

JBiCバイオDBシステム

JBiC Biotechnology Database System

あらまし

JBiCバイオDBシステムは、バイオ産業情報化コンソーシアム（JBiC）が運用するバイオ研究者のための統合データベースシステムである。本システムはコンソーシアム参加バイオ関連企業の研究者を主たる利用者として、多様なバイオ関連データベースの共同利用を目指している。

本稿では、まず本システムの概要とその特長を述べる。つぎにそのシステム要件の一つである、複数の多様なバイオデータベースを横断的に検索するためのエージェント技術の適用とその効果について紹介する。また、もう一つの重要な要件である、共同利用システムにおける、秘匿情報のセキュリティの確保のための、内部データ暗号化とワンタイムパスワードによる検索・解析結果の取得の仕組みとその背景について紹介する。

Abstract

The JBiC biotechnology database system is an integrated database system operated by the Japan Biological Informatics Consortium (JBiC) for biotechnology researchers. This system enables joint use of various kinds of biotechnology-related databases and is intended mainly for researchers at the biotechnology-related companies participating in the JBiC. This paper first describes the outline and features of the JBiC biotechnology database system. Next, it describes the application and effects of the system's agent technology for cross-sectional searching of multiple biotechnology databases of various kinds, which is an essential part of the system. Then, this paper describes the mechanism and background of data encryption and search-analysis results acquisition within the system, both of which are done using one-time passwords. Lastly, this paper describes how this system is made available to multiple users while maintaining the security of the information it contains.



塩原立也（しおばら たつや）
ライフサイエンス推進室 所属
現在、バイオ研究関連システム構築に従事。

ま え が き

バイオ産業情報化コンソーシアム（JBIC）は、バイオ産業の情報化による研究開発のスピードアップとバイオ分野における新産業の創出を目的に、1998年11月に日本の代表的なバイオ企業とIT企業の参加のもとに設立され、2000年7月にバイオ関連5省庁（現4省）共管の社団法人に改組されている。

1999年12月に総理大臣が決定したミレニアム・プロジェクトでは、JBICは、バイオ産業の情報化を確実にかつ効率的に行うための産業界の窓口として中心的役割を担う。⁽¹⁾

JBICはその活動の一環として、研究開発事業で開発されたデータベースの高付加価値化を推進するとともに、協力機関と連携してオリジナルなデータベースを整備し、多様な分野で活用できる統合的なデータベース事業を推進している。

JBICバイオDBシステムは、この事業の中核を占めるバイオインフォマティクス関連データベースシステムである。1998年に実施された「バイオ研究情報高度利用システム開発」を基盤に、今後の技術開発に必要な、国内外のバイオインフォマティクス関連データベースの統合化、データベースの検索・解析機能の高度化を行い、バイオ研究者およびバイオ産業界に広く利用可能な環境を整備・運用しているものである。2002年4月より、JBIC会員に向けて、システムの正式公開を始めている。⁽²⁾

本稿では、本システムの概要を述べ、システムの特長である横断検索機能とセキュリティ機能とその技術的背景について述べる。

システムの特長

本システムは以下のような特長を持つ。

(1) 高度なセキュリティ環境を提供

パブリックな公開サイトと異なり、高度なセキュリティによって、利用者の検索情報（query）などは他者からは隠蔽いんぺいされており、安心してアクセスできる。

(2) 最新の公開・商用DB群へのアクセスをサポート

JBICサイトをアクセスするだけで、PIR（Protein Identification Resource：蛋白質配列データベース）、PROSITE（蛋白質配列モチーフデータベース）、GenBank（DNA配列データベース）、OMIM（Online Mendelian Inheritance in Man：ヒト遺伝病データベース）、PubMed（医学文献データベース）など、最

新の公開・商用DB群の情報を一度に入手できる。

(3) 独自の多彩な検索・解析ツール群を用意

JBICが独自に開発した、特長ある種々の検索・解析ツール群が利用可能である。

(4) 継続的なDBサービスを提供

多くの研究機関と連携し、産業界が必要とする独自DBの開発や、継続的なDB登録を実施している。

JBICバイオDBシステムの目的は、近年のバイオ情報の爆発的な増大と蛋白質の立体構造から、文献情報までの多様性に対して、複数の利用者が運営コストを分担することで、個々の企業では抱えきれないデータベースを維持していくことである。⁽³⁾

