

生物種同定のためのデータベース検索システムの開発

— 汎用同定支援システム GENESIS —

矢 田 豊

I はじめに

森林を構成する樹木やその他の生物の同定は、森林を取り扱う上でもっとも基礎的な事柄のひとつである。生物の同定作業の労力を軽減するためのコンピュータ・ソフトウェアの研究開発は、以前より行われていた(1, 2, 3)が、特に近年のパーソナルコンピュータの「マルチメディア化」に伴い、いくつかの「図鑑ソフトウェア」も市販されはじめている。著者は1991年に、冬芽による樹種名検索のためのハイパーテキストシステムを作成し、公開した(5)。これは従来研究されていたプロダクションルールによるエキスパートシステム(4)とほぼ同一の原理で、最小限の労力で実用レベルのシステムを構築したものであった。しかし、ハイパーテキストやその他のプロダクションルールを用いたシステムにおいては、途中一ヶ所でも判断できない「分岐点」があると、基本的には同定作業はそこで中断されてしまう。また、別の生物群の同定支援システムを作成しようとした場合には、当然「検索ツリー構造」を最初から作らなければならない、多大な労力を必要とする。また市販のソフトウェアでは一般に、樹木と野鳥など、全く異なる複数の対象を単一のソフトウェアで扱うことはできず、使用者が利便性を考えてデータを一部改変・追加することもできない場合が多い。

これらの問題点を持たないソフトウェアとして、従来のデータベース検索システムの概念を元にした汎用同定支援システム GENESIS を開発し、公開した(41回日本林学会中部支部大会(1992)口頭発表)。本報では、GENESIS Ver. 1.1 (以下 GENESIS) の仕様と操作法の詳細について紹介する。

II 基本特性

GENESIS は、NEC のパーソナルコンピュータ、PC-9800 シリーズの MS-DOS 上で動作する、データベース検索システムである。GENESIS 本体はマイクロソフト社のプログラミング言語 Quick Basic Ver. 4.5 の統合環境により作成し、スタンドアロンタイプの実行可能ファイルとしてコンパイルし、公開している。データベースファイルは MS-DOS のテキストファイル形式であり、検索作業時には各フィールドの情報をコード化して、データベース全体をメインメモリ上の配列変数として保持するため、検索速度は比較的高速である。

GENESIS の処理の流れを図-1 に示す。各処理は固定されたウィンドウで表示され、ユーザーはウィンドウ内のメニューから必要な選択肢を選んで処理を実行する。各処理の機能の詳細は VI で述べる。また、本報で用いるデータベース構造用語の定義を表-1 に示す。

表-1 本報で用いるデータベース構造用語

検索項目	データベースの検索処理に用いる個々のフィールド(名) (例: 葉の形、葉の長さ、分布等)
(項目) 属性	各々の検索項目内の、分類名称 (例: 卵形、円形、楕円形等)
備考	検索処理に用いないフィールド

GENESIS の特徴を以下に示す。

・同定対象に対する汎用性

データベースファイルを交換することによって、異なる視点からの、または異なる対象の同定作業に使用できる。現在までに、樹木の同定のためのデータベースとして、葉、冬芽、樹皮による検索データベースを作成している。また、精英樹のア

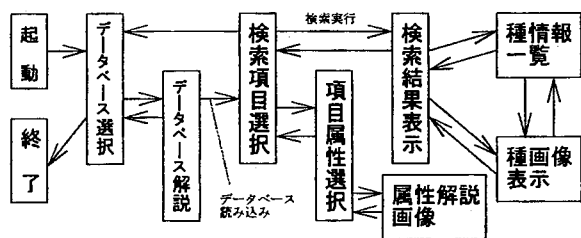


図-1 処理の流れ

イソザイムバンドパターンによる検索データベースや、野鳥の同定のためのデータベースも作成しており、その他の視点や対象への適用も可能である。

・検索作業の柔軟性

抽出条件は「備考」フィールド以外の任意のフィールドに対して設定できる。また、一つのフィールドに対する抽出条件として、複数の属性を「OR条件」で設定できる。データベースファイル上でのフィールドに対しても、最大4個までの属性を指定できる。これらの仕様により、不完全な情報や、種内変異の幅の広さに対しても柔軟に対処することができる。

・種または検索項目解説の画像表示

画像表示のための専用プログラムであるグラフィックローダを子プロセスとして実行することができ、それぞれの種の画像や、検索項目の補足説明のための画像表示が可能である。

