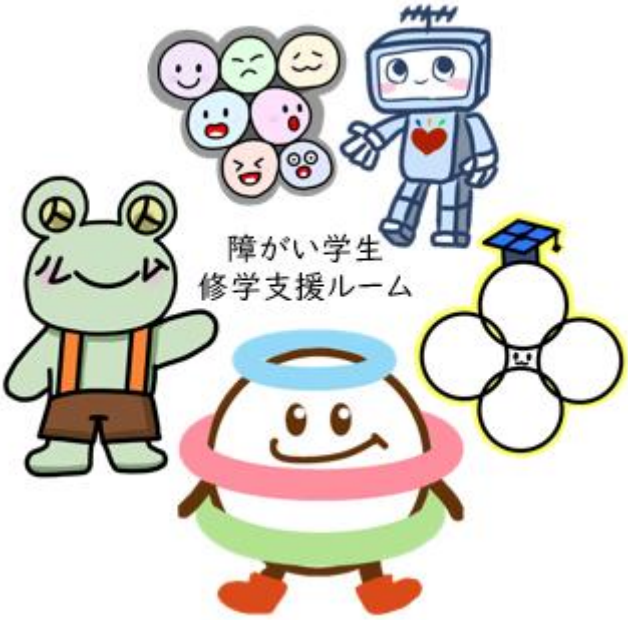
テキストデータ化

マニュアル

Ver.3.0



2023年10月16日 小馬 作成

2024年11月20日 柏原 改訂

もくじ

はじめに 2

テキストデータ化とは 2

作業工程 2

作業内容について 4

準備:PDF から 4

準備:Word/Excel/PowerPoint から 5

校正:文字の校正 6

校正:表の校正 9

校正:図/画像の校正 10

確認:ページ数の表記 11

確認:最終確認 11

テキストデータ化例 13

時間割表 13

目次と段組み 14

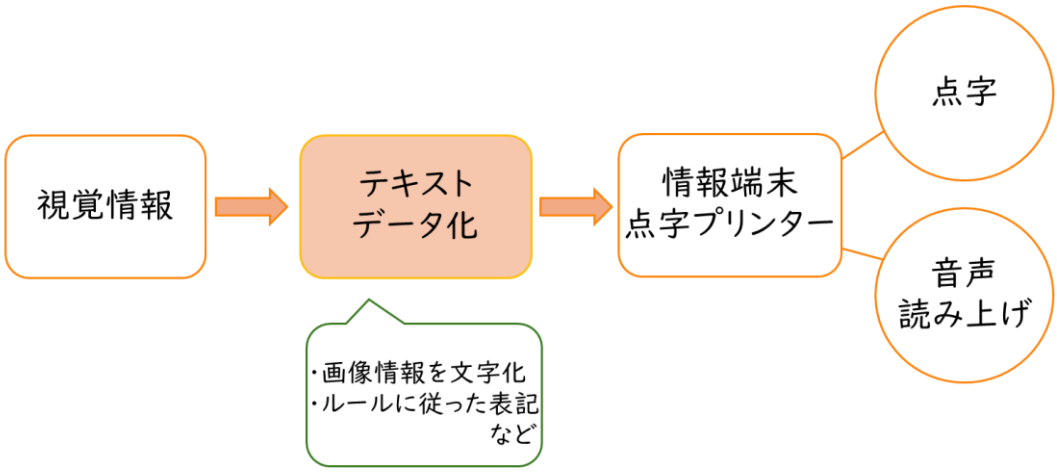
ポスター資料 16

はじめに

テキストデータ化とは

視覚障がい学生が墨字（印刷された、または書かれた文字）や画像から情報を獲得する際、音声での読み上げや点訳（点字に翻訳すること）が必要になる場合があります。現在はデータを読み込むことで、音声や点字に変換できる携帯情報端末や点字を紙に打ち出す点字プリンターも活用されています。端末やプリンターに読み込まれたデータが正しい音声や点字に変換されるために、データ上の表記におけるいくつかのルールがあります。

