

# データ・書式の共有による書類作成支援システムに関する研究

安田 愛† 藤澤 公也†

東京工科大学大学院 バイオ・情報メディア研究科†

## はじめに

会社や学校では、書類提出を求められる機会が多い。その中でも、一度出せば済むものだけでなく、その都度作成し、提出する必要がある書類もある。

既存の製品では、配布された書類を Excel や Word に変換し、電子データ化を行うものがある。電子データ化を行うことで、入力や保管が容易になる。しかし、電子データ化を行う際に書類の見た目が変わってしまうことがある。書類の見た目が違うことで書類が受理されない可能性がある。

最近では、書類も紙で配布されるのではなく、電子ファイルで配布されることも増えてきている。電子ファイルで配布されることによって、電子データとして保管することが容易となり、部分的に変更して提出することも多くなった。しかし、まだ書類の提出は電子データではなく、印刷した紙を提出する場合も多い。書類の処理も含めたすべての仕組みを電子化するシステムも提案されている[1]が、この場合はすべての利用者は新たなツールの使い方を覚える必要がある。

これらの問題点を解決するために、書類作成支援システム（図 1）の開発を行っている[2][3]。このシステムでは、利用する書類及び処理の手順を変えずに、書類作成者の手間を省くことを目的として、設計・実装を行った。

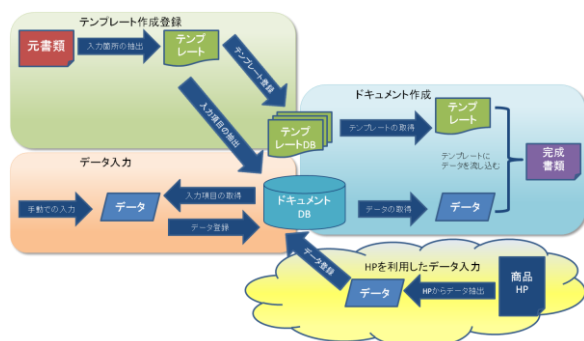


図 1 書類作成支援システム構成図

The documents generating support system with sharing templates and data

†Ai YASUDA, Kimiya FUJISAWA  
Graduate School of Bionics, Computer and Media Sciences, Tokyo University of Technology

入力データの共有を可能とするために、データベースを利用して入力データと書類テンプレートを管理する。紙に手書きで記述し、書類を作成する方法と、このシステムを利用して入力データを差し込むことで作成する方法が共存できるようになる。書類テンプレートとは、書類の入力箇所にフォームが付加されている PDF ファイルのことである。このシステムでは、データが入力された PDF ファイルが出力される。これを印刷することで、手書きで作成した書類と同様に、紙媒体で提出が行える。

このシステムの主な機能としては、「テンプレート作成登録」「データ入力」「ドキュメント作成」である。

## 簡易実装

入力データの共有を行い、書類の作成を行うシステムとして、書類作成支援システムの 3 つの機能を実装していたが、問題点として書類テンプレート作成を既存の市販のソフトウェアである「やさしく PDF へ文字入力 PRO v. 6.0」[4]を使用しているのがあげられる。このソフトウェアをインストールしているコンピュータを利用しないと書類テンプレートを作成できない。

この問題点を解決するために、Web ブラウザ上から書類テンプレート作成を行えるようにした（図 2）。サーバでは、元となる PDF を画像化し、HTML に埋め込んでいる。この画像に対して、

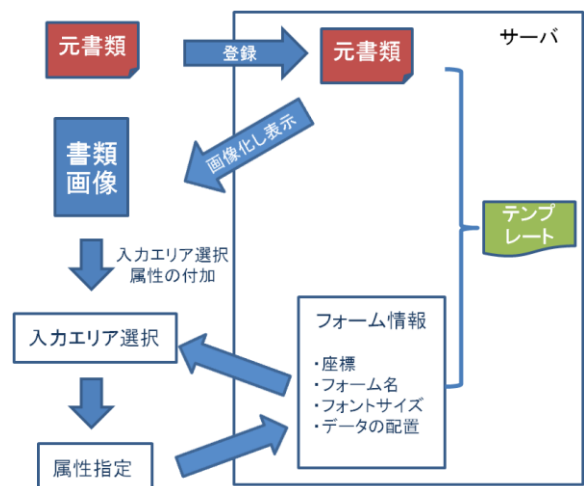


図 2 書類テンプレート作成の流れ

どこにどのようなフォームを埋め込んでいくかを指定していく。データ入力エリアの選択を範囲選択により行い、同時にフォーム名・データを差し込む際のデータの配置・フォントサイズ等を属性として付加し、これらの情報をサーバに送信している。サーバ側では、元の PDF の書類に対して、入力フォームを付加した PDF を生成する。フォームを付加する際には、送られてきた属性により、フォントサイズ等の設定を行う。ここで作成された PDF は、書類テンプレートとなる。