・フリーソフトウェア

GENESISの配布は無料で行っており、所定の条件を守る限り、誰でも自由に利用できる。また、データベースファイルの仕様も公開しているので、新たなデータベースの作成等も、自由に行える。

III 使用環境

GENESISを使用するために、以下のハードウェアおよびソフトウェアが必要である。

・NECのパーソナルコンピュータ、PC-9800シリーズ（またはその互換機）

ただし、ハイレゾリューションモード、およびPC-98 LT/XAを除く。

・MS-DOS Ver. 3.3x以上

・ハードディスク

GENESISはフロッピーディスクベースで使用

可能であるが、多量の画像ファイルを扱う場合には、ハードディスクへインストールすることで効率的な運用が可能となる。

・グラフィックローダ

画像表示を行う場合に、使用する画像ファイルに対応するものが必要である。GENESISは、標準設定状態でメインメモリを248KB必要とするので、これに加え、その他のメモリ消費量とグラフィックローダのメモリ消費量の合計が640KB以内である必要がある。グラフィックローダによっては、EMSメモリを使用することによりメインメモリの消費を抑える仕様のものもあり、それらを使用する場合には、EMSメモリを用意する必要がある。

IV 環境設定

GENESIS起動時に環境設定ファイルGNSS 11.CNFが読み込まれ、運用環境が自動的に設定される。GNSS 11.CNFはテキストファイルであり、エディタ等での編集が可能である。以下に、Ver. 1.1 CにおけるGNSS 11.CNFの標準設定値と各パラメータの意味を示す。

- 500 [CR] 一つのデータベースファイルで扱える種数の上限値
- a : [CR] GENESISがインストールされているドライブ
- ¥gnss¥ [CR] GENESISがインストールされているサブディレクトリ
- a : [CR] データベース付属画像ファイル群がインストールされているドライブ
- b : [CR] 使用システムの最後のドライブ（[DB選択]の「D.ドライブ変更」で使用）
- hsx /w 1 [CR] 使用するグラフィックローダと起動時オプション
- 3 [CR] 項目/属性 『選択』時のカラーコード（1：青
2：緑 3：水色 4：赤
5：紫 6：黄）

[EOF]

V データベースの構造

種名とその特性を記述したデータベース本体(拡張子: TEX)と、そのデータベースファイルに関する補足情報を記述した環境設定ファイル(拡張子: CNF)の2ファイルで、ひとつのデータベースを構成する。いずれのファイルとも、MS-DOSテキストファイル形式であり、エディタ、ワープロ、表計算ソフト、データベースソフト等で容易に加工、作成できる。

データベース本体

データベースファイル本体は、CSV形式と呼ばれる形式のMS-DOSテキストファイルである。以下にデータベースファイルの例を示す。

種名, 裂目, 模様, シワ, 皮目, 粗滑性, 色, 裂目の深さ, 皮目の色形 [CR]

アオハダ, なし, なし, 横シワ, 散在, 滑らか, 灰色系, なし, 灰色系・黒色系・多形 [CR]

アカシデ, なし, 縦縞, 横シワ, 不明, 滑らか, 黒色系・灰色系, なし, なし [CR]

アカメガシワ, 網目状, なし, 横シワ, 不明, やや粗, 黒色系・灰色系, 浅い, なし [CR]

.
. .
.

[EOF]

このように、1行目に項目名を指定し、2行目以降から同じ順番でデータを記述する。検索項目の数は10以内とする。データの区切りは1バイトのカンマ(,)を使用する。また、1つの種の、1つの項目には4つまでの属性を指定でき、複数の属性の区切りには2バイトの midpoint (・)を使用する。この、属性の複数指定の方法は、GENESIS固有のものである。

GENESIS内での表示上の都合から、項目名は14バイト以内(または14バイトで区切りがよいような文字列)、属性の名称は12バイト以内、備考は66バイト以内にする。また、データベース全体における各項目内の属性の合計は、20以内に設定する。

データベース制作者の情報不足で、ある種のある項目の属性を特定できない場合、その位置には何も記述しない。GENESISはその場合、「情報なし」と判断し、検索時にはその項目に何が指定されても、その種は、その条件には該当するものとして処理する。

環境設定ファイル

各データベースファイルの環境設定ファイルもMS-DOSのテキストファイルで、以下のフォーマットになっている。

データベースファイルの解説文 [CR]