この要求に応えるべく、JBICバイオDBシステムでは、図-1に示すように、多様なバイオ情報をインターネットなどを使って、一括収集・管理・提供を行っている。利用者は特別なソフトウェアのインストールを必要とせず、基本的にはインターネットに接続された端末から、Webブラウザベースで、本システムのすべての機能（図-2）が、24時間（システムメンテナンスによる休止時間を除く）利用できるようになっている。

以下、本システムの技術的な特長について述べる。

エージェント技術による横断検索

一般にデータベースの利用目的、対象の範囲、ジャンル作成目的などに違いがなく、単に分散したデータベースを連携しようとすることは比較的簡単である。しかし、バイオデータベースのように、別々の組織で構築されたり管理されたりするサービス同士では、違いがないことを仮定することは困難である。

エージェント技術は、こうした様々な「違い」を吸収し、複数の（必ずしもデータベースとは限らない）違ったものも含んだ様々なサービスを統合して一つに見せる機能を提供する。これにより、それぞれが単独で存在するときよりも便利なサービスを実現することが可能となる。

本システムの開発に当たっては、AgentPROのエージェント機能とInterstageのCORBA基盤を適用した。⁽⁴⁾ 情報を提供するアプリケーション（データベースなど）や、要求者（利用者、ほかのアプリケーションなど）と知識をやり取りするエージェントにするというアプローチをとっている。エージェント同士が知識レベルのやり取りを行い、ACL（Agent Communication Language）と呼ばれる強力な表現力を持つ言語を通信に採用し、知識のレベルのやり取りを可能にしている。これにより、利用

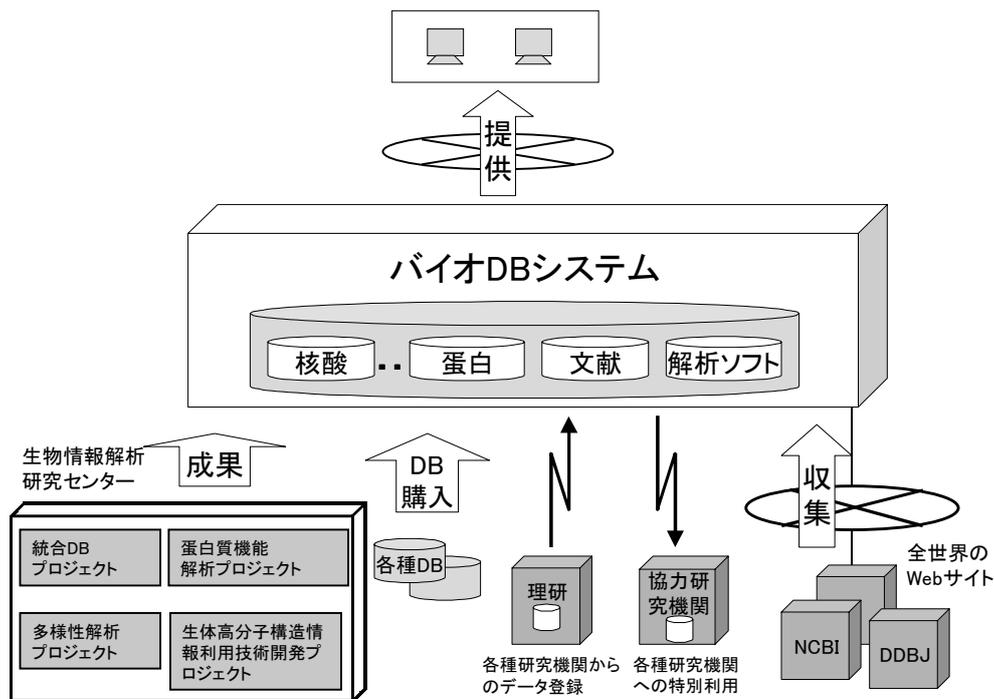


図-1 バイオDBシステム概要
Fig.1-Biotechnology DB system outline.



図-2 バイオDBサブシステム概要
Fig.2-Biotechnology DB sub-system outline.