授業やガイダンスなどで使用する資料を、そのルールに従ってテキストデータファイルに変換することを「テキストデータ化」と言います。



作業工程

テキストデータ化は「見てわかりやすい」ではなく、「**読み上げてわかりやすい**」ものにすることが目的です。実際に声に出して読んでみたり、どのように読まれるか想像したりしながら作業しましょう。

このテキストデータをもとに、視覚障がい学生は授業を受講したり、課題を提出したりします。

校正にあたる際の大原則は「**原本に忠実に**」です。表の表記や挿入されている画像などが漏れなくテキストデータになっているか、ズレや誤字がないか等に注意しながら進めてください。

作業途中で判断に迷うことや、分からないことがあれば遠慮なく支援ルームに相談してください。また、このマニュアルの手順や画像は作成当時のものです。それぞれの環境では別の表示になったり、よりよい方法が見つかったりすることがあります。適宜修正や追記を行うのでぜひ、報告・共有をしてください！よろしくお願いします。

作業工程

準備

- PDFから
- Word/Excel/PowerPointから

校正

- 文章の校正
- 表の校正
- 図/画像の校正

確認

- ページ数の表記
- 最終確認

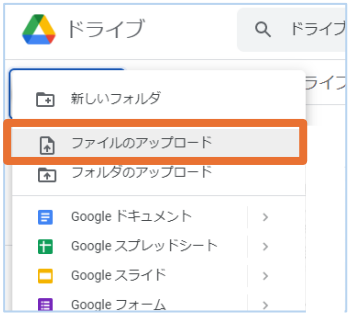
※支援ルームに登録しているサポート学生は、作業後報告をする必要があります。詳しい手続きは別資料「報告の流れ」をご確認ください。

作業内容について

準備:PDF から

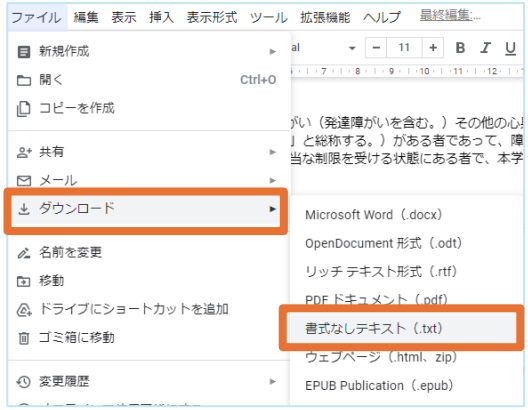
PDF 等の画像上の文字をデジタルテキストにする技術が OCR です。今回は Google ドライブの OCR 機能を使った方法を紹介します。

① PDF データを Google ドライブにアップロードする。



② アップロードした PDF を開き、上部にある「Google ドキュメントで開く」からテキスト化する。

③ 開いた Google ドキュメントの「ファイル」>「ダウンロード」>「書式なしテキスト(.txt)」で保存する。



〈注意〉

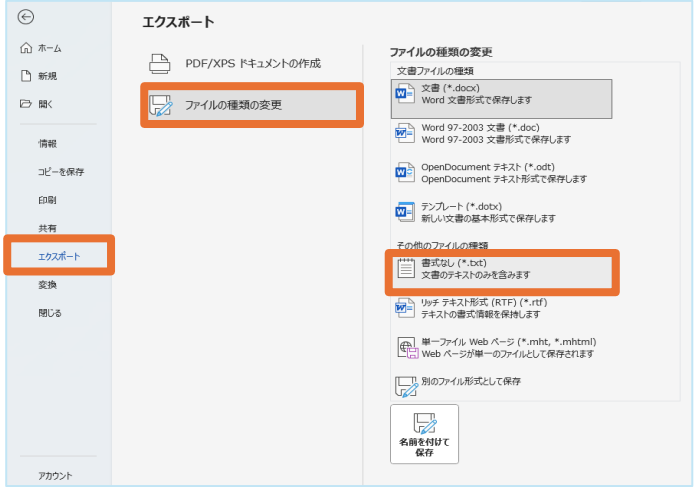
・すべての文字が完璧にテキスト化されるとは限りません。必ず「校正」「確認」の工程を行ってください。

・元データによっては、OCR で読み込めない場合があります。その際はキーボード入力で打ち込んでいく必要があります。

準備:Word/Excel/PowerPoint から

① 元データを開く。

② 「ファイル」>「エクスポート」
>「ファイルの種類の変更」
>「書式なし(.txt)」
でテキストファイルとして保存する。



図表を PDF で貼っていたり、テキストボックスを多用したりしているデータは一度 PDF にした後、OCR でテキストデータ化する方がよりスムーズに作業を行えます。その場合、コピーしたファイルで PDF 化を行い、**元データは必ずそのままのファイル形式**で残してください。

「準備」工程 完了!

