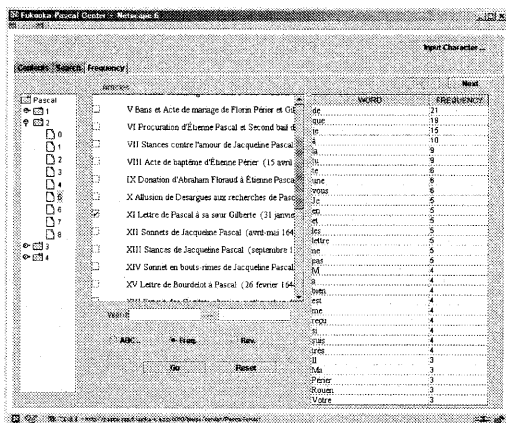
5 頻度表

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。

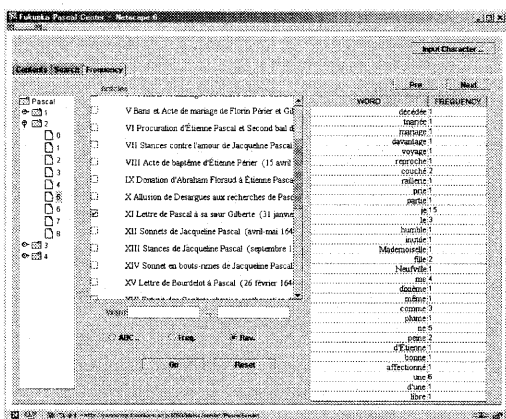
頻度表を作成するには、Main 画面で“frequency”タブをクリックする。頻度表条件入力画面が現れる。各作品ごとの頻度、またはいくつかの作品をまとめて頻度計算できるようになっている。abc 順、多寡順、逆引き辞書の3種類が計算できる。さらに、単語の範囲をたとえば、アルファベットが Q から R までのように絞り込むことができる。次は、2 巻 5 章 11 節の作品の頻度例である。



ABC 順



多寡順



逆引き辞書

6 おわりに

現在においては、当データベースの精度を高めるために、原著とテキストデータをつき合わせての細部のチェックを行っている。

またそれに加え、大小さまざま、多種多様な個々の修

正をデータベースに反映させるためには、迅速に処理できるようなデータベース保守管理用のインターフェースが必要との観点より、それに対応するインターフェースの作成を行っている。

パスカルの中で欠かせない数学や物理学の論文で使用されている数式や図表等は画像データベースの方で、すぐに確認できるというレベルではあるが、もっと、一步進めた画像処理の方向などを手探り中である。

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: 白石修二, 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: 白石修二, 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), LACE, 1999年8月 163-165. (第4回「言語・認識・表現」研究会) (共著: 白石修二, 輪田裕, 柴田勝征, 藤村丞)
- [5] パスカル頻度表 第2巻の1, 多賀出版, 1999年3月. (共著: 白石修二, 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.
- [7] パスカルデータベースシステム (I), 第6回公開シンポジウム人文科学とデータベース, (2000) 35-42. (共著: 白石修二, 藤村丞)