者はデータベース種別を気にすることなく、データベースの横断的な検索が行えるようになった(図-3)。

さらに、AgentPROでは、三階層のエージェント機能を持ち、仲介機能を利用して、システム構成の変更にフレキシブルに対応することと、データの抽象化によって異なるシステムを容易に追加可能とする機能を持つ。AgentPROは、UA(ユーザエージェント)、FA(ファシリテータ)、AA(アプリケーションエージェント)から構成される。これにより将来的に新たなデータベース

の追加があった場合にも、ユーザインタフェース部とAAアプリケーション部を作成・追加することで、容易にシステムの構築と追加を可能としている。このことも先に述べたシステムの特長の中の(4)項の「継続的なDBサービスを提供」に対して大いに役立っている。

セキュリティ機能

先に述べたシステムの特長のうち、とくに利用者にとって重要なのは、(1)項の「高度なセキュリティ環境を提供」である。利用者である各バイオ企業はデータ

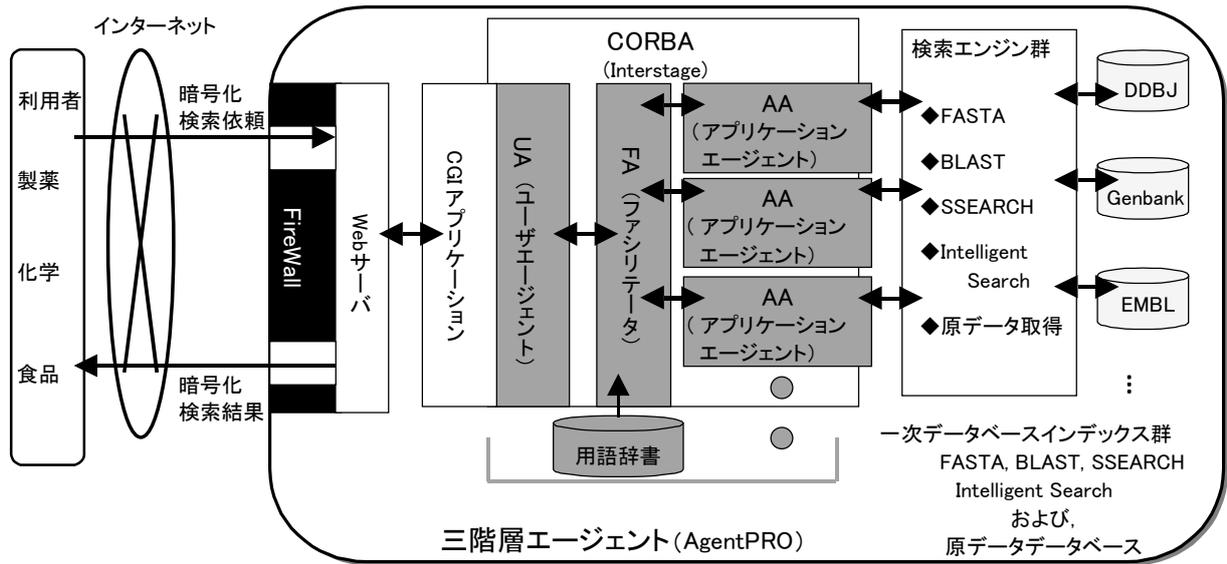


図-3 バイオDBシステムへのエージェント技術適用
Fig.3-Agent technology application to biotechnology DB system.

