

PDF 利用申請データの DB 登録処理方法の紹介

総合情報基盤センター 技術専門職員 畑 篤

1. はじめに

富山大学総合情報基盤センターの一部利用申請で、PDF を利用した利用申請を導入しています。この申請方法は、フォーム付き PDF 利用申請様式に必要な事項を入力し、PDF ファイルをメールでセンター宛に送付する申請です。

センターでは、PDF フォームに入力されたデータを抽出し、データベースへインポートするまでの作業を自動で行っています。

ここでは、Microsoft Visual Basic.net を利用し、PDF 入力データをデータベースに登録する自動化処理について紹介します。

2. Acrobat の起動

データの抽出方法として、一括に抽出する方法と 1 件ごとに抽出する方法があるため、抽出方法によって、Acrobat ファイルが異なります。そのため、利用ファイルに応じた、アプリケーション (Acrobat pro) を起動させます。また、Acrobat 起動後は、処理が終了するまで次の処理には進みません。

リスト 1 はファイルに関連付けられたアプリケーションを開き、終了するまで待機するプログラムソースです。

リスト 1

```
p = System.Diagnostics.Process.Start("PDF ファイル名")
p.WaitForExit()
```

System.Diagnostics.Process.Start : ファイルに関連付けられたアプリケーションで開く

Object. WaitForExit : 起動したアプリケーションが終了するまで待機する

2-1 一括での PDF 申請データの抽出

PDF ファイル名を PDF 収集用データセットファイルとし、Acrobat pro を起動させます。起動後、PDF 収集用データセットファイルに各申請 PDF ファイルを格納します。

