

メタデータ検索システムについて

本資料では、実際にメタデータ検索システムを構築する際の留意点などのついて記述します。

1. メタデータ検索システムについて

(a)必要な要件

日本においてメタデータ検索システムを構築する場合、以下の要件を最低限満足する必要があると考えます。

- (1)XML 形式のデータを格納・検索できること。
- (2)日本語（2バイトコード）が扱えること。
- (3)Web サーバやクライアントから検索・結果表示が可能であること。
- (4)各サーバ（ノードサーバ）間で横断的にデータの検索が可能であること。
- (5)データ管理が容易であること。
- (6)特定の OS、ハードウェア、アプリケーションになるべく依存しないこと。

(b)Biosafety Clearing-House(BCH)の TOOLKIT について

BCH では、生物多様性関連のメタデータ検索システムを開発中です。その中に Toolkit と呼ばれる開発用ツールが含まれています。詳細については開発中のため基本概念（データの登録や検索手順等）しか公開されていませんが、将来的に日本の CHM と海外の CHM を連結する際に検討が必要になるかもしれません。

2. 検索システム構築における注意事項

(a)ネットワークセキュリティ

現在、一般的なネットワーク構成では、内部ネットワークとインターネットの間はファイアウォール等で通信を制御して、外部からの不正アクセス防止やウイルス侵入検知などのセキュリティ対策を講じています。WWW サーバなどの外部公開用サーバもファイアウォールの内側に設置するのが一般的です。

このため、CHM サーバ（ノードサーバ）についても、他のノードサーバと通信をするためには間にファイアウォールなどの通信制御装置が介在することを考慮しなくてはなりません。

インターネットで通信を行うためには通信用のポート番号を確保する必要がありますが、ユーザが利用可能なポート番号は限られております。これらのポート番号はファイアウォールなどで閉じられている場合があり、また、他のアプリケーションで使用されている可能性があります。

以上のことを考慮すると、ノードサーバ間の通信には HTTP 通信用のポート番号（80）を利用し、通信プロトコルについても HTTP を利用することが望ましいと考えます。

(b)プラットフォーム（ハードウェア、OS、データベースソフト等）

ノードサーバを構築するためのアプリケーションについては、Java 言語など複数の OS、ハードウェアで動作する言語で作成する必要があると考えます。また、メタデータを格納するデータベースソフトについても、フリーウェアや製品版など各種存在しますが、個々のデータベースソフトウェアの独自機能を利用するのではなく、汎用的な機能でメタデータ検索システムを構築し、ノードサーバを構築するユーザが各自の環境や予算にあったシステム構成を選択できるよ

うにする必要があると思われます。

3. 生物多様性センターの環境

環境省の生物多様性センターでは、基礎調査データや RDB 種データをデータベースに格納し、Web から検索できるシステムを一般公開しています。当該センターのサーバ構成は以下の通りです。

- ・ WWW サーバ (UNIX サーバ, OS : Solaris)
- ・ 地図用 WWW サーバ (WindowsNT サーバ, OS:WindowsNT4.0)
- ・ データベースサーバ (UNIX サーバ, OS : Solaris, DB : Oracle7)

これらのサーバを利用することにより、異機種間の接続テストや、データベースと WWW サーバ連携テストなどが実施可能です。

なお、センターでは内部ネットワークとインターネット間にはファイアウォールが存在し、HTTP 通信や電子メールなどで使用するポート以外は全て通信を遮断しています。