

Japan Collection ウェブサイト“社史”ページプロジェクト

Prepared by Asako Shiba, March 18, 2003

(本ドキュメント保存場所：特殊流通 > 社史 > shashi.doc)

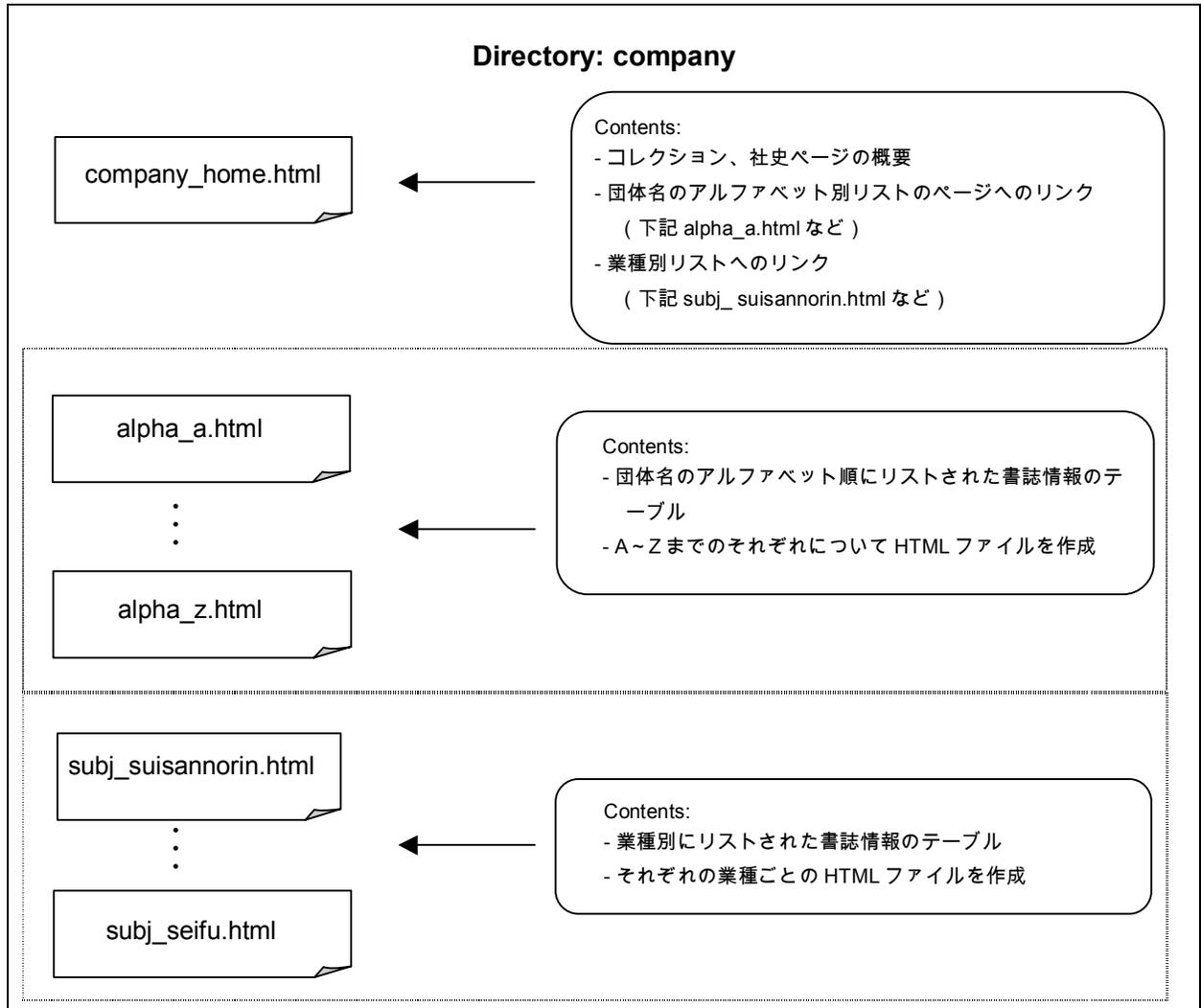
プロジェクトの概要

Asia Collection が所蔵する“社史”の書誌情報を含むデータベース (excel 形式) の内容を、Japan Collection のウェブサイトに載せることによって利用者に提供しようとするもの。プロジェクトは、Summer 2002 から Fall 2002 にかけて行われた。2002 年 12 月に、第 1 段階のウェブが完成した。

