電子マニュアルの文書構造に対する評価メトリクス

谷口 真也† 川口 真司† 松下 誠† 井上 克郎†‡

†大阪大学大学院 基礎工学研究科 情報数理系専攻
〒 560-8531 大阪府豊中市待兼山町 1-3
Phone: 06-6850-6571 Fax: 06-6850-6574
‡奈良先端科学技術大学院大学 情報科学研究科
〒 630-0101 奈良県生駒市高山町 8916-5
E-mail: {s-tanigu, s-kawagt, matusita, inoue}@ics.es.osaka-u.ac.jp

あらまし CALS 等に代表される文書の電子化においては，文書の再利用性や検索性等を向上させることを目的として，文書を構造化して記述することが一般的である．また，ソフトウェアが幅広い分野で利用されるようになってきたため，質のマニュアルを供給することは重要となってきている．しかし，マニュアルを評価する際には，文書の持つ構造が評価対象となることは少ない．このため，再利用性や検索性等に着目した評価を行うことは困難であった．そこで本研究では，HTML で記述された電子マニュアルを対象として，文書構造の良さを定量的に評価するためのメトリクスを，既存の文書に基づいて統計的手法によって導出した．また，提案するメトリクスによって検出された，質の悪い電子マニュアルに対する改善手法を提案し，具体的に修正を行うことによって文書構造を改善できることを示す．
キーワード 構造化文書，電子マニュアル，メトリクス

Statistics Based Metrics for the Structures of Electric Manuals

Shinya Taniguchi†, Shinji Kawaguchi†, Makoto Matsushita† and Katsuro Inoue†‡

†Graduate School of Engineering Science, Osaka University
1-3 Machikaneyama-cho, Toyonaka, Osaka 560-8531, Japan
Phone: +81-6-6850-6571 Fax: +81-6-6850-6574
‡Graduate School of Inforamtion Science,Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-0101, Japan
E-mail: {s-tanigu, s-kawagt, matusita, inoue}@ics.es.osaka-u.ac.jp

Abstract Most of electric documents such as CALS documents are described with some structured format to improve reusability and searchability. It is important that qualified documents are available with software, since there are lots of areas to apply in the world. However, existing metrics for electric manuals are focused to its contents; it is hard to evaluate reusability and searchability with the point aimed at its structures. In this study, we propose a new metrics of document structures for manuals written in HTML. Our metrics is based on statistics of lots of existing manuals. We also propose a methodology to improve a structure of manual which is detected by our metrics. With our metrics and methodology, we also show that how to improve a low-qualified manual.
Key words Structured document, Electric manuals, Metrics

1　まえがき

近年，ソフトウェアが分野を問わず広く開発，利用されるようになり，その数は増加傾向にある．それに伴って，ソフトウェアの開発・利用に必要なマニュアルもまた膨大な数に及んでいる．ソフトウェアに限らず，マニュアルは，それらが説明すべき事柄に関する知識をほとんど持たない人を対象に記述される文書であるために，その品質にはある一定以上の水準が求められている [7].

一方，効率のよい情報の保存と配布を行うために，さまざまな局面で，文書を電子化する動きが急速に広がっている．このような文書の電子化においては、文書の再利用性，検索性等を向上させることを目的に，文書を構造化して記述することが一般的である．

ソフトウェアのマニュアルは，元来紙媒体で提供されていたが，CD-ROM に代表される大容量記憶媒体や，インターネットなどの普及によって，最近では電子媒体で提供されることが一般的である．そのため，マニュアルの品質を評価する際には，その構造に対する評価も重要となってくる．しかし，現状では文書の内容に関する評価 [8] は行われていても，構造に対する評価はそれほど行われていない．

本研究では、電子マニュアルの構造の良さを定量的に評価することを目的に，大量のマニュアルから品質の劣る文書を検出し，そのマニュアルをより品質の高いものにするための修正法を示す手法を提案する．