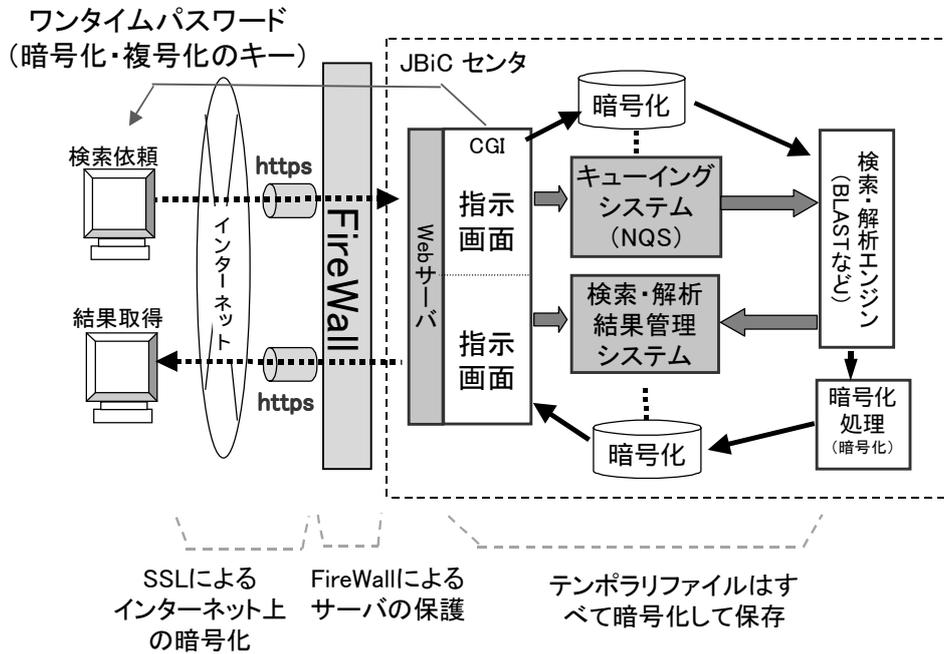


図-4 セキュリティ方式概要
Fig.4-Security method outline.

ベース運用コストを低減したいのはやまやまであるが、同時に、共同利用システムであることによる、秘匿情報の他社への漏洩については極めて敏感である。とくに、特許取得以前の遺伝子配列情報については、その配列情報が漏洩することは論外である。また、現在研究対象としているテーマそのものも重要な秘匿情報となる。したがって、どのようなキーワード（遺伝子の名前、生物の

名前など）を用いて、データベース検索を行っているか自体も、そこから研究目的を推測することが可能であるため、その情報のセキュリティについては十分配慮する必要がある。

本システムにおいては、セキュリティ要件として、保護すべき情報に対する不正なアクセスの規制、不正なアクセスの検出、情報の破壊防止、漏洩した情報の機密保

し指定データを暗号化し、ファイルリダイレクトにより暗号化ファイルを格納する。

NQSはスケジュールに順じUIアプリケーションを呼び出す。

UIアプリケーションは「復号化コマンド」を使用しユーザクエリ・クエリ管理情報を復号化し、作業用クエリを作成し、クエリ管理情報の状態を更新する。UIアプリケーションは作業用クエリをAgentPROに渡す。

作業用クエリは処理中のみ存在し、処理が完了直後に削除する。

AAアプリケーションは作業用クエリ内容を電文として受け取り、各種フリーソフトに渡す作業用クエリを作成する。このファイルを解析ソフトに渡す。処理結果はパイプインタフェースを通してAAアプリケーションが受け取る。

AAアプリケーションは、受け取った処理結果内容を「暗号化コマンド」を用いて暗号化する。

【ワンタイムパスワード】

本システムの検索・解析の結果取得には大きく二つの方法がある。その仕組みの概要を図-6(~)に示す。

(1) リアルタイム方式

検索や解析を依頼したら、即時に検索や解析を実行し答えを返す。利用者は、検索を依頼したらそのまま答え

を待つ。

(2) E-mail方式

いったん、検索・解析依頼をシステムが受け取り、検索・解析の完了後、E-mailで利用者に通知する。利用者は、解析が完了した後で結果を参照する。

リアルタイム方式については、通常のWebアプリケーションによる検索・解析システムと同じである。サーバとの通信は先に述べたようにSSLにより暗号化されており、セキュリティが保たれている。

しかし、タイムアウトが発生するような長い時間を要する検索や解析は、E-mail方式を利用者は選択する以外に方法はない。この場合に問題となるのは、検索・解析終了時に、検索・解析の依頼をした利用者以外にその結果を盗みとられる可能性があることである。ユーザ認証やあるいはCookieの利用、また固定IPアドレスなどにより、端末や利用者を特定することは技術的に可能である。しかし、利用者がブラウザを閉じず、席を離れた場合には、解析結果を盗み見られる可能性が高い。また、端末が固定されるのは使い勝手の面でも問題がある。

本システムでは、この問題をワンタイムパスワードにより解決している。ワンタイムパスワードとは配列情報などの利用者データをJBiCサーバ内部で暗号化するときの暗号キーである。実際には、依頼ごとにランダムに暗号キーを発生させ、結果取得時の利用者認証にも利用