引き続き格納したファイルを CSV 形式で出力し、PDF 収集用データセットファイルを閉じます。

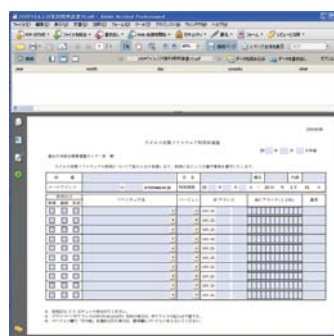


図 1
PDF 収集用ファイル

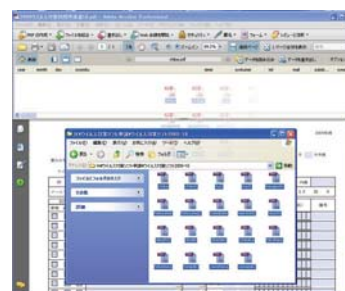


図 2
PDF 収集ファイルへ
申請データをコピー

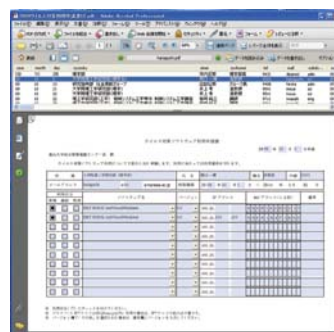


図 3
PDF 収集ファイルへ
格納された申請データ

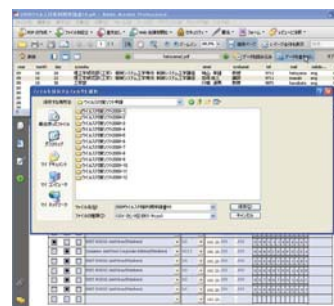


図 4
申請データを CSV 形
式での抽出

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	year[0]	month[0]	day[0]	syozoku[0]	simei[0]	syokumei[0]	tel[0]	mail[0]	subdomain[0]	year[0]	month[0]	day[0]	year[0]	month[0]	day[0]	ip1[0]	ip2[0]	mac[0]	ip1-1[0]	ip1-2[0]	mac[0]	bikou1[0]	n2[0]	cc[0]	cc[0]
2	chiba.pdf	9	10	5	工学部機械系 志子 助教	9685	chiba	ent	9	10	1	10	1	0	0	Symantec 111.0.4	CCC	CCC	9982901F843				0	0	
3	chiba2.pdf	9	10	5	工学部機械系 志子 助教	9685	chiba	ent	9	10	5	10	1	0	0	Symantec 111.0.4	EEE	EEE	997A2A757A6				0	0	
4	fuji.pdf	9	10	15	地域連携推進課 小淵 准教授	6390	obuchi	cte	9	10	15	10	1	0	0	Symantec 111.0.4	184	255	998F16FC8D07				0	0	
5	fuji2.pdf	9	10	13	研究開発推進課 岡田 副学長	9408	okada	adm	9	10	13	10	1	0	0	Symantec 110.2.1	184	179	998BC6981806				0	0	
6	hataesu.chip	9	10	28	経営学部経営学 立 助教	9221	shiotani	eco	9	10	28	10	0	1	0	0	1-1-2	7	AAA	993747E411				9	0
7	hatoyama.pdf	9	10	5	工学部 亀田中 秀征 助教	9608	stanaka	ent	9	10	5	10	1	0	0	ESET NOD32	4		99D99FAD0182				1	0	
8	hatoyama2.pdf	9	10	8	人間発達科高原 須美子 講師	9242	takahara	edu	9	10	7	10	1	0	0	ESET NOD32	4	255	999D0EE68D10				0	0	
9	kawabata.pdf	9	10	2	大学院理工学 川村 正樹 助教	9664	takemura	ent	9	10	2	10	1	0	0	Symantec 111.0.4			99788AA86981				1	0	
10	morimoto.pdf	9	10	1	地域連携推進課 堀 准教授	9927	hanikiri	ent	9	10	1	10	1	0	0	ESET NOD32	4	255	997926AC4D04				0	0	
11	obuchi.pdf	9	10	29	理工学部水山 友夫 助教	9762	yamahara	ent	9	10	2	1	10	0	1	0	Symantec 111.0.3	255	255	998E6973959				0	1
12	obuchi2.pdf	9	10	22	大学院理工学 山崎 准教授	9471	yamazaki	sci	9	10	22	10	1	0	0	Symantec 111.0.4	255	255	998877B6AFF				0	0	
13	okada.pdf	9	10	21	大学院理工学 藤久 准教授	9541	fuji	sci	9	10	21	10	1	0	0	Symantec 111.0.4	255	255	998A381EAD05				0	0	
14	shiotani.pdf	9	10	22	大学院理工学 藤井 准教授	9541	fuji	sci	9	10	26	10	1	0	0	Symantec 111.0.4	255	255	999F1F33C8D6				0	0	
15	stanaka.pdf	9	10	5	大学院理工学 立 准教授	9262	hataesu	sci	9	10	5	10	1	0	0	ESET NOD32	4		999D0EE0217				1	0	
16	takahara.pdf	9	10	28	理工学部経営学 立 准教授	9711	hatoyama	ent	9	10	28	10	0	1	0	Symantec 110.2.1	255	255	998F6037E63				0	0	
17	takemura.pdf	9	10	22	理工学部経営学 純三 講師	9711	inasaki	ent	9	10	22	10	0	1	0	Symantec 110.2.1	255	255	9988F6037E63				0	1	
18	tanigaki.pdf	9	10	30	工学部 川原 達男 助教	9871	kawabata	ent	9	10	30	10	0	0	1	Symantec 110.2.1	255	255	998A8AE6B40B				1	0	
19	yamahara.pdf	9	10	26	工学部 桑本 晃司 助手	9652	morimoto	ent	9	10	26	10	1	0	0	Symantec 111.0.4	255	255	99785AA04C00				1	0	
20	yamazaki.pdf	9	10	27	地域連携推進課 小淵 准教授	6390	obuchi	adm	9	10	26	10	1	0	0	Symantec 111.0.4	255	255	9983B641BE70				0	0	
21																									
22																									

図 5 取り出した CSV データ

2-2 1件ごとの PDF 申請データの抽出

申請 PDF ファイルを選択し Acrobat pro を起動させます, 利用申請データを xml 形式で出力させます。引き続き申請 PDF ファイルを閉じます。