まず，参考文献 [9] を元に文書構造と文書構造の品質を評価するための基準を定義し，それを HTML 文書に対して適用した．次に，それらの基準を定量的に評価するための計測値を定め，Web 上から収集した HTML マニュアルを利用して計測値を集約した．最後に，各計測値に関して異常な値を示すデータの分析を行い，それをもとに文書構造を評価するためのメトリクスを定義し，それによって検出される文書の修正法ガイドラインを定めた．

2　構造化文書

構造化文書とは，学術論文，マニュアルのように，意識して文書構造を作成し，それを明示した文書のことである．構造化文書を記述するためには，TeX や SGML[2] に代表される文書構造化言語を用いるのが一般的であり，それらを利用することによって構造が明確で整合性を持つ文書を記述することが容易となる．

構造化文書を用いる目的としては，文書作成のスケジュールやコストの管理が容易になる，文書作成を複数人で分担して行える，文書の品質を一定に保てるといったようなことがあげられる．

実際に構造化文書が用いられている一例としては CALS[6] がある．CALS は米国防総省が資材調達の支援システムとして開発した規格をベースとした，開発者と顧客の間で製品やサービスに関する情報を共有し，設計，製造，調達，決済をすべてコンピュータネットワーク上で行うための標準規格である．CALS においては SGML で記述された構造化文書を用いることにより企業間の情報の共有を目指している．

3　構造化文書に対する構造評価

前章で述べたように，さまざまな目的に応じて文書の構造化が行われているが，その目的を達成するためには適正な構造化を行う必要がある．文書の構造化については，さまざまな文献でその必要性，及び，その方法論について述べられている [1, 3, 4, 5]．本研究では，各文献の要素が統合されて記述されていた参考文献 [9] において述べられている文書の構造化を基準にすえることとした．

3.1　構造化文書

参考文献 [9] における構造化文書とは，文書内容がモジュール単位で記述された文書を意味する．その模式図を図 1 に示している．

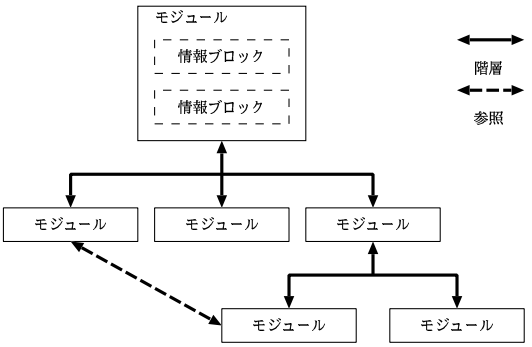

図 1: 構造化文書の模式図

モジュールとは，ユーザに対して一度に提供することが可能な情報量を表す単位であり，一つの機能，一続きの操作，または一つのテーマなどが説明できる程度の情報量である．各モジュール内は情報ブロックと呼ばれるさらに細かな情報量の単位で構造化される．情報ブロックは，意味が伝達可能な情報量を表す単位であり，その大きさはおおよそ小見出し一つ分程度となる．

これらモジュール間の上下関係を示すものが，階層である．一般的に階層は，段階的に層をなすもの全体，またはその各層を意味する言葉であるが，文書においては，章，節，そして，項などの上下関係や，章，節，項そのものを示す．また，文書には階層関係以外のモジュール間の関係を示すものとして，参照が存在する．これは，現在読んでいる文書と関連のある文書中のある箇所を指し示すためのものである．

3.2 文書構造の評価基準

構造化文書が持つ長所を損なわないように，文書の構造化を適切に行うための方法論が，参考文献 [9] では述べられている．3.2.1 節でモジュール，3.2.2 節で階層に関する評価基準について述べる．

3.2.1 モジュール

モジュールとは，ユーザに対して一度に提供することが可能な情報量を表す単位である．ユーザは文書内容をモジュール単位で理解することになるため，構造化文書作成時における文書内容のモジュールへの分割は適切に行わなければならない．

