

仕 様 書			
件 名	戦史叢書の生成 AI 用テキストデータの作成	作成	戦史研究センター

1. 総 則

1. 1 適用範囲

この仕様書は、戦史叢書を生成 AI による検索・分析に活用するためのテキストデータ作成について規定する。

2. 役務に関する要求

2. 1 作業内容

契約相手方は以下の作業を実施する。

同所が保持する戦史叢書中、下表に示す巻の PDF データの支給を受け、以下の作業を行う。

N o	巻	史料名	枚(見開き)
1	8	大本営陸軍部（1）昭和15年3月まで	343
2	20	大本営陸軍部（2）昭和16年12月まで	372
3	35	大本営陸軍部（3）昭和17年4月まで	366
4	59	大本営陸軍部（4）昭和17年8月まで	342
5	63	大本営陸軍部（5）昭和17年12月まで	334
6	66	大本営陸軍部（6）昭和18年6月まで	368
7	67	大本営陸軍部（7）昭和18年12月まで	333
8	75	大本営陸軍部（8）昭和19年7月まで	302
9	81	大本営陸軍部（9）昭和20年1月まで	319
10	82	大本営陸軍部（10）昭和20年8月まで	289
11	91	大本営海軍部・連合艦隊（1）開戦まで	307
12	80	大本営海軍部・連合艦隊（2）昭和17年6月まで	274
13	77	大本営海軍部・連合艦隊（3）昭和18年2月まで	295
14	39	大本営海軍部・連合艦隊（4）第3段作戦前期	307
15	71	大本営海軍部・連合艦隊（5）第3段作戦中期	325
16	45	大本営海軍部・連合艦隊（6）第3段作戦後期	329
17	93	大本営海軍部・連合艦隊（7）戦争最終期	267
18	65	大本営陸軍大東亜戦争開戦経緯（1）	290
19	68	大本営陸軍大東亜戦争開戦経緯（2）	312
20	69	大本営陸軍大東亜戦争開戦経緯（3）	361
21	70	大本営陸軍大東亜戦争開戦経緯（4）	323
22	76	大本営陸軍大東亜戦争開戦経緯（5）	345
23	100	大本営海軍大東亜戦争開戦経緯（1）	222
24	101	大本営海軍大東亜戦争開戦経緯（2）	308
		計	7633

(1) PDF のテキスト化作業

戦史叢書のフォーマット（縦書き、縦二段組）に合わせ、以下の要領に従い NDL-OCR 等の AI-OCR を用いてテキスト化作業を行う。精度に関して、単純な文字単位ではなく、縦書き横書き書字方向や行の抜け漏れがなく、文書を読み順に正確に認識できることとし、認識精度は 99% 以上を追求する。

ア 本文・表

漢字、その他の文字のテキスト化については、構造解析等により、正確な読み順にて文字化を行うとともに、極力対象文献の表記通り正確な再現に努めること。特に縦書きに対して横書き表現としての認識間違いやタイトル・本文等の行の認識漏れをなくすこと。また、項目と値から構成される箇条書き（例：□□□部隊、□□□人等）については行内でテキストが改行されないように留意すること。また、一行に意図的に一定のスペースが入っていたりする箇条書き（例：海軍中将 小沢治三郎）についても行内の改行をなくすよう努めることとするが、精度保証の対象外とする。

テキスト化の範囲に関して、ヘッダー、フッター内の文字（ページ番号、章番号、タイトル等）、脚注番号及びルビは、テキスト化から除外する。なお、書籍タイトル、章、節、項、本文、頁番号については構成要素として区別して表示できること。また、注についても注番号、および注の本文については、データ上、構成要素として区別して表示できることとする。

表、図については可能な限り表、図内のテキスト化を行うこととし、表構造としてのデータ保持及び読み順補正は不要とし、精度保証の対象外とするが、全文検索が可能な形のデータとして維持するものとする。

イ テキストのデータフォーマットは、マークダウン、JSON、XML、CSV など、AI 技術を利用した文字入力による質疑応答用途に利用するために変換できる形式とし、ファイル形式や構造については、官側の承認を得るものとする。

以上、上記によりがたい場合は官側と協議して承認を得るものとする。

(2) テキストデータ付き PDF の作成

PDF のテキスト化作業によって得られたテキスト情報を透かしとして入れた PDF ファイルを作成する。テキストデータについては画像上の各文字に対応した形で入れるものとする。