リスト 2 は出力した xml データです。

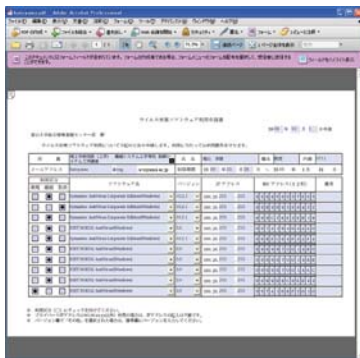


図 6 PDF 申請データファイル

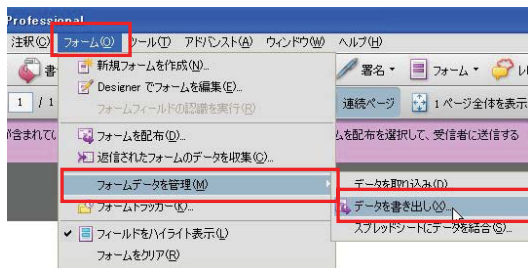


図 7 XML 形式データでの抽出

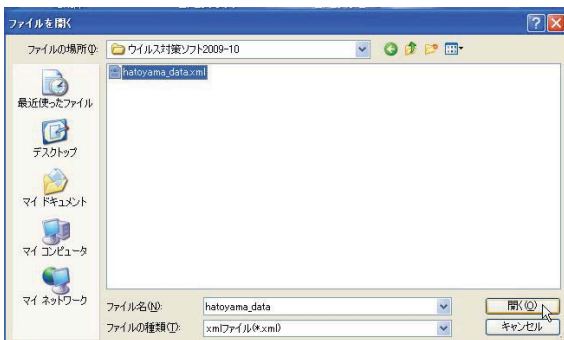


図 8 XML 形式データの保存

リスト 2

```
<?xml version="1.0" encoding="UTF-8" ?>
<topmostSubform>
<year>09</year>
<month>10</month>
<day>28</day>
<syozoku>理工学研究部(工学) 機械システム
工学専攻 制御システム工学講座</syozoku>
<simei>鳩山 幸雄</simei>
<syokumei>教授</syokumei>
<tel>9711</tel>
<mail>hatoyama</mail>
<subdomain>eng</subdomain>
<syar>09</syar>
<smmonth>10</smmonth>
<sdar>28</sdar>
<eyear>10</eyear>
<n1>0</n1>
<c1>1</c1>
<d1>0</d1>
<soft1>Symantec AntiVirus
Corporate Edition(Windows)</soft1>
<ver1>10.2.1</ver1>
<ip1-1>255</ip1-1>
<ip1-2>255</ip1-2>
<mac1>9988F6037E63</mac1>
<bikou1 />
<n2>0</n2>
<c2>0</c2>
<d2>1</d2>
<soft2>Symantec AntiVirus
Corporate Edition(Windows)</soft2>
<ver2>10.2.1</ver2>
<ip2-1>255</ip2-1>
<ip2-2>255</ip2-2>
<mac2>99B9D9369452</mac2>
<bikou2 />
----- 途中省略 -----
<n10>1</n10>
<c10>0</c10>
<d10>0</d10>
<soft10>ESET NOD32
AntiVirus(Windows)</soft10>
<ver10>3.0</ver10>
<ip10-1>255</ip10-1>
<ip10-2>255</ip10-2>
<mac10>997A1DBF2010</mac10>
<bikou10 />
</topmostSubform>
```

3. DB へのインポート

3-1 CSV データの読み込み

ADO.NET を利用し、CSV データをデータベーステーブルへ格納し、データ整形を行います。

リスト 3 は CSV データ読み込みソースです。

リスト 3

```
Dim cn As New OleDb.OleDbConnection()
Dim c As New OleDb.OleDbCommand()
fn = "ファイル名"
cn.ConnectionString =
    "Provider=Microsoft.Jet.OLEDB.4.0;"
    & "Data Source=" & dn & _ "Extended
Properties=""text;HDR=no;FMT=Delimited""

With c
    .CommandText = "Select * From " & fn
    .Connection = cn
End With

da.SelectCommand = c
da.Fill(ds)
```