手順

I. 紙上でのデザイン案作成

実際の作業に入る前に、どのような形で社史のセクションをウェブに載せるのか、そのファイル構造やハイパーリンクの動作も含めたデザイン案を作成。Japan Collection ホームページからの直接リンクで社史メインページへ跳び、そこから、社史の内容の対象となる団体名のアルファベット別のリスト (A~Z の各ファイル)、及び、業種別のリスト (後述の業種別の各ファイル) にアクセスできるような構造にすることを決定。以下はファイル構造とそれぞれのファイルに含まれる内容。



上記の構成のデザインで、登紀子さんによる承認後、以下の作業へ。

II. 下準備：エクセルファイルの編集

元のファイルの保存場所：**特殊流通 > 社史 > Holdings.xls¹**

このエクセル形式のデータベースに含まれていたのは、団体への依頼を通して Japan Collection に寄贈された社史と、収集を始める前から所蔵していた社史（ただし selective）の書誌情報。総レコード数は、プロジェクト開始時で、205 件。レコードに含まれる項目は、業種、社名、タイトル、Call Number、PY (published year)であった。

¹現在は **Holdings_old.xls** として**社史フォルダ**の中にあり：このファイルは現在使用しておらず、新着分を追加する際はファイル **shashi_holdings.xls** を使用。

II.における作業の目的は、元ファイルを、後に HTML ファイル作成の際に使用することになる Dreamweaver にエクスポートして HTML ドキュメントの中に組み込む作業に取り掛かれるレベルにまで持って行く事である。

まず上記のファイルを、HTML ファイル作成のための編集用にコピーし、フォルダ“workfile 社史”の中に収めた。

II-1 データ加工

元ファイルの状態ではソーティングが不可能なため（図 1 参照）、それを可能にするために各列に業種をインプットし、さらに団体名の romanization の列を加えた。同時にヘッダーも全て英語に変更：Corporate Name; Type of Industry; Title; Pub. Year; Call No.（図 2 参照）。

	A	B	C	D	E	F
1	業種	社名	タイトル	CALL NUMBER	PY	
2	ガス					
3		中部ガス(株)	社史 中部ガス株式会社	HD9660 .G37 C83	1981	
4	食品					
5		山崎製パン	From a corn of wheat = 一粒の麦から:山崎	HD9057 .J34 Y36513 1996	1996	
6		白鶴	白鶴二百三十年のあゆみ	HD9397 .J34 H344	1977	
7		日本製粉	日本製粉社史	HD9056 .J34 N545 2001	2001	
8	たばこ					
9		朝鮮煙草	朝鮮煙草元売別株式会社史	HD9146 .K67 Y35	1931	
10						
11	新聞社					
12		油縄タイノマ	新聞二十年・油縄タイノマが生きた油縄	HD894.99 .O3785 O38 1979	1979	

図 1 元ファイル

	A	B	C	D	E	F	G
1		Type of Industry		Corporate Name	Title	Pub. Year	Call No.
2	1	機械	aishin takaoka	アイシン高丘	ともに歩んだ40年:アイシン高丘社史	2001	HD9710.3 .J34 A38 2000
3	2	新聞社	asahi shinbunsha	朝日新聞社	朝日新聞社史	1990	PN9409 .T63 A83 1990
4	3	新聞社	asahi shinbunsha	朝日新聞社	朝日新聞社70年小史	1949	PN5404 .A83 A83 1949
5	4	新聞社	asahi shinbunsha	朝日新聞社	大阪朝日新聞50年の回顧	1929	PN5404 .O83 O83
6	5	出版社	bungei shunju	文藝春秋	文藝春秋七十年史:本篇・資料篇	1994	Z463.3 .B86 1991
7	6	農協	Chichibubetsu Nogyo Kyodo Kumiai	秩父別農業共同組合	農業共同組合史	1958	HD1491 .J3 N65
8	7	金融	chosen ginko	朝鮮銀行	朝鮮銀行25年史	1934	HG3361 .K8 J.S45
9	8	金融	chosen shokusan ginko	朝鮮殖産銀行	朝鮮殖産銀行20年史	1938	HG3330.5 .C36
10	9	たばこ	chosen tabako	朝鮮煙草	朝鮮煙草元売別株式会社史	1931	HD9146 .K67 Y35
11	10	ガス	chubu gasu	中部ガス(株)	社史 中部ガス株式会社	1981	HD9660 .G37 C83
12	11	新聞社	chubu nihon shinbun	中部日本新聞	中部日本新聞二十年史	1962	KAJUYAMA 952
13	12	財界/財界	chubu zaibanki kankokai	中部財界中刊行会	中部財界中	1956	KA JUYAMA 1598