- モジュールのサイズは 1 ウィンドウ程度
  モジュールの具体的なサイズとしては，紙媒体で記述された文書であれば見開き 2 ページ程度，画面上で閲覧する電子化された文書であれば 1 ウィンドウ分，日本語に換算すると約千二百文字程度がよい．これは，ユーザがページをめくらずに情報を一覧できる情報量であり，これを保つことによってユーザが情報を把握しやすくなり，読みやすさが向上すると考えられる．
- 各モジュールのサイズは均等
  ユーザに対する情報の一覧性を保つために，文書中のモジュールのサイズは可能な限り均等に保つことが必要である．つまり，サイズの小さなモジュールは他の関連するモジュールと組合わせ，サイズの大きなモジュールはより小さく分割して文書を構造化することが望ましい．ただし，一続きの操作手順といったような内容的に複数のモジュールに切り分けるのが難しいものに関してはこの限りではない．
- モジュールは複数の情報ブロックから構成
  内容的な理由から一覧性を保つサイズに切り分けられないモジュールに関しては，モジュール内を適度な大きさの情報ブロックに分けることによって，ユーザに対する一覧性を提供する必要がある．また，そのようなモジュールに限らず，モジュール内を複数の情報ブロックに分割することで、読みやすさを向上させることに加えて，文書作成の柔軟性を高め構造化作業をやりやすくすることが可能である．

3.2.2 階層

モジュール単位に分割された文書を階層化することによって，情報のまとまりや上下関係を明確に表現することが可能となる．ただし，その階層化が適切でないと逆に情報の把握が困難になるという結果を生むことになる．

- モジュールが構成する階層は基本的に 3 階層にする
  文書が階層化され情報を把握しやすいように整理されていた場合でも，実際にユーザが文書を読む際には次ページに進むか，前ページに戻るかという二者択一となる．この結果，どんなに階層が深くなってもユーザは情報の並び順に従って読むことになる．しかし，階層が増えることは，ユーザが現在読んでいるモジュールの階層関係を直感的に認識することを困難にするため，情報が細かく分類された階層の深い文書よりは，情報の並び順が十分に考えられた階層の浅い文書のほうが，ユーザにとっては読みやすいといえる．
- 各モジュールの子供は適切な数にする
  必要以上に長い階層は，ユーザが階層関係を理解を妨げる上に，ユーザの文書を読む意欲を減少させるため，文書を構成するモジュールが持つ子は適当な数 (1 桁以内程度) に抑える必要がある．ただし，リファレンス系の文書では 1 つの階層内の項目数が 1 桁を越える場合がある．

4 電子マニュアルへの適用

本研究では，電子マニュアルに対する文書構造の品質に対する定量的な評価をすることを目的としている．そのためには前章で述べた文書構造についてより明確な定義をした上で，その定義をを電子マニュアルに対して適用する必要がある．

4.1 文書構造の定義

前述のように，文書構造を決定する要素としてはモジュール，情報ブロック，階層，参照の 4 つのものが挙げられる．参考文献 [9] におけるこれらの要素に関する定義には曖昧な部分が含まれているため，そのままでは定量的評価に用いることは難しい．そこで，各要素に関してより厳密に定義を行った．

4.1.1 モジュール

参考文献 [9] の定義では読者に一度に提供するための情報量を表すための単位とあるが，このままでは文書中のどの部分からどの部分が一つのモジュールであるかということを決定する因子に欠ける．そこで，本研究ではモジュールを，文書中で見出しによって分割可能な一連の情報と定義する．つまり，モジュールを決定する因子として見出しを用いることにした．これによって、文書を見出しによって，章，節，項という形でモジュール単位に分割することが可能となる．

4.1.2 情報ブロック

参考文献 [9] の定義では意味を伝達可能な情報量を示す単位とあり，モジュールの場合と同様に，モジュール内を情報ブロックで分割するための決定的な因子が存在しない．よって，モジュールの場合と同様に情報ブロックを分割するための因子として，段落を用いることとした．つまり，情報ブロックを各モジュール内に含まれる段落と定義によりすることによって，例えば，一般的な文書で段落の表現に用いられる 1 文字のインデントを用いることでモジュール内を分割することが可能となる．

