

試験問題データベース検索・解析のための統合環境の構築

—センター試験問題データベースとのリンケージ—

石岡 恒憲, 橋本 貴充, 大津起夫 (大学入試センター)

著者らはセンター試験問題, 全国国公立大学入試問題を所内ネットワークを介して包括的に検索表示するシステムを構築し, センター試験問題については以前より作成してきた統計情報データベースとのリンケージを図る仕組みを作成した。この実装は現在のネットワーク・コンピュータ基盤技術の一つである Windows Server2008 を用いれば容易に構築できる。本システムと同等のシステムを作題部会に提供することにより, その作題支援に供することが期待される。

1 はじめに

著者ら大学入試センター試験評価解析研究部門では, いままで過去のセンター試験問題についての統計情報を整備してきた (大津ほか, 2008)

また著者らは株式会社JC教育研究所が作成・提供するセンター試験問題データベース (CenterTen) を用いて, 検索した結果に対してボタン一つで該当する統計情報を表示・提供するシステムを開発・試作した (石岡ほか, 2010)。つまり使い勝手のよい市販の検索システムと我々の統計情報データベースとのシームレスな統合を達成した。

今回, 我々はセンター試験問題に加えて, 全国の国公立大学3年分の試験問題 (問題文そのもの) を新規に導入したファイルサーバに格納し, 以下の検索システムをそのファイルサーバに構築した。

- Exam : 全国国公立大学試験問題DB検索システム(JC教育研究所, 2009b)
- センターTen : センター試験問題DB検索システム(JC教育研究所, 2009a)
- TrueTeller : 野村総合研究所 (NRI) 製, テキストマイニングツール
- Google Desktop : Google社製 Google インターフェイスによる検索システム (Google, 2011)

また, このサーバマシンにターミナルサービスGW (ゲートウェイ) の機能を持たせることで, 研究開発部ネットワーク (RD-Net) に繋がれている端末からファイルサーバ上にある上記アプリケーションをTS Webアクセスにて起動できる環境を構築した。現在の拡張システム構成を図1に示す。

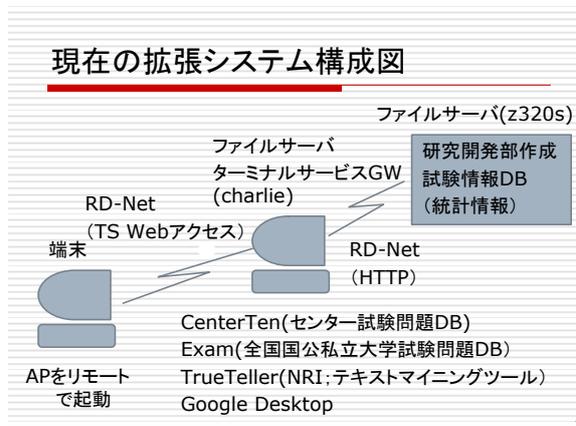


図1 拡張システム構成図

本稿で実装したシステムは, いわゆるネットワーク・コンピューティングといわれるものであるが, 巷で話題となっているクラウドコンピューティング, なかでも企業内のサービスであるプライベートクラウドの一形態と位置付けることもできる。

2節には, ネットワーク・コンピューティングの基盤技術であるターミナルサービスに

よるクライアントの仮想化について整理しておく。3節には、クライアントで実行できるアプリケーションの見え方や外部仕様についてまとめる。4節には本ファイルサーバを構成する上で注意すべき事項について記録する。5節はまとめと今後の拡張について述べる。

2 ネットワーク・コンピューティング

2.1 システム要件

実装したシステムはWindows Server 2008の有する機能に大きく依存している。その前バージョンであるWindows Server 2003では動作を保障しないことを予め断わっておく。