校正:文字の校正

OCR での読み取りでは十分に認識されない、元データそのままの表記では情報が正しく伝わらないといったケースがあります。
テキストデータ化の最も重要な作業工程です。
なお、元データにない文章の挿入は、注釈と分かるように「**（校注：　　）と記入**」してください。

① 誤認識の訂正

OCR で処理した文字は似ている形の文字を混同してしまうことが多々あります。特に、手書き文字や荒いスキャンの元データは読み取りの精度が落ちてしまいます。
OCR 処理をした後は、元データと見比べて（読み比べて）訂正を行います。
例）元データ「タ方（ゆうがた）」→OCR 処理後「タ方（たほう）」
漢字ではなく、カタカナとして認識されている。

誤認識されやすい文字（追加しますので、見つけ次第報告してください！）

び・ぴ	人・入	待・侍
る・ろ	日・目	東・束
l（える）・l（いち）	O（おー）・0（ぜろ）	伸・仲

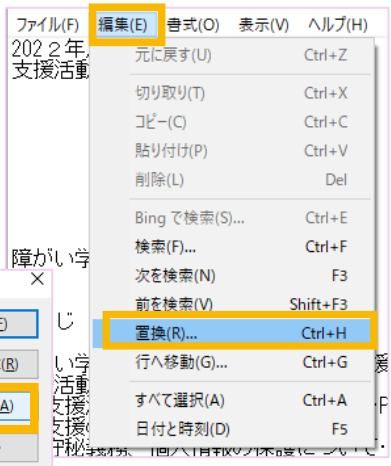
など

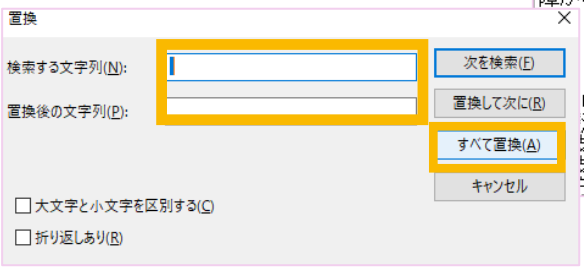
② 意図しない空白や改行の削除

OCR で処理を行った場合、意図しない場所にスペースが入ることがあります。また、元データに体裁を整える目的や改行時の字下げにより、スペースが入っていることがあります。
このようなスペースがあると、正しく点訳・読み上げがされないため、不要なスペースは消去してください。

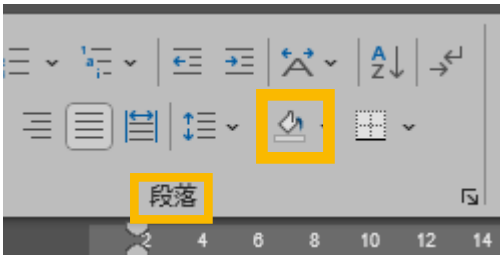
元データ	読み上げ（誤読例）	修正後	修正後読み上げ
日　時 （スペースあり）	ひ　とき	日時	にちじ

その際には、1 つずつ消していくよりも、「編集>置換」を選択し、「検索する文字列」にスペース、「置換後の文字列」には
なにも入力せずに「すべて置換」を行うと、空白をすべて消すことができます。※ただし、必要なスペースが消えてしまうことが
あるので注意してください。





スペースが使われている場所は、Word の機能で確認ができます。
「ホーム」>「段落」>
「編集記号の表示/非表示（右上のマーク）」をオンにすると、スペースが□と表示されます。この確認をする場合、必ず元データをコピーしたファイルで行い、**元データはそのままの状態**で残してください。