フィールド数, データ数, [画像ファイルのサブディレクトリ名], [属性解説画像ファイルの有無], [種画像ファイルの有無], [備考の有無], [CR] [EOF]

データベースファイルの解説文は、[データベース解説]で表示されるものであり、文の長さは216バイト以内にする。ここに、データベースの名称、解説、著作者、参考文献等を記載する。なお、72バイト目または144バイト目から2バイト文字が始まっている場合には、GENESISの表示時の改行のためにその文字が切れてしまい、「文字化け」を起こすので、注意が必要である。

フィールド数には、検索項目数に種名の分を加えた数を指定する。「備考」フィールドの有無は、ここでは考慮しない。データ数は入力した種数である。画像ファイルのサブディレクトリ名は、画像ファイルがない場合には指定する必要はない。

有無を指定する項目では、「有」の場合には1バイトの1を入力する。何も入力しない場合、または1バイトの1以外が入力されている場合、その項目は「無」と判断される。

備考が「有」の場合、データベースファイル本体の、最後のフィールドが備考とみなされる。

画像ファイル

GENESISで使用できる画像ファイルには、種画像ファイルと属性解説画像ファイルがある。GNSS 11. CNFで指定したグラフィックローダで使用できる画像フォーマットのファイルを用意す

る。現在までに公開しているデータベースでは、Q4フォーマットの画像を使用している。種画像ファイルは種名、属性解説画像は検索項目名をファイル名とし、GENESIS 11. CNF およびデータベース環境設定ファイルで指定したドライブ、サブディレクトリに置く。種名または検索項目名が8バイト以上ある時は最初の8バイトをファイル名とする。

VI インストール

プログラムファイルおよびデータベースファイルすべて（別表参照）を同一サブディレクトリ（標準設定時 A: ¥GNSS）に格納する。画像データを扱う場合には、同じディレクトリにグラフィックローダをコピーする。ドライブおよびサブディレクトリ名は、GNSS 11. CNF の内容を書き換えることにより、変更できる（IV参照）。また、各データベースに対応する画像ファイルは、別個のサブディレクトリに格納できる。これも、GNSS 11. CNF および各データベースファイルの環境設定ファイル（*.CNF）の内容を書き換えることにより、変更できる（IVおよびV参照）。

MS-Windows Ver. 3.1 に対応した PIF ファイル（GNSS 11. PIF）を用意しているので、ファイルマネージャからのドラッグ&ドロップ等により任意のグループへのインストールができ、Windows 上でも使用できる。アイコンファイル（GENESIS 1. ICO, GENESIS 2. ICO, GENESIS 3. ICO）も用意しているので、必要に応じてプログラムマネージャのアイコンメニューの「登録情報」、「アイコン変更」で設定する。

VII 操作法

起動

カレントディレクトリを、GENESIS をインストールしたサブディレクトリ（標準設定時 A: ¥GNSS）に変更してから、GENESIS. BAT を実行する。このバッチファイル内で、環境変数 qbgrpnc をセットしている。これは Quick Basic 上での、グラフィック画面の処理に対する指定を行うためのものであり、これが実行されないと GENESIS のグラフィック処理が正常に機能しない。GENESIS. BAT 実行後に「環境のためのメモリが不足している」旨のエラーが出る場合には、

上記環境変数行を AUTOEXEC. BAT に記述するか、CONFIG. SYS の SHELL 設定行/E: オプションの数値を大きくする等の対処が必要である。

基本的な操作法

GENESIS の操作はすべてカーソルキーおよびリターンキーのみで行える。反転表示の部分をカーソルキーで移動させ、リターンキー（またはスペースキー）で『選択』状態が設定される。

『選択』されると、文字表示が水色になる（色は変更可能。IV参照）。ここで再度リターンキーまたはスペースキーを押すと『選択』が解除される。また、コマンドの先頭にアルファベットの表示がある場合には、対応するキーを押すことによってもそのコマンドを『選択』できる。

【データベース選択】（図-2参照）

GENESIS を起動すると表示されるこの画面で、まず使用するデータベースを指定する。ここには、カレントディレクトリにある拡張子が TEX であるファイルの一覧が表示されている。どれかを『選択』すると、下に【データベース解説】が表示され、読み込みの確認を求めてくる。また、ここで必要に応じて「D. ドライブの変更」、「E. 終了」を選択できる。通常は「ドライブの変更」は必要ないが、環境設定が正しく行われていない場合、このコマンドでドライブを変更する。なお、サブディレクトリ名は環境設定ファイルで指定したものにのみ対応している（標準設定状態で ¥GNSS）。データベースファイルの拡張子は TEX に固定されている。

