

# Framework for Tax Preparation

Charles Duan

April 14, 2019

## Contents

<b>1</b>	<b>Overview of Operation</b>	<b>2</b>
<b>2</b>	<b>Phase One: Computation</b>	<b>3</b>
2.1	Structure of Tax Form Data . . . . .	3
2.2	Form Submanagers . . . . .	4
<b>3</b>	<b>Phase Two: Form Field Marking</b>	<b>4</b>
3.1	Boxed Data . . . . .	5
<b>4</b>	<b>Phase Three: Form Filling</b>	<b>5</b>
<b>A</b>	<b>Additions to Forms</b>	<b>6</b>
A.1	1095-B . . . . .	6
A.2	1098 . . . . .	6
A.3	1099-DIV . . . . .	6
A.4	1099-R . . . . .	6
<b>B</b>	<b>Informational Forms</b>	<b>7</b>
B.1	Alimony . . . . .	7
B.2	Asset . . . . .	7
B.3	Biographical . . . . .	7
B.4	Business Expense . . . . .	8
B.5	Charity Gift . . . . .	8
B.6	Dependent . . . . .	9
B.7	Estimated Tax . . . . .	9

B.8 HSA Contribution . . . . .	9
B.9 Home Office . . . . .	9
B.10 State Estimated Tax . . . . .	10
B.11 State Tax . . . . .	10
B.12 Traditional IRA Contribution . . . . .	10
B.13 Partner . . . . .	11
B.14 Partnership . . . . .	11
B.15 Real Estate . . . . .	12

This is a software system for automating the computation and filling of tax forms. Much like commercial software for tax preparation, the object of the system is to allow a user to enter data from received tax forms and other sources, and then answer interview questions, to produce completed tax forms ready for filing.

Unlike commercial tax software, however, this system is not intended to comprehensively cover all tax forms and options. Instead, a second object of this system is to provide a framework to make it reasonably easy for programming-adept users to add tax forms as necessary to their personal situations and needs.

## 1 Overview of Operation

Use of the system proceeds in three phases.

The first phase is the computation of form values. The user enters relevant data from received tax forms and other sources into files, and then writes a short script that imports that data and computes the values for tax forms to be prepared. This produces a plain text file containing the line item data to be filled into the tax forms.

The second phase is the identification of form fields. For each tax form to be filled in, the user provides the PDF file for the blank form and also the file of line item data. A program displays a graphical interface where the user can click on the blank spaces where each line item should be entered. The result of this is a second plain text data file that lists the coordinates of each of the form fields.

The third phase is the actual filling of the forms. The user writes another short script that receives the data files from the first and second phases, and places the line items from the first-phase file into the PDF form. If any changes need to be made subsequently, the user may simply rerun the

first-phase computation script and then rerun the third-phase filling script, thereby updating all of the tax forms automatically.

## 2 Phase One: Computation

This section describes the operation of the first phase, computation of tax form line items.

### 2.1 Structure of Tax Form Data

The basic data structure used throughout the system is the `TaxForm`, which in essence is simply a named table mapping line numbers to values. Values may be of several types:

**blank zero** A single dash, representing a zero value.

**number** A number, possibly with a decimal point and possibly led with a minus sign.

**date** Formatted as mm/dd/yyyy.

**array** A list of values, each of which is one of the above types. Arrays may be specified in two forms: A comma-separated list surrounded by square brackets, or another way described below.

**text string** Anything else.

Another data type is “boxed data,” described below; it is represented in a format surrounded with angle brackets and is essentially a text string or number.

A plain text file may store one or more forms according to the following format. A form begins with a line starting with the word “Form”, a space, and the name of the form. Lines of the form follow; they must be indented with whitespace, and the line number should be separated from the value with whitespace. For example:

```
Form W-2
    first_name      John
    last_name       Doe
```

a	123-45-6789
b	98-7654321
c	Acme Widgets Co.
1	50000
2	15000

Further forms may be included in the same file, delineated with the start word “Form” at the beginning of a line.

An array may be constructed by placing each array element on a separate line in the file, replacing the line number with a double quote mark. For example:

```
A  15
"   18
"   20
```

is identical to “A [15, 18, 20]”.

## 2.2 Form Submanagers

The main FormManager for an entity may contain “submanagers,” namely references to other FormManager objects for different entities or tax years. This is a convenient way to draw information from a related filing.