4.1.3 階層

階層は，モジュール間の上下関係を示す，という参考文献 [9] の定義そのままで十分である．前述のモジュールの定義により，文書が見出しによって分割可能となるため，その見出しの情報をもとにモジュール間の階層関係は容易に判別することが可能である．

4.1.4 参照

階層の場合と同様に参照もまた，関連のある箇所同士を指し示すという定義で十分機能する．モジュールを見出しによって分割した後に，階層関係を判別してやれば，参照は階層関係以外を示すモジュール間の関係として認識可能である．

4.2 HTML で記述された構造化文書

本研究では，電子マニュアルの一例として HTML で記述されたマニュアルを対象とした．HTML 文書には紙媒体の文書にはない，ファイルとリンクという概念が存在する．HTML で記述された構造化文書の模式図を図 2 に示す．

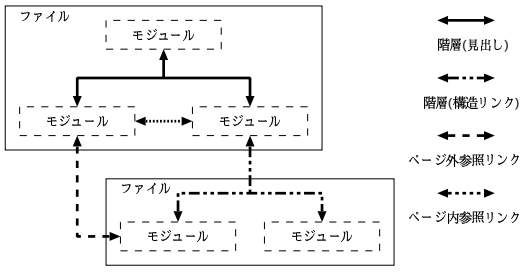

図 2: HTML で記述された構造化文書の模式図

HTML で記述された構造化文書では，単一，あるいは複数のファイルから構成され，そのファイル内に各モジュールが記述される．階層は，同一ファイル内でのモジュール間，あるいは，異なるファイル間のリンクによって構成される．この階層を構成するリンクを本研究では構造リンクと呼ぶ．また，HTML 文書においては参照をリンクによって表現することが可能である．本研究では，その参照先が同一ファイル内であるかそうでないかによって，ページ内参照リンク，ページ外参照リンクと区別して考えることにした．

4.3 HTML マニュアルの評価基準

前節で，HTML 文書には紙媒体の文書にはない，ファイルとリンクという概念が存在することを述べた．そのため，前述の文書構造の評価基準に加えて，それらを HTML 文書に適用したときに生じる構造評価基準についても考えておかなければならない．

- 1 ファイルに記述されるのは 1 モジュール
  HTML 文書においてモジュールをどのようにファイル上に構成するかは筆者の自由である．しかし，構造化文書の特色である，文書作成作業の分担や，再利用性の向上等を考慮すると，1 ファイル上に 1 モジュールがあることが理想的である．
- 1 モジュールにつき 1 つのページ内参照リンク
  一つのファイルに複数のモジュールが含まれるなどファイルに記述されている内容をウィンドウで一覧できない場合，ページ内参照リンクを利用することで，モジュール間の移動がやりやすくなる．そのため，少なくともモジュールの数と同じだけのページ内参照リンクは必要である．
- 関連のあるモジュール間でのページ外参照リンク
  ファイル数や木の深さが大きい文書では，ユーザが文書構成を把握しにくい上にモジュール間の移動もやりづらくなる．その問題を解決するために，関連のあるモジュール間にはできるだけページ外参照リンクを用いたほうがよい

4.4 HTML マニュアルへの構造定義の適用

文書構造の評価基準を HTML 文書に対して適用するためには，文書構造が HTML 文書に対してどのようにマッピングされるかを決定する必要がある．以下，モジュール，情報ブロック，階層，参照の HTML 文書への適用について述べる．

4.4.1 モジュール

モジュールは，文書中で見出しによって分割可能な一連の情報と定義した．HTML タグには，見出しタグ (⟨H1⟩,…, ⟨H6⟩) が存在するので，HTML 文書におけるモジュールを，文書中で見出しタグによって分割可能な一連の情報と定義する．