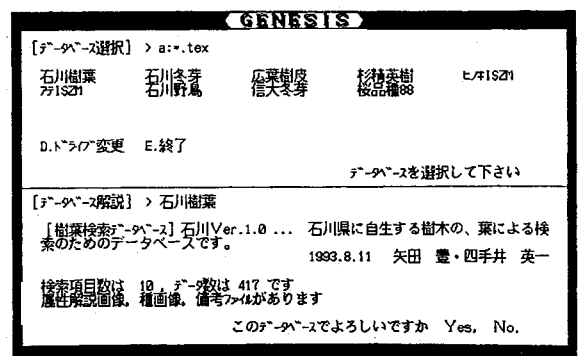


図-2 【データベース選択】および【データベース解説】画面

[検索項目選択] (図-3 参照)

データベースファイルを読み込むと次に検索項目の『選択』を行う。ここで、抽出条件を設定するデータベースのフィールドを指定し、次項の[項目属性選択]へ移動する。必要な設定が終わったら「S. 検索開始」で検索を開始する。「D. [DB 選択] へ」を『選択』すると[データベース選択]画面に戻る。

[項目属性選択] (図-3 参照)

ここで属性を『選択』し、実際の抽出条件の設定を行う。複数の指定が可能である。GENESISでは、一つの検索項目内の複数の属性を指定した場合には「OR 条件」で抽出が行われ、検索項目間は『選択』された(抽出条件が設定された)項目のみについて、「AND 条件」で抽出が行われる。すべての『選択』を取り消した状態で「E. 選択終了」を『選択』すると、その項目は『選択』を解除される。

また、属性解説画像のあるデータベースの場合には、ここで「G. 属性解説画像」を『選択』すると、指定画像ファイルが表示される(図-4)。

[検索結果表示] (図-5 参照)

[検索項目選択]画面で「S. 検索開始」を『選択』すると抽出処理が実行され、この画面になる。検索結果は一度に20種まで表示され、[ROLLUP]、[ROOLDOWN] キーで残りの種を表示させることができる。各種のデータベースの一覧を表示する「L. 種情報一覧」、および(画像ファイルのあるデータベースでは)種毎の画像を表示する「G. 種画像表示」があるので、必要に応じて『選択』する。種を先に『選択』しても処理を先に『選択』しても構わない。

「R. 再検索」は今の検索条件を残したまま[検索項目選択]へ戻る。「C. 検索条件消去」は今の検索条件をすべて消去して[検索項目選択]へ戻る。種数が多く、さらに条件を絞り込んで検索を行いたい場合は前者を、新たにまったく別の種を同定したい場合には後者を『選択』する。