図 2 修正後ファイル：業種、団体名 romanization 追加²

² このファイルは FullAlphabeticOriginal.xls として、workfile 社史フォルダの中に入っている。

この時点で、業種名 (type of industry) についても、見直しと修正を行った³。サイト内の他のページとの整合性も考慮し、書誌情報以外は基本的に英語の使用を徹底するため、業種名には英語を用い、日本語訳を付ける形を採った (社史メインページと各ページのタイトルのみ)。表 1 が修正前、表 2 が修正後の、業種の分類である。なお、業種名は、今後社史コレクションが増えるにつれて、新たに加えられる可能性がある。

備考: 表 2 に列挙した業種名は、最終的に決定したもの。HTML のコーディングの時点でも、業種について少し変更を加えたので、**workfile 社史**フォルダ内のファイルの内容とは若干違う可能性がある。

表 1 修正前に採用されていた業種名

ガス	政府官公庁	船舶	繊維	百貨店	農協	運輸
食品	特殊法人 / 団体	機械	林業	保険	証券	財閥 / 財界
たばこ	教育	鉄道	貿易	電気	化学	生協
新聞社	鉄鋼	軽金属	放送	自動車	製薬	製紙
出版社	化粧品	炭坑	レジャー	金融	建設	

³ 業種見直しの際、次の資料を参考にした: 1) Japan Company Handbook (東京経済新報社) [Asia HC461 .J354]; 2) オハイオ州立大社史サイト <http://pears.lib.ohio-state.edu/dbase/IndustrySch.html>; 3) 総務省日本標準産業分類 <http://www.stat.go.jp/info/seido/9-1-05.htm>

表 2 修正後決定した業種名とその英訳

Fishery, agriculture, forestry (水産・農林業)	Agricultural cooperatives (農協)	Mining (鉱業)	Construction (建設)	Foods (食品)
Beverage, tobacco (飲料・たばこ)	Textiles (繊維)	Paper, pulp (紙・パルプ)	Chemicals (化学)	Pharmaceuticals (医薬品)
Petroleum products (石油)	Ceramics, cement (硝子・土石)	Steel (鉄鋼)	Nonferrous metal (非鉄金属)	General machinery (機械)
Electric machinery (電気機械)	Transport machinery (輸送機械)	Other manufacture (その他製造)	Electricity, gas, water (電気・ガス・水道)	Land transport (陸運)
Water transport (海運)	Other transport (その他運輸)	Broadcasting (通信・報道)	Newspaper (新聞社)	Publishing (出版)
Wholesale merchant (卸売業)	Retail trade (小売業)	Banks (銀行)	Securities (証券)	Insurance (保険)
Real estate (不動産業)	Leisure (レジャー)	Other services (その他サービス業)	Economic organization (経済団体)	Other organizations (その他団体)
Education (教育)	Government (政府官公庁)			

II-2 ソーティング

業種の見直しが済んだスプレッドシートは、団体名のアルファベット順、業種順にソートし、別名保存。

次にソート済みの各ファイルをさらに加工：団体名とその romanization の列が別になっていたが、ウェブページ上では同じセル内にこの2種類の情報を表示するため、手間ではあったが、romanizeされた団体名を、日本語の団体名のセルへカット&ペーストする作業を、各ファイルについて行った。理由は、テーブルをウェブページに載せた時に、横幅をあまりに広く取りすぎずに済むようにするためである。図3が、編集後のスプレッドシートのサンプル。この作業済みのファイルは、**workfile 社史**フォルダの中に、**FullAlphabetic.xls** と **FullSubject.xls** として保存されている。