③ 半角・全角の統一

英数字は全角・半角の表記がバラバラになってしまうことが多いです。点訳・読み上げの都合上、大きな意図がない場合、英数字は**すべて半角で統一**してください。
その際には、②で説明している「置換」を活用すると便利です。

④ 記号や省略記号の校正

記号は利用する機器のソフトによって、読み取りが異なります。また元号をアルファベット1文字で表すなど、省略された表記のままだと意図しない伝わり方になることがあるので修正を行います。

元データ	読み上げ（誤読例）	修正後	修正後読み上げ
※1	きごういち	注1	ちゅういち
¥5,000	ごぜろぜろぜろぜろ	5,000円	ごせんえん
4/1	いちぶんのよん	4月1日	しがつついたち
14:30	Fourteen thirty	14時30分	じゅうよじさんじゅっぷん
3 日（月）	みっかつき	3日月曜日	みっかげつようび
R6	あーるろく	令和6年	れいわろくねん

⑤ 読みの補記

人名などの固有名詞や英語表記は、正しく読み上げられないことが多々あります。

学生が使用する読み上げソフトにもよるため、すべてを補完することは難しいですが、誤読するだろうと分かるものについては**（校注:　　）**で読みを補記するようにしてください。

元データ	読み上げ（誤読例）	修正後
大教大	おおきょうだい	大教大（校注:だいきょうだい）
DAISY 図書	でぃーえーあいえすわいとしょ	DAISY（校注:でぃじー）図書

※繰り返し登場するものについては、最初に登場するときのみ〇〇（校注:___）とし、その後は読みのみの表記でもかまいません。

例）大教大

１回目「大教大（校注:だいきょうだい）～」２回目以降「だいきょうだい～」

校正:表の校正

表は読み取り方が一定でなく、ソフトではうまく読み上げられないことが多いもののひとつです。
表のはじめと終わりに**(校注:表はじめ)**、**(校注:表終わり)**と記入し、各表の読み方に応じて表記を工夫します。
なるべく、1行に1項目になるように表記をしていきます。

例)

買い物メモ			
品名	個数	値段	買った
りんご	2個	¥150	✓
卵	6個入り	¥100	
	8個入り	¥130	✓
牛乳	1L	¥250	✓

このまま読み上げると……(漢字で表記しています)

買い物メモ 品名 個数 値段 買った りんご にこ ひゃくごじゅう 卵 ろっこいり ひゃく はっこいり ひゃくさんじゅう 牛乳 one litre にひゃくごじゅう

のように、表の左上から右に読み上げられてしまうため、
テキストデータ化では以下のように表記すると分かりやすくなります。

買い物メモ

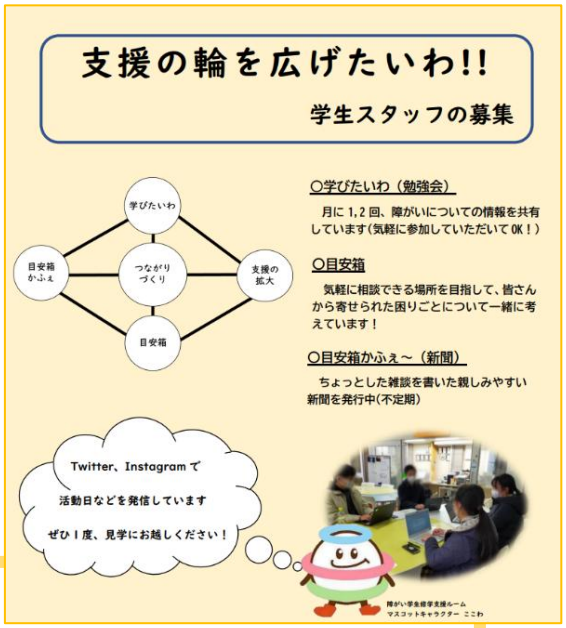
(校注:表はじめ)
品名 りんご 個数 2個 値段 150円 買った
品名 卵 個数 6個入り 値段 100円 買っていない
品名 卵 個数 8個入り 値段 130円 買った
品名 牛乳 個数 1リットル 値段 250円 買った
(校注:表終わり)