2.2 ターミナルサービス

本節ではターミナルサービスの基本的な知識について整理する。ターミナルサービスは一言でいえばデスクトップ環境を仮想化するための仕組みである。これによりアプリケーションをリモートに設置したターミナルサービス上で動作させ、アプリケーションの画面だけを手許のコンピュータで表示させることができる。ターミナルサービスを使用するとクライアントからはマウスとキーボードの操作情報のみが送信され、ターミナルサービスからは画面のみの転送されてくる。この際に使用されるプロトコルはRDP(Remote Desktop Protocol)と呼ばれる。

Windows Server 2008ターミナルサービスの主要な新機能および拡張機能には以下がある(安納, 2008)。

- Terminal Services RemoteApp (TS RemoteApp) によるシームレスなアプリケーション実行環境
- Terminal Services Web Access (TS Web Access) によるブラウザを使用したターミナルサービスへのアクセス
- Terminal Services Licensing (TS ライセンス) によるターミナルサービスへのク

ライアント・アクセス・ライセンス (TS CAL) の一元管理

- Terminal Services Gateway (TS ゲートウェイ) によるインターネットからのターミナルサービスへの接続
- Terminal Services Session Broker (TS セッション・ブローカ) によるターミナルサーバのロードバランス
- Terminal Services Easy Print (TS Easy Print) によるサーバ側のプリンタ・ドライバなしでの印刷の実現
- ターミナルサービスへのシングル・サインオン

2.3 RemoteAppプログラム

前節で述べたターミナルサービスのうちRemoteApp プログラムは、ターミナルサービス経由でリモートアクセスされるプログラムであり、エンドユーザーのローカルコンピュータ上で実行されているかのように表示される。ユーザは RemoteApp プログラムとローカルプログラムを同列に並べて実行できる。1人のユーザが1つのターミナルサーバで複数のRemoteAppプログラムを実行している場合、それらのRemoteAppプログラムは同じターミナルサービスセッションを共有する(安納, 2009)。

Windows Server 2008では、ユーザは、管理者が選択する展開方法に応じて以下に述べるいくつかの方法でRemoteAppプログラムにアクセスできる。

- TS Web アクセスを使用して、Web サイト上のプログラムへのリンクにアクセスする。
- 管理者が作成および配布したリモートデスクトッププロトコル(.rdp)ファイルをダブルクリックする。
- 管理者がWindowsインストーラ(.msi)パッケージを使用して作成および配布した、デスクトップまたは[スタート]メニュー上のプログラムアイコンをダブルクリッ

クする。

- ファイル名の拡張子がRemoteAppプログラムに関連付けられたファイルをダブルクリックする。これは、管理者がWindowsインストーラパッケージを使用して構成する。

著者らはこれらの選択肢の中から、最初のTS Webアクセスを介して、RemoteAppを配布することにより、これ（RemoteAppプログラム）をWebサイトで利用できるように決めた。これだと一般ユーザにはURLのみを通知するだけでよく、またユーザは自らがこの環境で実行できるアプリケーションが何であるかを意識することができると考えたためである。

3 クライアント側の操作

3.1 操作手順

本節ではクライアント側の操作、すなわちRD-Netに繋がれたコンピュータからのアプリケーションの実行方法についてマニュアルに準じて記録する。

- 1) TS WebアクセスをインストールしたWebサーバ(charlie:202.16.123.167)にアクセスする。

<http://charlie.rd.dnc.ac.jp/ts/>

- 2) TS Webアクセスに接続するためにはクライアントコンピュータでリモートデスクトップ接続(RDC) 6.1以上が実行されている必要がある。RDC 6.1は次のオペレーティングシステムに含まれている。

Windows Vista SP1

Windows XP SP3

RDC 7.0はWindows 7に含まれている。さらにターミナルサービスActive Xクライアントコントロールが有効になっている必要がある。Active XはこれらのRDCに含まれている。

- 3) Windows XP SP3を実行している場合、最初にTS Webアクセスサイトにアクセスすると、[ActiveX コントローラがイン

ストールされていないか、有効ではありません]というエラーメッセージが表示される。この場合は、たとえばInternet Explorer 7を使用しているならば[ツール]メニューの[アドオンの管理]をクリックし、[アドオンの有効化または無効化]をクリックしActive Xを有効にする。