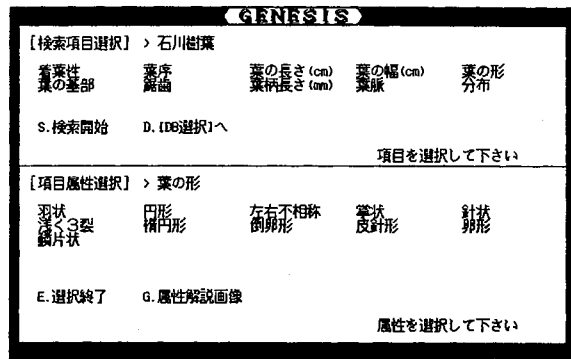


図-3 [検索項目選択] および [項目属性選択] 画面

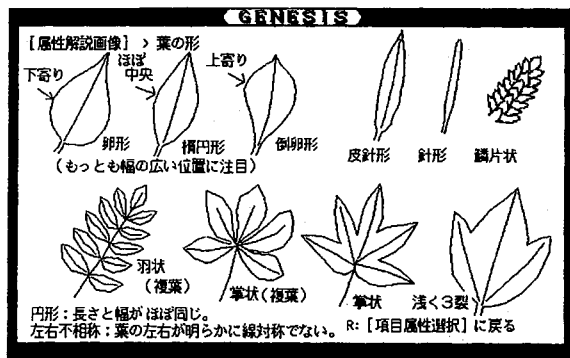


図-4 [属性解説画像] 画面

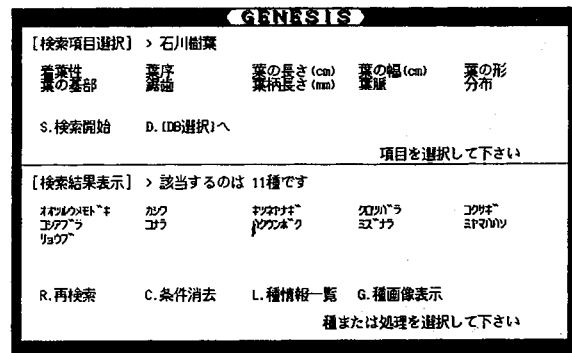


図-5 [検索項目選択] および [検索結果表示] 画面

[種情報一覧表示] (図-6 参照)

[検索結果表示] 画面で『選択』した種の、すべての検索項目の属性と、「備考」フィールドの設定のあるデータベースでは「備考」を表示する。「N. 次の種を表示」を選択すると、[検索結果表示] 画面での順番で、該当する種を表示する。「R. 検索結果表示」、「G. 種画像表示」を選択する事により、それぞれの画面へ移動する。

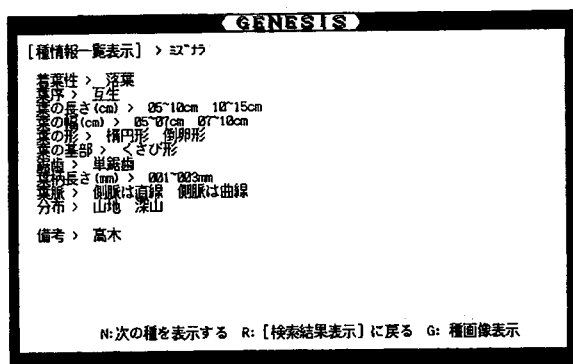


図-6 [種情報一覧表示] 画面

[種画像表示] (図-7 参照)

種画像の用意されているデータベースで、該当する画像ファイルを表示する。

「N. 次の種を表示」を選択すると、[検索結果表示] 画面での順番で、該当する種を表示する。「R. 検索結果表示」、「L. 種情報一覧」を選択する事により、それぞれの画面へ移動する。

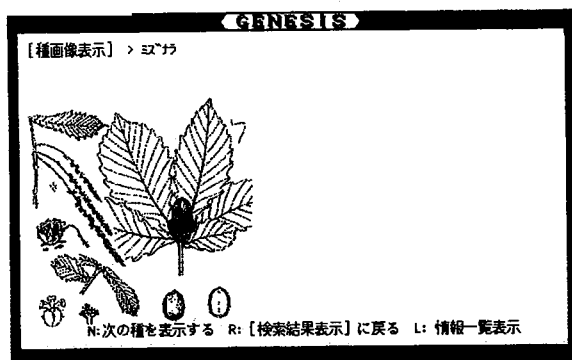


図-7 [種画像表示] 画面

VIII 著作権および配布

GENESIS の著作権は「石川県」にある。企画・制作は石川県林業試験場 矢田 豊が行った。

GENESIS はいわゆるフリーソフトウェアとし、以下の条件内での再配布、転載は自由としている。

- ・各ファイルの内容、名称を変更しない。
- ・再配布は無料で行う。ただしディスク代金、郵送料等の実費の徴収は構わない。

また、データベースファイルおよび種画像ファイル等は、それぞれ、GENESIS 本体とは独立した著作物と考える。参考までに、[樹葉検索データベース] 石川 Ver 1.0 の画像ファイルの取り扱いの条件について、次に示す。これは石川県林業試験場発行の「石川県樹木誌図譜」より画像データ化したものである。

- ・GENESIS 上での使用を前提とした場合の取扱いは、GENESIS 本体の取扱いに準じて行う。
- ・(GENESIS を介さない) 画像データ単独での使用については、以下の条件内で、許可する。
 - ・営利を目的としない。
 - ・1 著作物に複数種の画像を使用する場合は、その旨石川県林業試験場または矢田へ連絡をした時点で、使用を許可する。なお、その場合には石川県林業試験場発行の「石川県樹木誌図譜」より転載した画像である旨を明記する。
 - ・1 著作物に1種の画像を使用する場合は、連絡は不要とする。