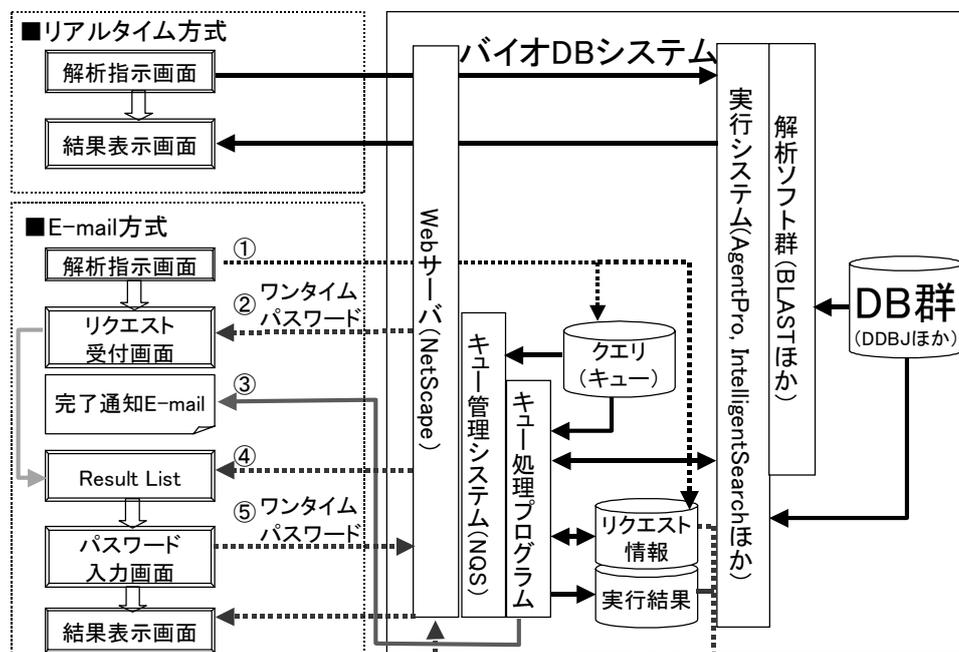


図-6 検索・解析の結果取得の仕組み

Fig.6-Mechanism of acquisition as a result of searching/analysis.

している。

以下にE-mail方式での流れを述べる。

E-mail検索での依頼時には、必ず受付画面が表示される。

受付画面には、依頼時の詳細とリクエストID・ワнтаイムパスワードが表示される。本情報はこの画面でしか表示されないため、利用者のノートなどに記録が必要である。

検索が完了した旨の通知がユーザ登録時に申請したアドレスにE-mailで送られる。

検索の完了はメールによって知ることができるが、ResultList画面の監視機能でも知ることができる。ResultList画面の中の検索・解析完了ジョブの中から一つを選択し、Submitボタンで表示指示（ログインしたユーザIDの依頼ジョブのみを表示する。ほかのユーザIDのジョブは表示されない）。

ワнтаイムパスワード入力画面でパスワードを入力し、検索結果画面へ移る。

検索・解析結果は一定期間保存された後、システムによって自動的に消去される。

む す び

バイオ情報データベースは個々のデータベース提供機関で作成されているために { 例えば, DNA配列データ

ベースの一つであるDDBJ (DNA Data Bank of Japan) は国立遺伝学研究所で作成されている } , 格納形式, 用語などが統一されておらず, 情報の相互活用が困難な状況にある。また, 解析ツールについても個々のシステムにおいての利用を前提に開発されており, 連携処理が困難である。

一方, バイオ情報は研究者にとって, 重大な機密情報を含むものであり, そのシステムにおいて, 機密保護の実現が重要な課題となっている。

本システムでは, これらの課題を横断検索機能とセキュリティ機能により解決している。本システムが日本のバイオ産業の振興に貢献できることを願っている。

参 考 文 献

- (1) バイオ産業情報化コンソーシアム (JBIC) ホームページ
<http://www.jbic.or.jp>
- (2) JBiCバイオDBシステムのホームページ
<http://jbic1.jbic.or.jp/ec3/index.html>
- (3) 佐藤千秋: 膨張するバイオデータベースを共有システムで乗り切れ. 日経バイオビジネス, 2002年6月号, p.93-96, 東京, 日経BP社.
- (4) 牧本治男: アプリケーションサーバ: INTERSTAGE . FUJITSU , Vol.50 , No.3 , p.116-120 (1999) .