- 4) ログイン資格情報を入力する。

ユーザ名 : dnc

パスワード : ***** (伏せ字にしてある)

認証がされると図2に示すウィンドウが表示され、実行可能な以下のアプリケーションがアイコンで表示される。



図2 TS Webアクセスの起動画面

- 5) リモート接続警告画面

アプリケーション・アイコンをクリックすると警告画面(図3)が表示されるので「接続」をクリックし、再度、資格情報を入力する(図4)。資格情報のユーザ名は4.の手順で入力したユーザ名がそのまま継承されるので、パスワードのみを入力すればよい。

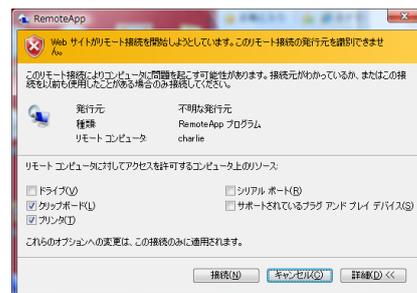


図3 リモート接続警告画面

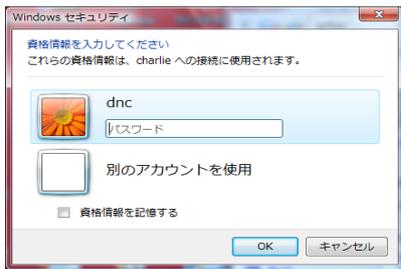


図4 資格情報の入力画面

3.2 実行可能なアプリケーション

TS Web アクセスで実行可能なアプリケーションには以下が含まれる。その動作仕様について詳細をここで述べることはしない。それぞれのマニュアルを参照されたい。

- Exam : 全国国公立大学試験問題DB検索システム (図5はトップメニュー, 図6は英語をクリック選択した画面)



図5 Exam トップメニュー

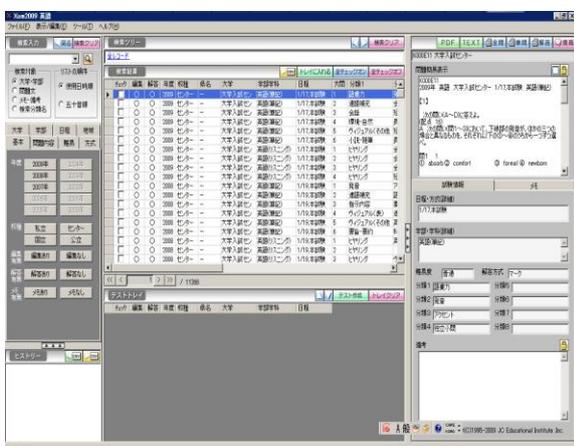


図6 Exam 「英語」 起動画面

- センターTen : センター試験問題DB検索システム (図7はトップメニュー, 図8は英語をクリック選択した画面)



図7 センターTen トップメニュー

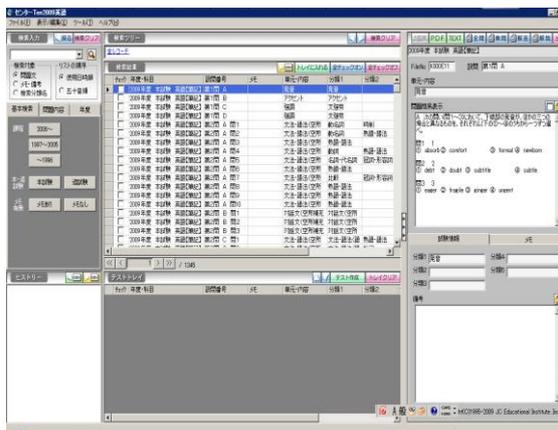


図8 センターTen 「英語」 起動画面

- TrueTeller : 野村総合研究所 (NRI) 製, テキストマイニングツール (図9は起動画面)

