

文書モデルの作成と文書診断に向けた予備的考察： 自動車の修理書を対象に

杉野 峰大[†]宮田 玲[‡]松崎 拓也[‡]佐藤 理史[‡]
[†]名古屋大学工学部電気電子・情報工学科 [‡]名古屋大学大学院工学研究科

1 はじめに

文書作成において、何をどのような配置で書くかは執筆者の裁量に任せがちであり、効果的な情報提示を可能とする文書の内容と構造は明文化されていないことが多い。我々は、内容要素（何を書くか）と文書構造（どのように内容要素を配置するか）を合わせて文書モデルと呼び、自動車関連の技術文書用の文書モデルの作成を目指している。文書モデルは、文書の執筆及び診断の指針として使える有用なものであるが、どのように作成すればよいかは自明ではない。

技術文書の構造を一般的に定義したものとして DITA (Darwin Information Typing Architecture)[1] が広く知られているが、特定の文書ドメインにそのまま適用できるとは限らず、また書くべき内容についても細かく規定されていない。そこで本研究では、比較的構造が明確な自動車の修理手順書（以下、修理書）を調査対象として、技術文書用の規格として実績のある DITA をベースとしながら、それを拡張する形で文書モデルの作成を試みた。具体的には、自治体文書を対象とした [2] で提示されている方法を参考に、実際の文書に書かれている内容要素をコーディングにより類型化し（2 節）、修理書向けに調整した DITA の構造に内容要素を対応づけた（3 節）。さらに文書モデルの応用の 1 つとして、文書診断に用いる可能性を考察し、テキストに内容要素を自動で付与する簡易な機械学習モデルの作成と評価を行った（4 節）。

2 内容要素の調査

2.1 調査方法

自動車の修理書には、これまでどのような内容が書かれてきたかをオープン・コーディングの手法 [3] を参考に調査した。

調査対象として 1 つの車種の修理書¹（合計 2614 文書）の中で、最も出現頻度の大きい「取り外し」というタイトルを持つ 325 文書のうち 31 文書をランダムに選び、それらの対になっている「取り付け」と合わせた 62 文書を選定した。

¹トヨタ自動車株式会社, PRIUS 修理書, 2017 年 11 月版。

使用用品	操作
クリップリムーバーを使用して、	クリップを取りはずす。

図 1: 内容要素のコーディングの例

図 1 のように、最小単位を節として、対象テキスト全体を重複なく網羅するように、文書に書かれている内容を抽象化したコードを手作業で付与した。コードはあらかじめ用意したものではなく、文書を順次分析しながら作成した。また並行してコードの類型化も行った。

2.2 調査結果

対象 62 文書に対してコードの付与と全体の検証を 3 サイクル行った結果、26 種類の内容要素コードにまとめ、合計 2907 個のテキストスパンに対してコードが割り当てられた。表 1 に内容要素のタイプを出現頻度、例とともに示す。

修理書は、自動車の修理手順を段階を追って説明するものであるため、具体的な動作に関する要素が多く記述される。動作は大きく、具体物の状態を変化させる操作、具体物の状態を観察・評価する判断、具体物に関連して行う行為の 3 つに大別した。

また動作に類似した要素として、方法がある。方法は動作を補足するもので、「～ように～する」と具体的なやり方を指定するものである。

3 修理書の文書モデルの作成

DITA を基にして自動車の修理書に適した文書構造の定義を行い、2 節で得られた内容要素との対応づけを行った。

3.1 修理書の文書構造

DITA とは技術文書の構造を定義した XML 標準規格である。DITA には Concept, General Task, Strict Task, Machinery Task, Reference などの構造が定義されており、自動車の修理書の構造は Machinery Task 型と類似している（表 2 参考）。しかし、この構造は必ずしも自動車の修理書に最適化されているわけではないため、一部を変更し、自動車の修理書に合わせた文書構造を定義した（表 3 左部分）。