4.4.2 情報ブロック

モジュールの場合と同様に，HTML には段落タグ (⟨P⟩) が用意されているため，HTML 文書における情報ブロックを，各モジュール内に含まれる段落タグにより記述された段落として定義する．

4.4.3 階層

HTML 文書における階層を次の二つのパターンで定義する．まず，一つ目は同一ファイル内に複数のモジュールが存在する場合である．この場合，階層は見出しタグ (⟨Hn⟩:n は 1 から 6 の整数) の大小 (整数 n の大小) と定義する．一方，モジュールが複数のファイルに存在する場合は，ファイル間のリンク，即ち，前述の構造リンクを階層として定義する．

4.4.4 参照

既に述べたように，HTML 文書においては文書中に明記した参照以外にもリンクによる表現が可能である．しかし，文書中で明記された参照にもリンクを用いることが一般的であるため，HTML における参照を前述のページ内参照リンク，ページ外参照リンクとして定義する．

5 構造化文書の評価手法

4 章において，3 章で述べた文書構造，及び，その評価基準を定量的な評価が可能となるように HTML マニュアルに対して適用した．その適用の結果得られた文書構造を評価するための基準は以下のようなものである．(以下，この番号によって各基準を表す．)

1. モジュールサイズは 1 ウィンドウ程度
2. 各モジュールのサイズは均等
3. モジュールは複数の情報ブロックから構成
4. モジュールが構成する階層は基本的に 3 階層
5. 各モジュールの子供は適切な数にする
6. 1 ファイルに記述されるのは 1 モジュール
7. 1 モジュールにつき 1 つのページ内参照リンク
8. 関連のあるモジュール間でのページ外参照リンク

本章では，これらの評価基準を定量的に評価するためのメトリクスの定義とその結果検出される文書の修正ガイドラインについて述べる．

5.1 文書構造から算出可能な計測値

まず，文書構造から算出可能であると考えられる計測値を導出し，それらと各評価基準との関連を考察した．その変数を以下に示す．() 内は，その計測値が関連すると考えられる評価基準を示す．

- ファイル数 (6)
- 階層の深さ (4)
- 構造リンク数 (4，5)
- ページ内参照リンク数 (7)
- ページ外参照リンク数 (8)
- 外れファイル数:構造リンク (4，5)
- 外れファイル数:ページ内参照リンク (7)
- 外れファイル数:ページ外参照リンク (8)
- 文字数/モジュール (1，2)
- 情報ブロック数/モジュール (1，2，3)
- 子供の数/モジュール (5)
- 文字数/ファイル (7，8)
- モジュール数/ファイルの (6)

各リンク数は 1000 文字単位でその種類のリンクがいくつあるかを示している．また，外れファイル数は，各要素の平均から標準偏差が 2 倍以上離れているファイルの総数である．文字数/モジュール，情報ブロック数/モジュール，子供の数/モジュール，文字数/ファイル (7，8)，モジュール数/ファイルの (6) については，最大値，最小値，平均，分散，標準偏差，正規化分散，尖度，変動係数についてそれぞれ計測する．

次に，無作為に収集した 142 件 (7885 ファイル) の HTML で記述されたマニュアル文書に関して，計測ツールを利用して，前述の計測値を算出した．その結果から，主成分解析などの統計的手法を用いて相互に関連性の高い評価値を集約した．その結果得られた計測値を以下に示す．() 内は，その計測値が関連すると考えられる評価基準を示す．

- 階層の深さ (4)
- 構造リンク数 (4，5)
- ページ内参照リンク数 (7)
- ページ外参照リンク数 (8)
- 文字数/モジュール (1，2)
- 情報ブロック数/モジュール (1，2，3)
- 子供の数/モジュール (5)
- 文字数/ファイルの平均 (7，8)
- モジュール数/ファイルの平均 (6)

文字数/モジュール，情報ブロック数/モジュール，子供の数/モジュールについては，平均，標準偏差，変動係数についてそれぞれ計測する．

5.2 データ分析

収集した HTML マニュアルから得られた各計測値に関して，有意水準 5%で両側検定を行ったときに棄却域に属するデータの分析を行った．その結果を検出された文書の特徴を簡単に述べる．