OLE DB : Microsoft 社によって開発された、データベースの種類によらず統一的手法でデータベースにアクセスするためのプログラミングインターフェース

OleDb. Connection : プログラムでデータベースに接続するための接続文字列 (設定を文字列で指定)

OleDb.OleDbCommand : データソースをデータベース管理システムに保存

ConnectionString : データベースに接続する為の情報を設定するプロパティ

CommandText : 読み取り専用でのテーブルの内容を参照

SelectCommand : データアダプター (データベースとデータセットの間をつなぐ「アダプタ」) と Select コマンドを関連付ける

Fill(ds) : データセットにデータを格納

3-1 xml 形式の取り込み

1 行目は XML の宣言、2 行目はルート要素であるため 3 行目以降のデータを読み込みます。

ReadString で XML を読み込み XmlReader で読み込んだ XML ストリームから文字列を取り出します。

リスト 4 は xml データ読み込みソースです。

リスト 4

```
reader = System.Xml.XmlReader.Create(ファイル名)
reader.Read()
reader.Read()
reader.ReadString()

For i = 1 To 103
    reader.Read()
    dat(i) = reader.ReadString
Next i
reader.Close()
```

XmlReader.Create(ファイル名) : URI で指定された XML データ ファイルを読み取る

reader. Read : XML を読み込む

XmlReader.ReadString : 要素ノードまたはテキスト ノードの内容を文字列として読み取る

3-3 データベースの指定とテーブルを開く

Access にデータをインポートするため接続先のデータベースを指定し、当該テーブルを開きます。リスト 5 はデータベースとして Access 選択し、当該テーブルを開くソースです。

リスト 5

```
cn2.ConnectionString =
    "Provider=Microsoft.Jet.OLEDB.4.0;" & _
    "Data Source=" & "データベース名"

cn2.Open()
With rs
    rs.CursorType =
        CursorTypeEnum.adOpenDynamic
    rs.LockType =
        LockTypeEnum.adLockOptimistic Update
    rs.Open("master", cn2, , ,
        CommandTypeEnum.adCmdTable)
End With
```

CursorTypeEnum.adOpenDynamic : Recordset 内でのすべての動作を許可

LockTypeEnum.adLockOptimistic : Update メソッドを呼び出した場合にのみ、共有的ロックを使ってそのレコードをロック

3-4 テーブルにデータを格納する

データを新しいレコードに格納し、レコードに書き込みます。

リスト 5 は Access へのデータインポートのソ

ースです。

```

For i = 1 To m
    With rs
        .AddNew()
        .Fields("申請日").Value = dbdata(i, 1)
        .Fields("所属").Value = dbdata(i, 2)
        .Fields("氏名").Value = dbdata(i, 3)
        .Fields("職名").Value = dbdata(i, 4)
        .Fields("電話番号").Value = dbdata(i, 5)
        .Fields("e-mail").Value = dbdata(i, 6)
        .Fields("利用開始日").Value = dbdata(i, 7)
        .Fields("利用期限").Value = dbdata(i, 8)
        .Fields("利用区分").Value = dbdata(i, 9)
        .Fields("ソフト名").Value = dbdata(i, 10)
        .Fields("Ver").Value = dbdata(i, 11)
        .Fields("IP").Value = dbdata(i, 12)
        .Fields("Mac").Value = dbdata(i, 13)
        .Fields("備考").Value = dbdata(i, 14)
        .Update()
    End With
Next i
rs.Close()
cn2.Close()
    
```

Object.AddNew : 新しいレコードを作成する
Object.Fields("フィールド名").Value : レコードの各フィールド(項目)にデータを格納