Submanagers are added using the `add_submanager` method, which requires a name for identifying the submanager. Common names are `:last_year` and `:spouse`. The `add_submanager_from_file` method creates a new FormManager, imports data from the file, and adds a new submanager.

## 3 Phase Two: Form Field Marking

The program that operates this phase is “`mark_fields.rb`”. It must be supplied with the tax form data produced in phase 1, using the “-i” argument. With no further arguments, the program lists the forms for which the fields have not been marked yet.

To mark fields in a form, enter the form name as the first non-option argument to the command. If the form has not been marked yet, you must

provide the name of the PDF file for the form as a second command line argument.

The program will automatically save the location of the PDF file and the coordinates of the marked fields to a file, which by default is named “pos-data.txt”. This file name may be changed using the “-p” argument.

### 3.1 Boxed Data

In some cases, the PDF form will have individual boxes for each character or portion of text, rather than a single field for all the text. The “boxed data” feature enables filling in these form fields correctly.

To use this feature, the relevant TaxForm should call the method “boxed\_line(line, count, split)” where line is the line number to be placed in boxes, count is the number of boxes available, and split is how to separate the line’s data into segments (the default is an empty string, which means each character is separated individually). All this does is mark the relevant line in the form as a boxed-data line; computation proceeds unchanged.

When the form field marking program is invoked, it will detect boxed-data lines and treat them like arrays. Thus, each box will need to be marked individually. However, if there are more letters than boxes available, the program will truncate data values to fit.

## 4 Phase Three: Form Filling

To fill in the forms, create a script that creates a FormManager object with the tax forms from phase 1 imported. Then create a MultiFormManager to handle the filling of forms. For each form to be filled in, use the fill\_form method of the MultiFormManager.

The MultiFormManager can automatically produce continuation sheets when there is insufficient space. This requires providing two options:

**continuation\_bio** Biographical text for the continuation sheets, just for identification purposes.

**continuation\_display** Either :show, which will print the continuation page to the screen; :raw, which will print the raw troff code; or :append, which will compile the troff code and append the continuation sheet to the filled tax form.

## A Additions to Forms

### A.1 1095-B

This is the health coverage form. No information other than the items listed below is needed.

**hdhp?** Whether this plan was a high deductible plan that qualifies for HSA contributions.

**coverage** “individual” or “family”.

**months** The months of coverage for this plan. (Currently it is assumed that all persons on the form are covered for the same months.) These should be all-lowercase three-letter abbreviations of months.

### A.2 1098

**property** The name of the Real Estate form with which this 1098 corresponds.

**balance** The average mortgage balance on this loan. See Pub. 936. You can compute this by averaging the balances on January 1 and December 31, or you can multiply the interest paid by the lowest interest rate of the year.

### A.3 1099-DIV

**qualified\_exception** If this form shows qualified dividends, this flag must be set to indicate whether an exception applies. See 1040 instructions.

### A.4 1099-R

**2b.not\_determined?** The relevant checkbox under line 2b.

**2b.total\_distribution?** The relevant checkbox under line 2b.

**ira-sep-simple?** Whether the checkbox on the form with this name is checked.

**destination** Where this distribution went. It can be “Roth conversion” (currently the only one supported).

## B Informational Forms

### B.1 Alimony

Alimony received. See Pub. 504.

**amount** The amount of alimony received.

### B.2 Asset

A form for a capitalized asset.

**date** The date the asset was put into service.

**amount** The dollar value of the asset when purchased.

**179?** Whether the asset is a section 179 deductible asset.

**listed?** Whether the asset is a listed asset, see Pub. 946 chapter 5.

**dc\_category** The depreciation category for the asset in DC, see the FP-31 instructions, under Depreciation Guidelines.

**dc\_type** Choose from “reference”, “fixed”, or “other” to categorize the asset for DC Form FP-31, lines 1–3.

**description** A description of the asset.

### B.3 Biographical

Biographical information for an individual tax filer. Spouses should have a separate Biographical form.