- 階層の深さ
  - 線形に構成された部分を含む
- 構造リンク数
  - 線形に構成された部分を含む
  - サイズに対して章構成が詳細
- ページ内参照リンク数
  - 単一ファイルに全ての内容を含む
- ページ外参照リンク数
  - トップページに全てのノードのリンクがある
  - ナビゲーションリンクが詳細にはってある
- 文字数/モジュール

平均	文書内をモジュールに分割していない
標準偏差	文書中に極端に大きいファイルがいくつかある
変動係数	トップページに文書内容のほぼ全てが含まれる 様々な大きさのモジュールが含まれる

- 情報ブロック数/モジュール

平均	モジュールが大きいため必然的に情報ブロックが大きくなっている 段落タグの使い方に誤りがある
標準偏差	一部分だけに段落タグが使われている 文書の中に巨大なモジュールが存在し，そのモジュール内に多数の情報ブロックが含まれている
変動係数	一部分だけに段落タグが使われている 文書の中に巨大なモジュールが存在し，そのモジュール内に多数の情報ブロックが含まれている

- 子供の数/モジュール

平均	トップページが全てのノードへのリンクを持つ
標準偏差	トップページがほとんどのノードのリンクを持つ
変動係数	トップページがほとんどのノードのリンクを持つ

- 文字数/ファイルの平均
  - 単一ファイルで全てが記述されている
- モジュール数/ファイルの平均
  - 単一ファイル構成で，サイズが大きい

標準偏差と変動係数に関して検出される文書は概ね似た傾向を示すが，結果となってあらわれる文書には違いがみられた．標準偏差の場合は，ある程度文書が大きく，かつ，ちらばりが大きい時に値が大きくなる文書が検出された．一方，変動係数の場合は，全体的に小さい値のなかに一つだけ大きな値がある文書が検出された．

5.3 構造評価メトリクス

分析したデータをもとに，各評価基準に関して品質の低い文書を検出するための評価メトリクスを定義した．(各番号が本章始めの評価基準のリストと対応している．)

1. この基準に関連のあるデータとしては，文字数/モジュール，情報ブロック数/モジュールの平均値がある．分析の結果，情報ブロック数/モジュールは予想以上に文字数/モジュールとの関連が深く，検出されるデータは類似していた．そこで，この評価基準を検出するメトリクスとして，文字数/モジュールの平均値を用いる．

2. 各モジュールのサイズは均等
   文字数/モジュール，情報ブロック数/モジュールの標準偏差と変動係数をこの基準を計測するための計測値として用意しておいたが，情報ブロック数/モジュールの標準偏差と変動係数から，検出されている文書は評価基準とはあまり関係のない文書であった．これは，情報ブロックの文字数の散らばりがモジュールの文字数の散らばりとそれほど関連が深くなかったためであると思われる．よって，評価のためのメトリクスには文字数/モジュールの標準偏差と変動係数を利用する．同じ因子に属する標準偏差と変動係数から検出される文書は，似た傾向を示すが，結果となってあらわれる文書には違いがみられるということは既に述べたとおりである．よって，評価メトリクスとしては，各計測値において外れ値となった文書の和集合を利用する．

3. この基準を評価するための計測値は，情報ブロック数/モジュールのみであり，これをそのままメトリクスとして利用する．

4. この基準は，深さと構造リンクが関連している．構造リンクは文書構造の階層を構成する要素であるが，対象となる HTML マニュアルでは構造リンク以外にもモジュールの論理構造によって，階層が構成される．本研究では階層を構成する二種類の方法が同じ確率で生じると考え，この基準の評価メトリクスについて，深さは構造リンクの 2 倍の重みづけをして扱うことにした．よって，深さと構造リンクの偏差値を 2:1 の割合で加算し，平均をとったものを評価基準として用いる．