内容要素		頻度	例		
メタ	手順見出し	764	1. アウトミラー RH 取り外し		
	作業対象	45	取り付け位置 RH 側:		
	指示的概要	14	以下に示す整備要領は LH 側を示す。		
	同一作業	110	LH 側および RH 側を同要領で整備する。		
	誘導	17	作業後の必要作業 リヤアクスルキャリア LH 脱着作業時に行う部品の脱着・交換作業後に必要な調整・初期化・登録作業を以下に記載		
	注意事項の確認	10	作業する前に、必ず SRS エアバッグシステムの注意事項を確認すること。		
モノ	パーツ	機能	19	補機バッテリーターミナル脱着時には、一時的に機能しなくなるが、各システムを使い始めることで学習が完了する機能がある。	
		物理	2	エアバッグコンピュータ ASSY にヨーレートセンサ & G センサが組み込まれている。	
作業	実施条件	条件	イベント	57	クランクポジションセンサを落下させたり、強い衝撃を与えた場合は、
			アイテム	107	デジタルTVアンテナあり:
		時期	開始	11	シールパッキングラック塗布後、
			終了	2	接続作業の直前まで、
	動作	操作		770	ブレーキチューブクランプをブレーキアクチュエータブラケット ASSY に取り付ける。
		判断		40	補機バッテリーマイナスターミナルが切り離されていることを確認する。
		行為		8	ブレーキチューブ再接続のためのメモをする。
	説明	方法		68	クリップはイラストの範囲になるように取り付ける。
		禁止事項	内容	101	エンジンおよびエキゾーストパイプなどの高温部に触れない。
			危険性	38	鏡面部が割れるおそれがあるため、
		徹底事項		16	必ず修理書記載手順に従って作業を行う。
		推奨事項		2	図のようにコイルスプリングストッパーベルトの取り付けを推奨。
	理由			40	ナットは回り止め処理されているため、
	目的			10	スパイラルケーブル断線防止のため、
	補足	使用用品		120	ロープなどを使用して、
		基準		59	トルク:45 N・m (459 kgf・cm)
		参考情報		2	リヤディスク交換時は不要。
参照				475	要領は総説>総説>作業上の心得および注意事項>注意事項 参照

表 1: 修理書の内容要素

DITA 要素		説明
prelreq		事前条件
context		背景情報
steps	step cmd	手順
	info	補足情報
	tutorialinfo	チュートリアル
	stepresult	step の結果
result		Task の結果
closereq		終了条件

表 2: DITA の Machinery Task 型の構造 (一部省略)

大きな変更点として、**attention** と **conditional-step** という 2 つの要素を追加した。**attention** は、注意や警告についての情報を扱う。自動車の修理において注意や警告は人命に関わるものが多いので他の補足的な内容とは別に扱うべきだと考えた。**conditional-**

step は、条件付きの作業を指すものである。自動車の修理書では、パーツの有無、駆動方式の違い、さらにパーツの破損といったイベントにより実施すべき作業が変わることが多いので、条件の有無によって **step** と **conditionalstep** を使い分けることにした。**conditionalstep** を用いることで、作業の分岐を構造レベルで明示的に保持できるため、あらかじめ選択された条件に適する手順のみを読み手に提示したり、箇条書き記号に連番ではなく「1-a」「1-b) を使ったりと、表示上の工夫を行うことが可能となる。

なお、**steps** の各 **step** に入っている作業 (**cmd**) がより細分化できるときは **steps** の入れ子構造をとることで表現する。

DITA に基づく手続き型の文書構造				対応する内容要素
attention				禁止事項, 徹底事項, 推奨事項, 注意事項の確認
info				モノ, 補足
steps	step	cmd		動作, 使用用品
	conditionalstep	case	condition	実施条件
			cmd	動作, 使用用品
			attention	禁止事項, 徹底事項, 推奨事項
			tutorialinfo	方法
			info	モノ, 補足
postreq				作業後の必要作業

表 3: 文書構造と内容要素の対応

3.2 内容要素と文書構造の対応づけ

定義した文書構造に 2 節で得られた各内容要素を対応づけた (表 3)。これにより、どのような内容をどのような配置 (構造) で書けばよいかの指針となる文書モデルの大枠が完成した。メタレベルの内容要素である手順見出し、作業対象、指示的概要、同一作業は、DITA の外側で定義すべきものと考え、DITA 要素とは対応づけなかった。また理由、目的は、どの DITA 要素にも出現する汎用的な要素であるため、表では省略している。