校正:図/画像の校正

特に、PowerPoint の資料やポスターなどは、図や画像が多く用いられています。
表記を工夫したり、代替テキストを入れたりすることで対応します。
（校注:図）、（校注:写真）、（校注:イラスト）を
適宜使用し、どこに配置されているかも表せると
より伝わりやすい情報となります。

※別資料「非テキストコンテンツ編」では、
さまざまな図や画像のテキストデータ化を紹介しています。
ぜひご確認ください。

例）学生が作成した修学支援ルームのポスター（→）



テキストデータの一例）

（ポスタータイトル）
支援の輪を広げたいわ!!学生スタッフの募集

（校注:左上につながりづくりを中心に、学びたいわ、目安箱カフェ、支援の拡大、目安箱がつながった図）

（校注:図の右側に説明）
学びたいわ（勉強会） 月に1、2回障がいについての情報を共有しています。（気軽に参加していただいて OK!）
目安箱 気軽に相談できる場所を目指して、皆さんから寄せられた困りごとについて一緒に考えています!
目安箱かふぇ～（新聞） ちょっとした雑談を書いた親しみやすい新聞を発行中（不定期）
（校注:交流スペースで学生が集う様子の写真）

（校注:写真の横で障がい学生修学支援ルームマスコットキャラクターのここわのイラストと、そこから吹き出し）Twitter、Instagram で活動日などを発信しています。ぜひ一度見学にお越しください!

「校正」工程 完了！

確認:ページ数の表記

元データを参考にテキストデータに追加します。元データの形式によって表記を分けています。
紙資料の場合、〇枚目
書籍の場合、〇ページ
スライド資料の場合、〇スライド
複数ページにまたがる場合、〇～□ページ
〈注意〉
・どの表記であっても**数字から**表記を始めてください。
・内容のキリが悪くても、**ページが変わったらすぐに**追加してください。
　ただし、表などが複数ページにまたがる場合にはまとめて表記しても OK です。
例）

> ～～障がいのある学生は、障がいの種類やその程度により個
> （2 ページ）
> 人差が大きく、一人ひとりに様々なニーズがあります。このため～～

その他、大きくレイアウトが変わった際なども、校注を用いて表記できると伝えられる情報が増えます。

確認:最終確認

□ 誤植・脱字を確認した。
□ 意図しない改行や空白がない。
□ 数字やアルファベットがすべて半角表記になっている。
□ 省略・記号・外国語・固有名詞が正しく表記されている。
□ 図表、画像の注釈や校正を行い、その位置がずれていない。
□ ページ数やスライド番号が正しく表記されている。
□ テキストデータ形式（.txt）で保存されている。

「確認」工程　完了！

作業は完了となります！
お疲れ様でした！
サポート学生は「報告の流れ」に従って、**データの提出や謝金に関する報告**を行ってください。

【参考】

総務省　東海総合通信局

https://www.soumu.go.jp/soutsu/tokai/siensaku/accessibility/L4_text2.html

テキストデータ化例

時間割表

令和4年度 授業時間割

	第1時限 (8:50～10:20)				第2時限 (10:35～12:05)			
	授業コード	授業科目名等	教員名	講義室	授業コード	授業科目名等	教員名	講義室
月								
火								
水					C02014	病弱者の心理・生理・病理 2 (後)	宇塚 大内田	C6-203
					C02006	視覚障害教育 2 (前)	正井	A-215
木	C02016	発達障害指導論 2 (前)	湯浅	C6-203	C02018	聴覚障害教育 2 (前)	井坂	C6-203
					C02001	特別支援教育総論 2 (後)	冨永光	C6-203
金								

	第3時限 (12:55～14:25)				第4時限 (14:40～16:10)			
	授業コード	授業科目名等	教員名	講義室	授業コード	授業科目名等	教員名	講義室
月					C02005	病弱教育 2 (前)	冨永光	A-307
火								
水								
木	C02007	発達障害教育 2 (前)	後藤	A-216	C02002	聴覚障害心理 2 (前)	西山	C6-203
					C02015	障害者の心理・生理・病理 2 (後)	[illegible]	C6-203
金					C02011	障害児研究法B 2 (後)	全員	C7-101