5. 子供の数/モジュールの平均，標準偏差，変動係数がこの評価基準に関連する計測値である．同じ因子に属する標準偏差，変動係数で検出される文書の違いについては 1. でも述べたとおりである．よって，メトリクスとしては各計測値で外れた値をとった文書の和集合をとることにする．

6. これに関連する計測値は，モジュール/ファイルのみであり，これをそのままメトリクスとして利用する．

7. 文字数/ファイルとページ内参照リンクが関連する測定値である．評価基準からメトリクスとしては，文字数/ファイルが 5000 以上で，かつ、ページ内参照リンクが 1 以上を基準として用いる．

8. この評価基準に関連する測定値は，ページ外参照リンクと文字数/ファイルである．しかしながら，これらの測定値において外れ値となったファイルを実際に検証してみたが，特に問題は感じられなかった．この理由としては，参照がモジュール間の相互関連を示すためのものであり，その性質が文書にとって必要十分なものではないからと推定できる．

5.4 文書構造の修正ガイドライン

各評価基準に対するメトリクスにより検出されたデータの修正するためのガイドラインを定めた．(各番号が本章始めの評価基準のリストと対応している．)

1.
   - 見出しタグが使われていない
     見出しタグを利用してモジュールに分割．ファイル規模が大きい場合は，さらに構造リンクを用いた構造化．

2.
   - 極端に大きいモジュールが存在する
     ファイル内を見出しタグで分割し，分割されたモジュールを，ファイル単位で再構成．
   - モジュールのばらつきが大きい
     文書内容を再考し，モジュールを再度分割．

3.
   - 段落タグが大量に使われている
     文書内容を再考し，段落を再構成．
   - 段落タグが使われていない
     段落ごとに段落タグを適切に記述．

4.
   - 文書に線形な部分が含まれる
     線形になっている部分を木構造に再構成．

5.
   - 子供が大量にある文書が含まれる
     まとめられる子供があれば，親モジュールを作成し木構造に構成．

6.
   - サイズの大きい単一ファイルで記述されている
     1 モジュール単位にファイルを分割．
   - モジュールをファイルに配置する基準が未定
     明確な基準を定め，ファイルに配置．

7.
   - 複数のモジュールをもつファイルが含まれる
     モジュール単位でページ内参照リンクを記述する．
   - サイズの大きいモジュールが含まれる
     約 1200 文字ごとにページ内参照リンクを記述．

5.4.1 構造化文書の修正例

ここでは，基準 4 に関して，実際に修正した例を示す．

モジュールが構成する階層構造の深さに関する基準で検出される品質が低いとされるマニュアルは全部で 3 件である．そのうち最も大きく外れていたマニュアル A について，その問題点と修正法について述べる．

マニュアル A は 86 個のファイルから構成されたマニュアル文書である．このマニュアル A は一見しただけでは，特に問題のない構造化文書に思えるが，実際にメトリクスを評価すると深さが 15 という値を示し，平均から大きく外れた値を示した．この理由としては，このマニュアル A がある手順部分に関して，その説明を HTML 的に線形に記述してることにあった．(図 3)

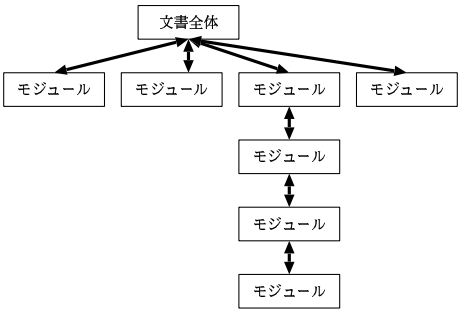

図 3: 修正前のマニュアル A の文書構造の模式図

このため，ユーザはその手順の概要を把握しづらく，また，各手順を直接参照することができなくなっている．よって，ガイドラインに従い，線形部分に関して，その概要を

親モジュールに記述した上で，線形部分が全て親モジュールの異なるような形でマニュアル A の文書構造を再構成した．(図 4)