4 文書診断への応用

文書モデルの活用方法の 1 つとして、文書の診断が考えられる。たとえば、表 4 のように禁止事項が並んでいるところに 1 つだけ作業の方法が含まれていることを可視化することで、文書の改善を促すことができる。具体的には、自動的な文書診断を行うためには、(1) テキストへの内容要素・DITA 要素の付与、(2) 要素の配置が適切であるかのチェックが必要である。本稿では、予備的な検討として (1) に着目し、テキストに対する内容要素の自動付与実験について報告する。

4.1 内容要素の自動付与

内容要素ごとに、入力文がその要素を含んでいるか否かを判定する 2 値分類器を作成した。表 1 の内容要素の内、基準は数値のみの断片で記載されることが多いため今回は対象から外した。

訓練データの形態素 (表層形) の Bag of Words を素性として SVM で分類器を実装した²。SVM の実装は scikit-learn³ で提供される SVC (RBF カーネル) を用いた。

4.2 実験データ

分類器の構築には、人手で内容要素を付与した「取り付け」「取り外し」の 62 文書を用いた (2 節参照)。

²形態素解析には MeCab ver.0.996 (<http://taku910.github.io/mecab/>) と mecab-ipa-NEologd (<https://github.com/neologd/mecab-ipadic-neologd>) を用いた。

³<https://scikit-learn.org/stable/>

ただし、調査時には最小単位を節として内容要素を割り当てていたが、今回の分類器では入力を文とするため、ある内容要素が割り当てられている節を持つ文は、文としても内容要素を持つものとした。データの合計は 2359 文である (Base データとする)。

また、他の手続き型文書への適用可能性の検証のため、修理書中の「点検」というタイトルの文書を 10 文書無作為に選び、2 節と同様の方法で内容要素のコーディングを行った (合計 158 文、New データとする)。コーディングに用いた内容要素は表 1 に列挙したもので十分であり、内容要素の追加や変更は行っていない。

4.3 実験結果

表 5 に自動付与 (分類) 実験の結果を示す。なお Base データでは 5 分割交差検定を行った。全体としては、Base データにおける精度・再現率は高い値に達している。

操作や判断の要素は Base データでは高い精度で分類できているが、New データでは大きく精度が下がっている。New データ中には Base データでは 1 度も使われていなかった「点検」という単語が頻出しており、今回用いた Bag of Words を素性とした簡易な機械学習手法では、このような未知語への対応が十分でできなかったといえる。

また、Base データにおいても比較的、目的の内容要素の分類精度が低い。これは目的は理由と同様に「～のため」という形が用いられているため、件数の多い理由に引きずられて精度が低くなったと考えられる。

5 おわりに

本研究では、自動車の修理書を対象に、何がどのように配置されているか、すなわち内容要素と文書構造を定めることで、文書モデルを作成した。修理書の中でも、「取り付け」「取り外し」の文書の分析により作成した文書モデルであるが、手続き型の修理書に広く

テキスト	内容要素	文書構造	
1. ブレーキースタポン ASSY 取り付け	手順見出し	step	cmd
a. ブレーキースタポン ASSY、ブレーキースタポンプシュおよび、ブレーキアクチュエータケースカラーをブレーキアクチュエータブラケット ASSY に取り付けける。	操作	step	cmd
注意： ・運搬時、ブレーキースタポン ASSY は落下させない。落下させたブレーキースタポン ASSY は使用しない。 ・コネクターの内側にブレーキフルードが付着しないように注意する。 ・ブレーキースタポン ASSY 取り付け時、ブレーキースタポンカラー (*1) およびブレーキースタポンプシュ (*2) がブレーキースタポン ASSY に確実に取り付けられていることを確認する。 ・新品のブレーキースタポン ASSY にはブレーキフルードが充てんしてあるため、ブレーキチューブを接続するまで、ホールプラグをはずさない。 ・ブレーキースタポン ASSY を取り扱う場合、イラスト内指示文字 (*a、*b、*c および*d) で示す部分を持たない。	禁止事項/内容		attention
	禁止事項/内容		attention
	方法		tutorialinfo
	禁止事項/内容		attention
	禁止事項/内容		attention
b. ナットを取り付ける。 トルク：5.4N・m(55kgf・cm)	操作 基準	step	cmd info