	第5時限 (16:25～17:55)				
	授業コード	授業科目名等	教員名	講義室	
木	C02020	知的障害教育 2 (前)	今林	C6-203	

(備考)

〈Point〉
「ない」ということも大切な情報です。
空きコマを作業者がテキスト化しなければ、
「データが抜け落ちている」のか、「授業がない」
のか、利用学生が判断できません。
「授業がない」ことをテキスト化しましょう。

令和４年度 授業時間割

（校注:表はじめ）

前期

月曜日

第１時限（8 時 50 分から 10 時 20 分）:授業なし

第２時限（10 時 35 分から 12 時 5 分）:授業なし

第３時限（12 時 55 分から 14 時 25 分）:授業なし

第４時限（14 時 40 分から 16 時 10 分）

授業コード:C02005 講義名:病弱教育 単位数:2 教員名:冨永光 講義室:A-307

火曜日

第１時限（8 時 50 分から 10 時 20 分）:授業なし

第２時限（10 時 35 分から 12 時 5 分）:授業なし

第３時限（12 時 55 分から 14 時 25 分）:授業なし

第４時限（14 時 40 分から 16 時 10 分）:授業なし

水曜日

、、、（省略）

目次と段組み

学生生活案内目次

	Page
はじめに	1
お知らせ	2
注意喚起	5
学内規則の遵守	
学生相談	7
01. 学生生活に関する窓口	9
02. 諸手続	11
03. 証明書の発行	
04. 学生証	13
05. 授業料	14
06. 授業料免除等	
07. 奨学金	15
08. 課外活動・その他施設	17
09. 附属図書館	19
10. 就職支援	20
11. 教育研究施設	21
12. 学生宿舎・下宿	22
13. 学生教育研究災害傷害保険(学研災)・学生教育研究賠償責任保険(学研賠)	23
14. アルバイト	
15. 海外留学・国際交流	25
16. ボランティア活動	26
17. 大阪教育大学校友会	
18. 大阪教育大学教育振興会	27
19. 国民年金学生納付特例制度	28
20. 懲戒の標準	
国立大学法人大阪教育大学の基本理念と基本目標	29
沿 革	30
大阪教育大学学生規則	31
大阪教育大学学生懲戒規則	34
防災カード	巻末
インフォメーション	
緊急時の大学連絡先等	
学外の相談窓口	
学生生活に関すること	

はじめに

大学から学生への連絡等について

●大学が学生に対して行う通知及び連絡は，Live Campus（LCU）・大教 Gmail（Google Workspace）・Moodle 及び ANPIC（災害情報）を使用します。
●授業に関する連絡は，Live Campus（LCU）により行います。
●構内放送による学生呼び出しは行いません。
●郵便物や荷物は大学で受け取ることはできません。学生個人宛の郵便物等を，大学の住所に送らないでください。

掲示板の場所について

●柏原キャンパス：
・M 棟（大学会館）：
〈2F〉奨学金・授業料免除・就職・学研災保険・教育振興会等
〈1F〉課外活動・ボランティア・同窓会等
・N 棟（事務局棟）：〈3F教育実践支援ルーム前〉教育実習・介護等体験・学校インターンシップ・学校サポート活動
●天王寺キャンパス：中央館1F

車両による構内への入構

●柏原キャンパス
車・バイク等で入構するには「許可」が必要です。許可された場合は「車両入構許可証」が交付されます。
学生支援課学生企画係で手続きをしてください。
●天王寺キャンパス
車・バイクでの入構は原則禁止です。
自転車で入構する場合は正門守衛室で「車両入構許可証」交付手続きをしてください。