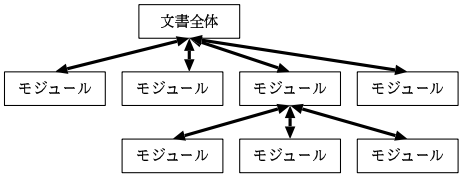

図 4: 修正後のマニュアル A の文書構造の模式図

この結果得られた文書は，先ほど述べた問題点が解消され，深さも 4 と大きく減少し，評価メトリクスにおいても検出されることはない．

6 考察

本手法では，ほとんどの評価基準において，品質の低い文書を検出することができた．また，その結果得られた品質の低い文書を今回示した修正法に従うことで，それらの文書の持つ構造の品質を高くなる．

しかし，基準 3，5，7 においてデータの検出が適切な結果を得られない場合があった．各基準について適切な結果が得られなかった理由について考察する．

基準 3 で検出されたデータは段落タグが必要以上に利用された文書である．この評価基準ではそのような文書の他にモジュール内に含まれる情報ブロックが極端に少ない文書の検出も意図していたが，実際には検出されなかった．これは，今回収集した HTML マニュアルに段落タグをあまり利用していないものが数多く含まれていたためと考えられる．

基準 5 において，実際に検出されるのは、texinfo に代表されるようなトップページに全てのノードへのリンクが張ってある文書である．これは，今回利用したツールが HTML の構造リンクによる階層と，モジュール間の論理構造から構成される階層のうち，前者を優先して判断しているために，孫ノードのように，本来ページ外参照リンクであるものまで構造リンク、すなわち子ノードへのリンクとして捉えているためである．

基準 7 では，大規模であるにも関わらず，単一ファイルに内容が全て記述された文書が検出された．この評価基準では，そのような文献のほかにモジュールをファイルに配置する基準が一定でない文書の検出も意図していたが，そのような文書は検出されなかった．原因としては，評価メトリクスに利用する評価値が十分ではなかったか，あるいは，サンプルデータが不足していたことが考えられる．

7 まとめ

本研究では、電子マニュアルの構造の良さを定量的に評価することを目的に，大量のマニュアルから品質の劣る文書を検出し，そのマニュアルをより品質の高いものにするための修正法を示す手法を提案し，その評価を行った．

その結果，本手法により実際に文書構造の品質が低い文書を検出することが確認できた．

今後は，更に大量のサンプルデータを集めて分析を進めるとともに，文書構造の品質と，HTML 文書の構文的正しさとの関連，あるいは，文書構造の品質と文書の再利用性との関連等について調べていく予定である．

参考文献

[1] Jacques Andre, Richard Furuta, and Vincent Quint, editors : “Structured Documnts”, Cambridge University Press, (1989).

[2] Martin Bryan: “SGML 入門”, 山崎 俊一監訳, アスキー, (1993).

[3] Fred Cole and Heather Brown : “Editing structured documents–problems and solutions”, Electronic Publishing: Origination, Dissemination, and Design, 5(4):209-216, (1992).

[4] Richard Furuta: “Defining and Using Structure in Digital Documents” , Digital Library ’94, (1994).

[5] Robert E. Horn: “Mapping Hypertext: Analysis, Linkage, and Display of Knowledge for the Next Generation ofOn-Line Text and Graphics”, The Lexington Institute, (1989)

[6] 水田 浩: “CALS の実践”, 共立出版株式会社, (1997).

[7] ニューメディア開発協会: “電子マニュアル評価ガイドラインの適正標準化に関する調査研究”, http://www.jtca.org/empg/report98/index.html, (1999)

[8] 高橋善文， 牛島和夫: “計算機マニュアルのわかりやすさの定量的評価方法”，情報処理学会論文誌, vol.32, NO.4, pp.460-469, (1992)

[9] 横河電気 CyberDoc プロジェクト: “デジタル時代のドキュメント企画と設計”，日本理工出版会, (2000).