申請日	所属	氏名
2009/10/01	理工学研究部<工	鳩山 幸雄
2009/10/01	理工学研究部<工	鳩山 幸雄
2009/10/01	理工学研究部<工	鳩山 幸雄
2009/10/01	理工学研究部<工	鳩山 幸雄
2009/10/01	理工学研究部<工	鳩山 幸雄
2009/10/01	理工学研究部<工	鳩山 幸雄
2009/10/01	理工学研究部<工	鳩山 幸雄
2009/10/01	理工学研究部<工	鳩山 幸雄
2009/10/01	理工学研究部<工	鳩山 幸雄
2009/09/29	経済学部	広中 和歌子
2009/09/24	地域連携推進機	甘利 明

図9 個別抽出したときのデータ

申請日	所属	氏名
2009/10/30	工学部	川端 達男
2009/10/30	工学部	川端 達男
2009/10/30	工学部	川端 達男
2009/10/30	工学部	川端 達男
2009/10/29	理工学研究部<工	山花 直夫
2009/10/29	理工学研究部<工	山花 直夫
2009/10/29	理工学研究部<工	山花 直夫
2009/10/29	理工学研究部<工	山花 直夫
2009/10/29	理工学研究部<工	山花 直夫
2009/10/28	理工学研究部<工	鳩山 幸雄
2009/10/28	経済学部経営学科	塩谷 立
2009/10/28	経済学部経営学科	塩谷 立
2009/10/28	理工学研究部<工	鳩山 幸雄
2009/10/28	理工学研究部<工	鳩山 幸雄
2009/10/28	理工学研究部<工	鳩山 幸雄
2009/10/28	理工学研究部<工	鳩山 幸雄
2009/10/28	理工学研究部<工	鳩山 幸雄
2009/10/28	理工学研究部<工	鳩山 幸雄
2009/10/28	理工学研究部<工	鳩山 幸雄
2009/10/28	経済学部経営学科	塩谷 立
2009/10/28	理工学研究部<工	鳩山 幸雄
2009/10/28	理工学研究部<工	鳩山 幸雄
2009/10/28	経済学部経営学科	塩谷 立
2009/10/27	地域連携推進機	小淵恵三
2009/10/26	工学部	森本 晃司
2009/10/26	工学部	森本 晃司
2009/10/22	理工学研究部<工	岩崎 純三
2009/10/22	大学院理工学研究	山崎 拓
2009/10/22	大学院理工学研究	藤井 博久
2009/10/22	理工学研究部<工	岩崎 純三
2009/10/22	理工学研究部<工	岩崎 純三
2009/10/22	理工学研究部<工	岩崎 純三
2009/10/22	理工学研究部<工	岩崎 純三
2009/10/22	理工学研究部<工	岩崎 純三
2009/10/22	理工学研究部<工	岩崎 純三
2009/10/22	理工学研究部<工	岩崎 純三
2009/10/22	理工学研究部<工	岩崎 純三
2009/10/22	理工学研究部<工	岩崎 純三
2009/10/21	大学院理工学研究	藤井 博久
2009/10/15	地域連携推進機	小淵恵三
2009/10/13	研究振興部 社会	岡田 勝也
2009/10/08	人間発達科学部	高原 須美子
2009/10/05	工学部 電気電子	田中 秀征
2009/10/05	工学部機械知能	千葉 圭子
2009/10/05	工学部 電気電子	田中 秀征
2009/10/05	大学院理工学研究	原口 一博
2009/10/05	大学院理工学研究	原口 一博
2009/10/02	大学院理工学研究	武村 正義
2009/10/02	大学院理工学研究	武村 正義
2009/10/01	工学部機械知能	千葉 圭子
2009/10/01	地域連携推進機	谷垣 禎一
2009/09/29	経済学部	広中 和歌子
2009/09/24	地域連携推進機	甘利 明

図10 一括抽出したときのデータ

4. おわりに

図9は、1つの申請ごとに、図10は一括で申請データをDBへ取り込んだ例です。このように、VB.netでAcrobatファイルを起動させ、DB(Access)へデータを取り込むことができます。Acrobatで保存するディレクトリを指定しておくことができないため、CSVやxmlファイルの保存先を確認する操作が必要になるといった手間が生じるので、今後は、このように保存先を指定せずにDBへ申請データをインポートできるよう改善を行っています。