正しい交通ルール・マナーを守り安全運転を心がけましょう。

大阪教育大学安否確認システム（ANPIC）について

大阪教育大学安否確認システム（ANPIC）は，大地震を始めとする大規模災害が発生した際に，本学の学生，教職員へ迅速に災害情報を発信し，受信者の安否状況の報告により安否確認を行うものです。
大阪府及び隣接する府県で震度5強以上の地震発生を気象庁が発表した際，安否確認のお知らせが LINE 等（事前に登録が必要）に自動で配信されますので，安否報告を必ず行ってください。
QR コードを読み取れない場合は，以下の URL を入力してください。

https://anpic-v6.jecc.jp/osaka-kyoiku/login/

本人または父母等の，氏名・本籍・住所・電話番号を変更した時の届出について

学生支援課奨学厚生係（天王寺キャンパス所属の学生は天王寺地区総務課）に届け出てください。手続には学生証が必要です。
氏名の変更及び本籍の変更には，変更の事実を確認できる書類を持参ください。（※生協に届けただけでは大学の登録は変更されません。変更の際は「大学」と「生協」両方に届けてください。）

1

（1 ページ）

（左側）

〈Point〉
段組みが使われている資料の場合、
左側の段→右側の段の順でテキスト化します。
また、段組みがある場合は OCR で正確に
読み込まれないことが多いです。

学生生活案内目次

はじめに 1ページ
お知らせ 2 ページ
注意喚起 5 ページ
学内規則の遵守 5 ページ
学生相談 7 ページ
01. 学生生活に関する窓口 9 ページ
02. 諸手続 11 ページ
03. 証明書の発行 11 ページ
04. 学生証 13 ページ
05. 授業料 14 ページ
06. 授業料免除等 14 ページ
07. 奨学金 15 ページ
08. 課外活動・その他施設 17 ページ
09. 附属図書館 19 ページ
10. 就職支援 20 ページ
11. 教育研究施設 21 ページ
12. 学生宿舎・下宿 22 ページ
13. 学生教育研究災害傷害保険（学研災）・ 学生教育研究賠償責任保険（学研賠） 23 ページ
14. アルバイト 23 ページ
15. 海外留学・国際交流 25 ページ
16. ボランティア活動 26 ページ
17. 大阪教育大学校友会 26 ページ
18. 大阪教育大学教育振興会 27 ページ
19. 国民年金学生納付特例制度 28 ページ
20. 懲戒の標準 28 ページ

〈Point〉
目次がある資料では、元データのページ数に
対応するように、テキストデータでも目次を
表記します。
資料のページ数に表紙等がカウントされておらず、
途中からページ数表記が始まる場合は、
・表紙・表紙裏・次ページ
等を用いて、ページが変わったことを
表記してください。

（右側）

はじめに

大学から学生への連絡等について

ポスター資料

〈Point〉
キャラクターやイラストなど、分かる範囲で（校注：）を用いて補足を入れます。
ただし、主観が入りすぎないよう、注意が必要です。

〈Point〉
QR コードは URL に変換して表記します。
方法が分からない場合は、報告時にお伝えください。

（タイトル）
サポート学生募集！
（校注：タイトルの右側にロボットの支援ルームキャラクター：むーむのイラストがありその上に支援ルームという掛け声）
（校注：ロボットのイラストの左側に説明）
4 月 25 日、5 月 12 日 12 時 20 分から柏原 C（キャンパス）支援ルームにて登録説明会開催！
「すべての学生が、共に気持ちよく学び合う活動」に参加しませんか？

PC テイク、字幕挿入、音声データ化、移動サポート、手話、学生活動
（校注：それぞれが丸で囲まれていてその中の字幕挿入のところに手を添えたカエルの支援ルームキャラクター：かえるーむのイラスト）

本学では、すべての学生が同じように教育を受けることができるよう取り組んでおり、障がいのある学生への修学支援活動を行う学生スタッフを募集しています。（時給 1,100 円）サポーター登録の後、それぞれの研修を受けることで支援活動に参加できます。学生活動や勉強会のみの参加も大歓迎！登録や活動については遠慮なく支援ルームにお問い合わせください。

登録は登録フォームまたは直接支援ルームへ
https://forms.gle/xxbziHGwVgXog8ZW8
（校注：この文章の右側に QR コードがありその右側に学生帽をかぶった支援ルームキャラクター：フォーレのイラスト）