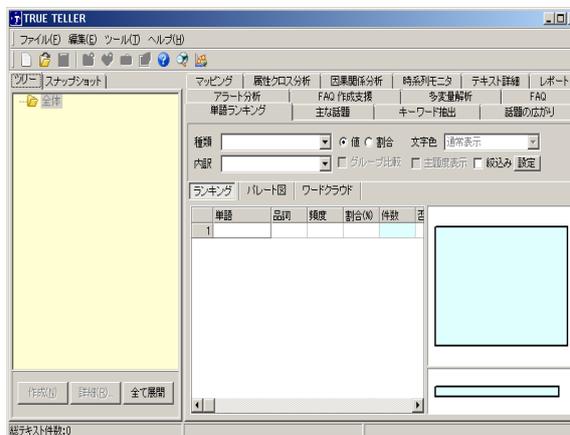


図9 TrueTeller 起動画面

- ・ Google Desktop : Google社製 Google インターフェイスによる検索システム。

Google Desktopのアイコンをクリックするとツールバーにそのアイコンが格納され、これを改めてクリックする。すると場面上に検索ウィンドウが表示され (図10), 検索文字列を入力するとその検索結果がブラウザにGoogleのフューチャで表示される。図11は検索文字列に「下関」を入力したときの検索結果画面である。



図10 Googleデスクトップ検索画面

Google デスクトップでの作業を終了させるときは、デスクトップのアイコンをクリックし「終了」させる。これにより利用者ユーザのロックを解除させることができる。

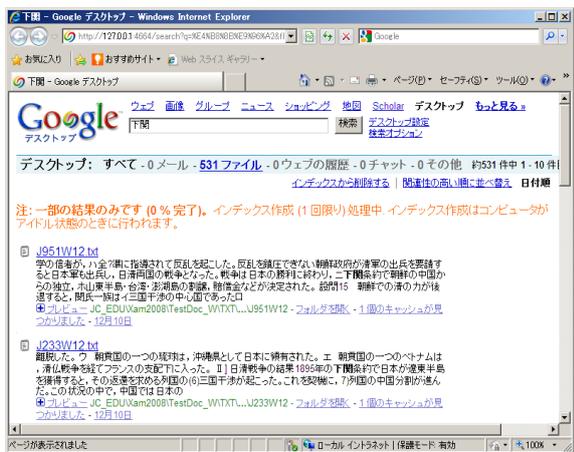


図 11 Google デスクトップ検索結果画面 (検索文字列として「下関」を入力した場合)

3.3 検索対象

センター試験問題データベース(Center Ten), 国公私立試験問題データベース (Xam) において検索対象となるファイルのサイズおよびファイル数 (問題数に相当する) は表1の通りである。Google デスクトップではこれらを瞬時で検索, 表示することができる。

表 1 格納されているデータ量

データベース名	容量	ファイル数
CenterTen	2.2 GB	23,941
Xam 2007	3.1 GB	47,738
Xam 2008	4.0 GB	51,008
Xam 2009	4.2 GB	53,271

4 ファイルサーバを構成させる上での注意事項

4.1 Google デスクトップのインデックス

(1) 格納場所 : インデックスファイルは, 標準では

C:\¥Documents and Settings¥Username
¥Local Settings¥Application Data¥Google

の中の「Google Desktop Search」というフォルダに保管され, このインデックスの容量はギガのオーダーで決して少なくない。検索の対象となるファイル, および検索インデックスをC以外のパーティションに置かないと, Cドライブの空き容量を圧迫し, サーバの動作が緩慢になってしまう。