希望者には郵送によりプログラムおよびデータベースファイル、画像ファイルを配布している。

IX バージョンアップ

Ver. 2 の仕様を現在検討中であり、ハイパーテキストタイプのヘルプシステムへの対応、サウンドプレーヤの子プロセス実行による音声情報への対応、多色画像表示への対応等を予定している。また、データベースに関しても既出のデータベースのバージョンアップ、画像ファイルの追加、現在開発中のデータベースの整備・公開等をいくつか検討しているが、ユーザーによる GENESIS 対応データベースの作成・公開も期待したい。

X 改変履歴

- '92.11.11 Ver.1.0 公開
- ・[落葉広葉樹冬芽検索データベース] 石川 Ver. 1.1、[樹皮による落葉広葉樹同定データベース] をバンドル
- '93.1.11 Ver.1.1 公開
- ・[石川の野鳥同定データベース] Ver.0.8 をバンドル
 - ・MS-DOS Ver.5.0 A への対応 (Ver.5.0 では動作未確認)
 - ・項目/属性 [選択] 時のカラーコードを任意設定可能に変更
 - ・上記事項にともなうセットアッププログラムおよび CNF ファイルの変更
 - ・セットアッププログラムを G11 SETUP.EXE (/BAT)、CNF ファイルを GNSS 11.CNF に変更。
 - ・[項目属性選択] から [検索項目選択] へ戻ったとき、カーソル位置を次項へ移動
 - ・[種情報一覧] と [種画像表示] に「N. 次の種を表示する」項を追加
 - ・データベースファイルの「情報なし」項目取りこぼしのバグフィックス
 - ・属性コードを格納したディメンジョン初期化のバグフィックス
 - ・その他微調整
- '93.8.11 Ver.1.1B 公開
- ・[樹葉検索データベース] 石川 Ver.1.0 をバンドル
 - ・MS-Windows Ver.3.1 への対応 (PIF ファイル、アイコンファイルをバンドル)
- '93.11.1 Ver.1.1C 公開
- ・ネットワークアップロード対応のための、ドキュメントファイル等の調整

XI 引用文献

- (1) 小林正吾・倉田和彦・竹内昌子：prolog による樹木名検索システム (I) - 広葉樹の場合 - . 99回日林論, 769~770, 1988
- (2) 小林正吾・井南哲司：prolog による樹木名検索システム (II) - 連想記憶モデルの利用 - . 101回日林論, 175~176, 1990
- (3) 古在豊樹・星 岳彦：農業・農学における知識工学的手法の応用. 農業および園芸, 61 (1), 3~10, 1986
- (4) 西川匡英：森林・林業へのエキスパートシステム適用の現状と問題点. 森林計画誌, 18, 45~66, 1992
- (5) 矢田 豊・四手井 英一・馬場多久男：「冬芽による樹木名検索システム」の構築 - ハイパーテキストソフトウェアの応用 - . 40回日林中支講, 255~256, 1992

別表－1 GENESIS Ver. 1.1 C に含まれるファイル

GENESIS. BAT	起動用バッチファイル
GNSS 11. EXE	GENESIS Ver. 1.1 プログラム
GNSS 11. CNF	GENESIS 環境設定ファイル
GNSS 11 C. DOC	GENESIS Ver. 1.1 C マニュアル
GNSS 11 DB. DOC	GENESIS データベース仕様マニュアル
HISTORY. DOC	GENESIS 改変履歴ドキュメント
TITLE. GRH	オープニング画像データ
FLAME. GRH	エンブレム画像データ
GENESIS. PIF	GENESIS Ver. 1.1 用 PIF ファイル
GENESIS 1. ICO	GENESIS 用アイコンファイル (その1)
GENESIS 2. ICO	GENESIS 〃 (その2)
GENESIS 3. ICO	GENESIS 〃 (その3)
石川樹葉. TEX	[樹葉検索データベース] 本体
石川樹葉. CNF	〃 環境設定ファイル
石川樹葉. DOC	〃 ドキュメントファイル
石川冬芽. TEX	[落葉広葉樹冬芽検索データベース] 本体
石川冬芽. CNF	〃 環境設定ファイル
石川冬芽. DOC	〃 ドキュメントファイル
広葉樹皮. TEX	[樹皮による落葉広葉樹同定 DB] 本体
広葉樹皮. CNF	〃 環境設定ファイル
広葉樹皮. DOC	〃 ドキュメントファイル
石川野鳥. TEX	[石川の野鳥同定データベース] 本体
石川野鳥. CNF	〃 環境設定ファイル
石川野鳥. DOC	〃 ドキュメントファイル
README 11. DOC	ファイルリスト