表 4: 文書診断のための内容要素と文書構造の可視化の例

内容要素	Base: 「取り付け」「取り外し」				New: 「点検」				正答率
	総数	P	R	F	総数	P	R	F	
メタ 手順見出し	764	0.999	0.999	0.999	38	0.974	1	0.987	0.994
作業対象	45	1	0.978	0.988	0	N/A	N/A	N/A	1
指示的概要	14	1	1	1	2	1	1	1	1
同一作業	114	1	1	1	3	1	1	1	1
誘導 注意事項の確認	10	0.950	1	0.967	0	N/A	N/A	N/A	1
作業後の必要作業	15	1	0.950	0.971	0	N/A	N/A	N/A	1
パーツ 機能	19	1	1	1	0	N/A	N/A	N/A	1
物理	2	0.800	0.800	0.800	0	N/A	N/A	N/A	1
作業 条件 実施	51	0.965	0.921	0.940	37	1	0.892	0.943	0.975
イベント アイテム	113	1	0.995	0.998	2	N/A	0	N/A	0.987
時期 開始	11	0.999	0.783	0.849	4	N/A	N/A	N/A	0.974
終了	2	0.800	0.800	0.800	0	N/A	N/A	N/A	1
動作 操作	612	0.978	0.978	0.978	59	0.678	0.678	0.678	0.759
判断	39	0.988	0.962	0.973	39	0.667	0.051	0.095	0.759
行為	8	1	0.900	0.933	2	N/A	0	N/A	0.987
説明 方法	67	0.969	0.798	0.861	5	N/A	0	N/A	0.968
禁止事項 内容	98	0.979	0.974	0.976	7	0.500	0.857	0.632	0.956
危険性	39	0.999	0.946	0.970	2	1	0.500	0.667	0.994
徹底事項	16	0.899	0.867	0.880	0	N/A	N/A	N/A	1
推奨事項	2	1	1	1	0	N/A	N/A	N/A	1
理由	39	0.998	0.913	0.948	1	N/A	0	N/A	0.994
目的	10	0.890	0.700	0.766	1	N/A	0	N/A	0.994
補足 使用用品	119	0.970	0.965	0.966	18	0.900	1	0.947	0.987
参考情報	2	0.800	0.800	0.800	1	N/A	0	N/A	0.994

表 5: 機械学習による文の分類結果 (P: 精度, R: 再現率, F: F 値, N/A: 精度または再現率の計算時に分母が 0)

適用可能だと想定している⁴。また文書モデルの応用として、文書診断の可能性を提示し、予備実験として内容要素の自動付与を行った。

今後は、手続き型以外の修理書を対象とした文書モデルの作成と各文書モデルの適用範囲の見極めを行う。また予備実験を発展させて、より頑健な文書診断手法の開発を進めたい。

謝辞 研究データはトヨタ自動車株式会社からご提供いただいた。本研究の一部は科研費 (17H06733) の支援を受けた。

⁴4 節で示したように、「点検」の 10 文書に対しては、準備した内容要素を追加・変更せずに用いることができた。

参考文献

- [1] OASIS. Darwin Information Typing Architecture (DITA) Version 1.3. <http://docs.oasis-open.org/dita/dita/v1.3/dita-v1.3-part3-all-inclusive.html>
- [2] 宮田玲, Hartley, A., 影浦峽. 自治体ウェブサイト文書の多言語展開を支援するシステム環境. 言語処理学会第 20 回年次大会, pp.812-815, 2014.
- [3] Corbin, J. & Strauss, A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory (4th ed.). Sage Publications, 2014.