以上、上記によりがたい場合は官側と協議して承認を得るものとする。

(3) 描画結果データの作成

OCR にて読み取り、テキスト化した箇所と該当テキストを原典画像上に示したデータ、PDF ファイルの 1 ページを 1 画像とし、その画像の読み取り対象箇所（行）を矩形で囲み、テキストの変換結果を当該箇所の傍に本文文字と異なる色を使用した小文字で表示すること。

2. 2 作業期間

契約締結日から令和 8 年 3 月 27 日（金）まで

2. 3 作業場所

契約相手方作業所等

ただし、官側から要請があった場合は速やかに協議する必要があるため、作業場所は関東甲信越地方に限定する。

2. 4 準 拠

契約相手方は、契約後速やかに官側と作業打合せを行い、作業内容等の詳細について確認決

定し、作業スケジュール表を作成し、官側の承認を得て作業を実施する。

作業スケジュール表は、作業項目ごとに作業人員、想定処理量、作業期間等を記述した週単位のスケジュールを記載し、速やかに官側に提出すること。

2. 5 権 利

本作業で作成されたテキストデータおよびテキストデータ付き PDF の著作権は官側に帰属する。当該データは官側が許諾した場合のみ、AI による学習等に使わせることができるものとする。

2. 6 秘密保持

本作業を実施するにあたり知り得た情報及び作成したデータについては、第三者に洩してはならない。

2. 7 成果物

契約相手方は以下に示す成果物を官側に納品するものとする。格納媒体は契約相手方で準備するものとする。細部については、官側との協議に基づき、実施するものとする。

- (1) 戦史叢書テキストデータ : データ格納媒体 [DVD-R]
- (2) テキストデータ付き PDF : データ格納媒体 [DVD-R]
- (3) 描画結果データ : データ格納媒体 [DVD-R]

2. 8 品質保証

官側へ納品した成果物に誤認識等の不具合が発見された場合は、納入後 5 年間、契約相手方の責任において適宜速やかに修正等の対応を行うものとする。あるいは事後、誤認識箇所を官側にて適宜修正できる措置を用意すること。

3 提出物及び納期等

- (1) 提出物及び納期は表 1 のとおり。

表 1

名 称	数 量	引渡時期	引渡場所
戦史叢書テキストデータ	1 式	令和 8 年 3 月 27 日	防衛省 防衛研究所
テキストデータ付き PDF	1 式		
描画結果データ	1 式		
テキストデータの修正要領を記したファイル	1 式		

- (2) その他提出書類

提出書類については表 2 のとおりとする。

表 2

名 称	部数	提出時期	様式等	提出場所	備 考
実施体制表	1 部	契 約 後 速 やかに	様 式 随 意 及 び 写 し 様式随意	防衛省防衛研究所 東京都新宿区市谷 本村町 5 - 1	本役務に携わる作業 の体制表
作業計画表	1 部				本役務の作業予定
主任技術者届	1 部				経歴やスキル

4 入札資格要件

入札参加希望者の資格要件は以下のとおりとし、要件を満たすことを証明する資料を官側の指定する日までに書面で提出し、承認を得るものとする。

(1) 実績

省庁及び関連機関において、当該機関が保有する文書等について AI-OCR の開発改善業務を、過去 3 年間で 5 回以上行った実績を有すること。上記の実績がありながらも件数が不足する場合は、大学等の研究機関や民間企業向けの縦書き段組み文書（一部横書き混在）用 AI-OCR の開発改善業務の具体的内容を明示することで実績としてみなすものとする。それらの実績についてリストを作成し、官側の承認を受けるものとする。

(2) サンプルデータの処理

官側が提供するサンプル 2 点（本文見開き 4 ページ）に付き、本業務利用の OCR を用いてサンプル処理と評価を行い、以下の内容を提出すること。

- ・評価結果報告と校正の方法
- ・テキストデータ付き PDF
- ・テキストデータ（マークダウンや JSON などの AI 向けフォーマット）
- ・描画結果ファイル

(3) セキュリティ

役務作業を行う場所について、入退出管理及び 24 時間警備体制が確立され、使用パソコンについては、社外への流出防止処置が施されているものとする。

(4) 作業従事者

主任技術者は、上記実務において、責任者又は副責任者等を務めた実績を有する者であること。

5 検 査

第 3 項に関する提出物に関して、対象データファイルをランダムで抽出、開き、目視検査を行う。

6 その他

この仕様書に疑義が生じた場合は、速やかに官側と協議するものとする。