	A	B	C	D	E
1	Corporate Name	Type of Industry	Title	Pub. Year	Call No.
2	アイシン高丘 Aishin Takaoka	機械	ともに歩んだ40年:アイシン高丘社史	2001	HD9710.3 .J84 A38 2000
3	朝日新聞社 Asahi Shinbunsha	新聞社	朝日新聞社史	1990	PN9409 .T63 A83 1990
4	朝日新聞社 Asahi Shinbunsha	新聞社	朝日新聞社70年小史	1949	PN5404 .A83 A83 1949
5	朝日新聞社 Asahi Shinbunsha	新聞社	大阪朝日新聞50年の回顧	1929	PN5404 .O83 O83

図3 修正後ファイル：最終

III. HTML ファイルの作成

HTML ファイルの作成にあたっては、Macromedia 社の Dreamweaver ver.4（日本語版）を使用。

III-1 テンプレートの作成

既存の Japan Collection の他のページの概観を基本に、アレンジを加え、社史のメインページ、書誌情報のテーブルを含む全てのページ共通のテンプレートを作成。

III-2 エクセルファイルの Dreamweaver への移行

II.の手順で用意したエクセル形式のスプレッドシートを、DWの機能を利用して、前述のテンプレートファイルにインポートする予定であったが、問題が発生。以下は、本プロジェクトとほぼ同時期に進行していた梶山プロジェクトドキュメントの引用。

前回の朝鮮関連タイトルの経験では、エクセルファイルのテーブルをDWへ直接取り込むことができたはずだったが、今回はそれが不可能であることが判明した（通常の読み込み方法は、DREAMWEAVER4 ユーザーマニュアル186ページに解説あり）。明確な原因は不明だが、漢字など日本語がいわゆる文字化けしてしまったのである。おそらく、今回は日本語版のウインドウズ98を使っていたのが、今回、英語版のウインドウズ2000を日本語化して使っていることに関連した問題だと思われる。よって、トキコさん経由で、他者の助けを借りた結果、エクセルファイルのテーブルをいったん、ワードファイルにコピーし、それをウェブページとして保存したうえで、DWで読み込むこととなった。手順は以下の通りである。...

上記の問題を本プロジェクトでも経験したため、先例にならって、以下の方法でエクセルファイルをDWに取り込んだ。

1. 編集済みのエクセルファイルを開き、インポートしたいスプレッドシートのテーブルを選択・コピーする。
2. MS ワードの新規ドキュメントを開き、1.でコピーしたテーブルをここにペーストする。その際、テーブルが新規ファイル上のページからはみ出るかもしれないが、それを無視したまま、File メニューにある **Save as Web Page** で別名保存する。
3. 新規 DW のウインドウ (ファイル) を開き⁴、File メニューから「読み込み」→「Word HTML の読み込み」を選択する。すると、小ウインドウが現れるので、そこで、先にテーブルを保存した MS ワードのファイルを指定して、OK をクリック。さらに別のウインドウが、読み込みかつワードファイルのクリーンアップをするかどうか質問して来るので OK をクリックし、実行する。
4. しばらくすると、テーブルが現れる。しかし、この段階では、テーブルにあった罫線が反映されておらず、これを復活させる必要がある。そこで、テーブル (スプレッドシート) が置かれたレイアウトテーブルをマウスで選択したうえで、オブジェクトウインドウにある「スタンダードビュー」を選択する。すると、そのレイアウトテーブルのプロパティが表示される (これが表示されない場合は、ウインドウメニューにある「プロパティ」を再度選択する)。そこにある「セル余白」と「ボーダー」という小さな小窓にある数字 0 をそれぞれ 1 (または他の整数) に変更すると、罫線が復活する。

III-3 DW での HTML ファイルの編集

上記の手順でインポートしたテーブルのファイルに、先に作成していたテンプレートを挿入していく、という作業を、各テーブルにつき行っていった。

具体的には：

1. DW の「表示」メニューから「コードとデザイン」を選択して、コードビューとデザインビュー両方を表示させる。
2. 先に作成していたテンプレートのコードをコピーし、編集するファイルのコードビューのフィールドにペーストする。
3. テーブルが適切な位置に来るように、コードを編集する。
4. 適宜編集を加えて、ファイルを完成させる。