(2) ファイル数の上限 : Google デスクトップの初回インデックスでは, ドライブにつき10万個までのファイルしか登録されない。10万以上のファイルを格納したドライブがある場合は, ファイルを移動させたりオープンしたりしたときにリアルタイムでインデックス作成することもできる仕様であるので, 検索の効率を考えれば, 10万個未満のファイル数単位でディスクパーティションを切るようにすることが望ましい。表1からわかるように概ね1年間の国公私立試験問題が5万件程度

であるから、1年ごとに一つのディスクパーティションにすればよく、初期設定時にそのようにしておくといよい。

(3) ファイルサイズの上限: Googleデスクトップの仕様により、ファイルの先頭から約75,000文字までしか検索しない。試験問題データベースでは1つの設問ごとにファイルとしているため、1つのファイルサイズが巨大となることはないためにその心配をする必要はないが、備忘のために記録しておく。

4.2 Microsoft 製品の起動

一般ユーザ向けに作られた Microsoft 製品はターミナルサービスを提供しているサーバ上では動作しない。なぜなら一つのライセンスでそのアプリケーションを複数の人が利用できてしまうからである。このため、このような利用の場合にはボリュームライセンスを予め取得しておく必要がある。Xam, センターTen, TrueTeller などのアプリケーションは Word などの Microsoft 製品のインストールを前提としており、TS アプリケーションで Xam 等を起動した場合には(クライアント側ではなく)サーバ側の Microsoft 製品を起動してしまう。

4.3 アプリケーションのライセンス

Xam, センターTen については購入したソフトを1台のマシンに限ってインストールできる。本利用はその範囲内である。TrueTeller については製品自体がネットワークを介して複数で利用することを想定している。

5 おわりに

大学入試センターの現在の唯一の製品は良質な試験問題であり、その作題支援に向けて我々研究開発部は日々腐心している。市販のアプリケーションはそれなりに多くのユーザをもち、また歴史もあるために改良も加えられている。したがって我々は同じような製品を再度作ることはせずに、これらを利用しつ

つ、我々だけが有する試験問題統計データベースとの連動を図るというアプローチを今回、採用した。その際に、できるだけデータ保守性を高めるよう努めた。たとえば JC 教育研究所が定める科目コードと統計情報 DB で設定している科目コードとの対応付けについては、簡単な Csh スクリプトが担当し、その保守、修正について容易になるよう工夫した。

本システムについては、ネットワークの利用環境においてまだ若干整備すべき事項が残っているが、本システムと同等のシステムを作題部会に提供することにより、今後作題支援に供することができると期待している。なお本研究については、科学技術研究費補助金(基盤研究(B), 課題番号 21300320)の助成を受けた。

参考文献

- 安納順一 (2008). ターミナル・サービスによるクライアントの仮想化, 連載: Windows Server2008 の基礎知識—新サーバ OS で何が変わるのか?—, アイティメディア株式会社.
- 安納順一 (2009). TS ゲートウェイの環境設定, 連載: Windows Server2008 ターミナル・サービス・ゲートウェイ実践構築入門, アイティメディア株式会社.
- Google (2011). 企業向けデスクトップ. 2011年9月2日 <<http://googledesktop.blogspot.com/>> (2011年12月9日)
- 石岡恒憲・大津起夫・橋本貴充 (2010). CenterTen(センター試験問題データベース)と研究開発部作成による統計試験情報データベースとの統合システムの開発, 研究開発部リサーチノート RN-10-01.
- JC 教育研究所(2009a). 大学試験問題データベース センターTen2009 ユーザーズガイド.
- JC 教育研究所(2009b). 全国大学入試問題データベース Xam ユーザーズガイド.
- Microsoft TechNet (2008). TS RemoteApp のステップバイステップガイド, 2008年7月, <[http://technet.microsoft.com/ja-jp/library/cc730673\(WS.10\).aspx](http://technet.microsoft.com/ja-jp/library/cc730673(WS.10).aspx)> (2011年12月9日)
- 大津起夫・橋本貴充・石岡恒憲(2008). センター試験に関わる統計情報の蓄積と利用に関する研究, 平成17年度~19年度共同研究報告書.