### システム設計

簡易実装を行って浮かび上がった問題点としては、入力箇所すべてに対して、入力エリア選択・属性指定を行わなくてはならない点である。

研究の目的は、書類作成支援システムの機能である書類テンプレート作成の部分に、画像処理を組み込むことで、書類テンプレート作成の処理の効率向上を行うことである。

書類テンプレート作成部分に、枠の検出機能を加える。枠の検出には、アップロードした元書類の PDF ファイルを画像化し、その画像に対して画像処理を行う。

枠の検出の流れ (図 3) は、まず Hough 変換を用いて画像に対して直線を抽出する。抽出した直線から、交点を算出する。算出された交点から、枠となる 4 点を算出する。それぞれ 4 点の間に線があるかを判別し、すべてに線があった場合は枠として検出される。

次に、検出された枠に対して、枠内に文字があるかの判別を行う。文字がないと判別された枠は入力エリアとなり、HTML に表示する際に入力エリアとして選択された状態になる。

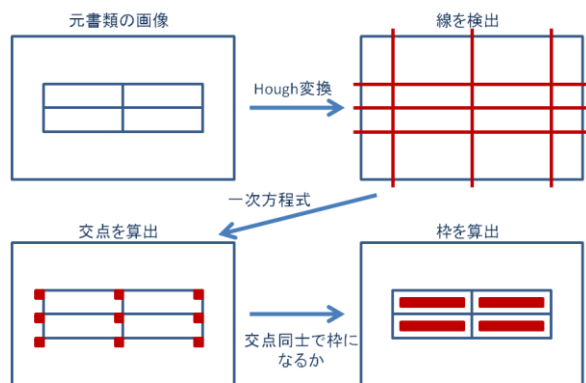


図 3 枠の検出の流れ

### 実装内容

書類テンプレート作成を Web ブラウザ上から行えるようにし、あらかじめ画像処理を行うことで枠を検出した。

元書類の PDF をサーバに送信した際、元書類の画像を作成し、HTML に埋め込む前に枠の検出する機能を実装した。この機能によって、HTML に表示される際には、検出された枠は入力エリアとして選択された状態となる。入力エリアとして足りない部分や修正を Web ブラウザ上で行い、書類テンプレートを作成する。

### 検証

本システムの検証としていくつかの書類を利用し、実際に書類テンプレートの作成を試みた。

対象とした書類は、Word から PDF 化を行った書類、Excel から PDF 化を行った書類、スキャナーから取り込んだ書類に対して行った。

元書類の PDF がどのように作成されたかによって、精度が異なった。二値化を行う段階での閾値の値によっても精度が異なった。閾値は、書類毎に大幅に異なり、適切な閾値を設定する必要があることがわかった。

### おわりに

本研究では、書類作成支援システムの機能である書類テンプレート作成部分に、画像処理を組み込むことで、書類テンプレート作成の処理の効率向上を行った。

今後の課題としては、二値化を行う際の閾値を書類によって、枠の検出が一番多くできたものに合わせられるようにすることによって、より精度の高い枠の検出が行えるようになる。

今後の展望は、システム全体を完成させることで、書類作成の簡易化が行えるようになる。

### 参考文献

- [1] 森廣, 宮崎, 靄, “Web 技術を利用したデータベースシステムによる業務効率化”, 大分工業高等専門学校紀要 第44号, PP.37-42, 2006
- [2] 安田愛・藤澤公也, “書類入力支援システムの設計と実装”, 第71回情報処理学会全国大会 1Q-1 (2009.3)
- [3] 安田愛・藤澤公也, “データ・書式の共有による書類作成支援システムの構築”, 第72回情報処理学会全国大会 5Q-4 (2010.3)
- [4] メディアドライブ株式会社「やさしくPDFへ文字入力 PRO v.6.0」 : <http://mediadrive.jp/products/ypdfp/index.html>