⁴ 先に作成してあったテンプレートへのインポートは不可能であった。文字化けが発生。

IV. 修正

テーブルをインポートし、HTML ファイルを編集する過程で、書誌情報およびデザイン面においても随時修正を加えた。代表的なものを以下に記す。

テーブルのタグ削除

MS ワードのファイルを DW にインポートした時点である程度のタグのクリーンアップは行われているが、それでもまだ余分なタグや属性がかなり見られたので、さらにクリーンアップを行った。DW の「編集」メニューの下にある「検索/置換」機能を利用して、例えば、以下のようなタグや属性（下線部分）を一括削除した。

例 1) <p>（<td>内にあったもの）

例 2) ここに内容

書誌情報の確認・修正

この時点まで、書誌情報の包括的なチェック作業を行っていなかったため、Voyager のレコードとのつきあわせを行った結果、かなり間違いが見つかり、随時 Call number、タイトルなどの修正を行った。

テーブルのヘッダー修正

Call no. のセルに「(Location: EAST)」を追加した。

上記のような修正を経て出来上がったページのサンプルが以下の図 4。

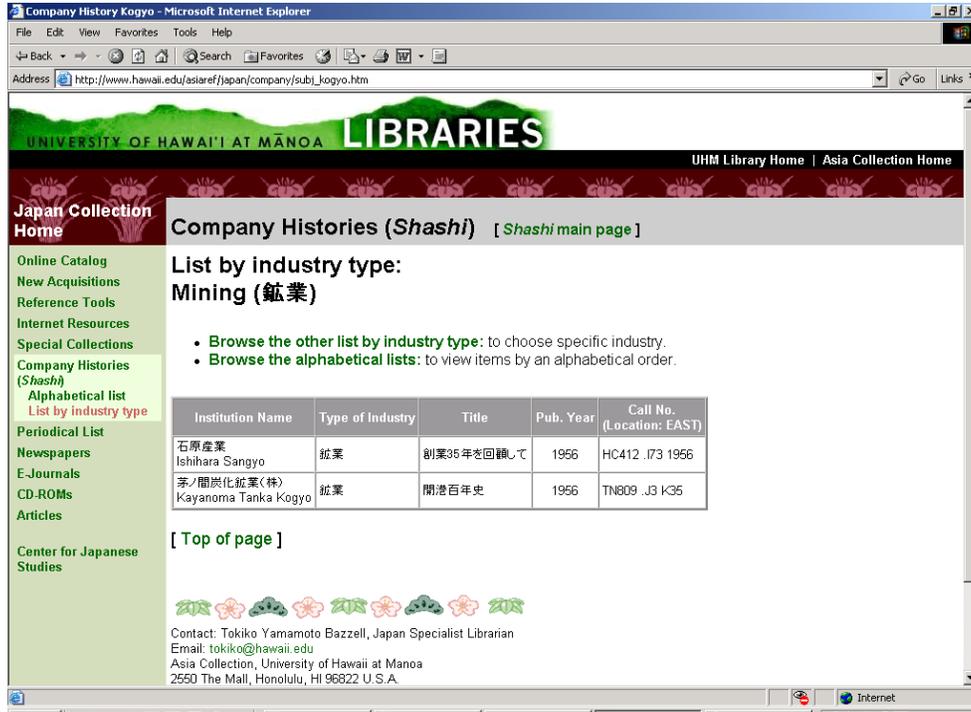


図4 完成ページスクリーンショット

全てのページの修正と登紀子さんの承認がおりた時点で、社史以外の Japan Collection の全てのページに、社史メインページへのリンクを挿入（図5参照）。



図5 社史ページへのリンクの挿入

反省

今回は、書誌情報のチェックを HTML ファイルの完成後に行ったが、エクセルファイル編集の時点でデータに関する修正は徹底して行うべきであった。HTML ファイルの編集は、1 タイトルにつき、2 回行わないといけなかったためである（アルファベット別リストと、業種別リスト）。