**whose** The individual to whom this form pertains. May be “mine” or “spouse”.

**first\_name** The first name and middle initial.

**last\_name** The last name.

**phone** Telephone number.

**ssn** Social Security number, with dashes.

**home\_address** The first line of the home address.

**apt\_no** The apartment number for the address.

**city\_zip** The city, state, and ZIP code.

**birthday** The person's birthday.

**blind?** Whether the person is blind.

**occupation** Your occupation

## B.4 Business Expense

A deductible business expense. See IRS Pub. 535, chapter 11.

**date** The date of the business expense.

**amount** The dollar amount of the expense.

**category** The category of expense. “Meals” and “Utilities” will automatically be halved, the former per the IRS 50% rule and the latter on the assumption that 50% of the expense was for non-business purposes. Other common values are “Supplies”, “Travel”, and “Membership”.

**description** A description of the expense.

## B.5 Charity Gift

A charitable contribution.

**amount** The dollar value of the contribution.

**cash?** Whether the contribution was in cash (as opposed to in-kind).

**name** The name of the charity.

**documented?** Whether the donation was documented. See Pub. 526. This is checked if the contribution is over the \$250 threshold.



## B.6 Dependent

A dependent.

**name** The dependent’s name.

**ssn** The dependent’s social security number.

**relationship** The dependent’s relationship with the form provider.

**qualifying** Whether the dependent qualifies for a tax credit. Options are “child” for the child tax credit, “other” for the credit for other dependents, and “none” where the dependent is not qualifying.

## B.7 Estimated Tax

Estimated tax payment made to the IRS.

**amount** The amount of estimated tax paid.

## B.8 HSA Contribution

Records of contributions to an HSA account.

**ssn** The SSN of the beneficiary of the HSA account.

**contributions** Contributions made to the HSA, but not including employer contributions or HSA/Archer MSA rollovers or HSA funding distributions.

## B.9 Home Office

A portion of a home used as an office for a particular business.

**type** The type of business associated with this home office. Currently the only supported type is “partnership”.

**ein** The EIN of the relevant business.

**method** Either “simplified” or “actual”.

**sqft** Square footage of the home office area.

**daycare?** Whether the relevant business was a daycare.

**property** The name of the Real Estate form with which this home office is associated.

## **B.10 State Estimated Tax**

Estimated tax payment made to a state.

**amount** The amount of estimated tax paid.

**state** The two-letter code for the state.

## **B.11 State Tax**

A record of state taxes paid in the relevant tax year (usually for the previous year). This is the amount actually paid to the state tax office, (i.e., the amount due), not the total tax including what was already paid via withholdings.

**amount** The amount of tax paid.

**name** A description (not used for computation).

## **B.12 Traditional IRA Contribution**

A contribution to a traditional IRA.

**amount** The amount contributed.

**tax\_\_year** The tax year of the contribution.

**date** The date of the contribution.

## B.13 Partner

A single partner within a partnership.

**name** The partner's name.

**liability** Either "general" or "limited".

**nationality** Either "domestic" or "foreign".

**type** The entity type of the partner. Currently only "Individual" is supported fully. Other possibilities include "Estate". (This should also, in the future, include corporations, partnerships, trusts, nonprofits, and foreign governments, among others.)

**share** This partner's share of profits and losses, as a decimal fraction such that all the partners' shares add to 1. Currently it is assumed that shares of profits and shares of losses are equal.

**capital** This partner's percentage contribution of capital to the partnership, as a decimal fraction such that all the partners' shares add to 1.

**ssn** The SSN of the partner.

**country** The country of citizenship of the partner, relevant to Form 1065 Schedule B-1, part II, column iii.

**address** The first line of the partner's address.

**address2** The second line of the partner's address.

## B.14 Partnership

Biographical information for a partnership.

**name** The name of the partnership.

**address** The first line of the address.

**address2** The second line of the address.

**business** The line of business of the partnership, Form 1065 line A.

**product** The product or service of the partnership, Form 1065 line B.

**code** The business code, Form 1065 line C.

**ein** The employer identification number of the partnership.

**start** The start date of the partnership.

**accounting** “Cash”, “Accrual”, or other accounting method.

## **B.15 Real Estate**

A description of real property, used in a variety of contexts and to link other forms that deal with a particular property.

**property** A name for the property that will uniquely identify the property across tax forms.

**basis** The basis price for the property, i.e., the purchase value. See Pub. 551.

**sqft** The square footage of the property.

**purchase\_date** The date on which the property was purchased.