

構造化文書(ODA)の処理: 文書の基本スタイル設計支援のための知識獲得

4C-6

藤本 正和、田中 剛

富士ゼロックス(株) システム技術研究所

1. はじめに

構造化文書体系では、見出し・本文などの文書スタイルを構成する要素が区別されており、要素に応じたスタイルの設定が容易である。このことから、構造化文書体系に基づく処理の一形態として、文書の高品質化、すなわち高品質な文書スタイルの作成支援が考えられる。

ここでは、WYWIWYG (What You Want Is What You Get)ドキュメンテーション機能^[1]と呼ばれる機能の具現化を考えた。これは、煩雑で専門的な知見を要する文書スタイル設計(文字組/レイアウトなど)の代わりに、目的や印象といった簡単な入力により、高品質な文書が容易に得られる機能である。

このWYWIWYGドキュメンテーション機能の具現化の一手段として、一般ユーザーへのアンケート調査により文書スタイル作成支援のための知識の獲得を試みた。この結果、モノクロでテキストのみの技術文書について、文書スタイルの各要素の設定と、いくつかの心理的な印象との間に、相関関係が見られた。

2. 心理的印象による文書評価

2.1. アンケート調査による評価

今回は、対象となる文書をモノクロでテキストのみの技術文書に限定し、文書スタイルの各要素の設定と心理的な印象との間でどのような関係があるのかを調べることに主眼を置いた。そこで、文章が同じでレイアウトの異なる文書をいくつか作成し、サンプル数11、設問14問、被験者14名という小規模なアンケート調査を行った。

サンプルのレイアウトの変更の際には、「本文の左右マージン」と「本文の文字の大きさ」「(本文の)段組数」「大見出しの位置」「小見出しの位

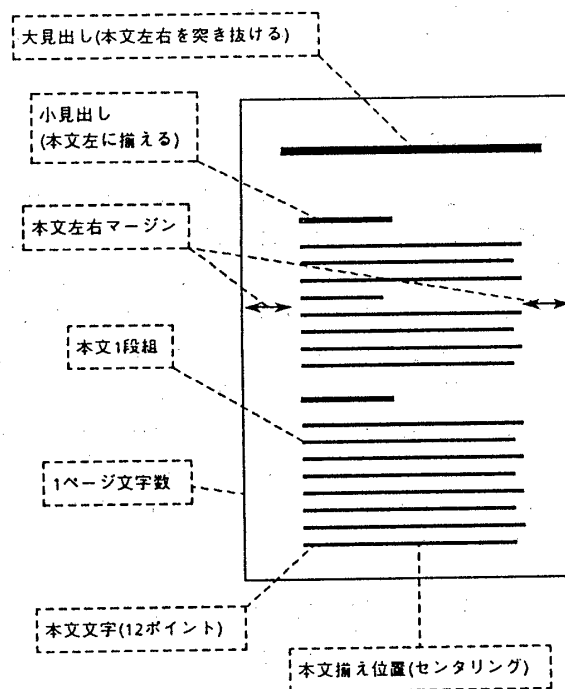


図1 文書スタイル構成要素設定例

置」「本文の揃え位置」「1ページの文字数」の7つを代表的な文書スタイルの要素と考え、これらの設定に変化を与えた。文書スタイルの各要素の設定例を図1に示す。心理的な評価要因については、文書表現品質の評価尺度の抽出^[2]での4つの尺度(注目度を表す因子、機能性を表す因子、ソフトさを表す因子、フォーマルさを表す因子)との相関関係などをもとにして、言葉をいくつか選定した。最終的に14の言葉を、セマンティック・ディファレンシャル法^[3]で用いられるアンケートと同様の手法により、言葉の対を用いて「非常に、かなり、やや」及びどちらともいえないの7段階評価で回答を求めた。

2.2. 分散分析

アンケート調査が、7段階評価を用いており、特性値が計数分類値であることと、分類の組の間に順序があることから、累積法を用いて分散分析を行った。サンプルによる有意差があるかどうかを、各評価要因ごとに、被験者とサンプルによる変動を検定した。この結果、サンプルによる変動は2つの評価要因については危険率5%で有意、その他の評価要因では危険率1%で有意であった。したがって、サンプルごとに評価の違いがあると考えてよいと判断した。

2.3. 重回帰分析

7段階評価を1~7として、各評価要因ごとの評価の被験者平均と文書の各要素の設定値との間で、重回帰分析を行った。この結果、一部の評価項目を除き1次項のみで寄与率0.93以上($r^2 > 0.87$)と非常に高かった。このため被験者平均評価のかなりの部分を文書スタイルの要素の設定値で説明できる。一例として「落ち着いた⇔活気のある」(寄与率0.935)および「良い⇔悪い」(寄与率0.994)という評価項目について得られた、文書スタイルの要素の設定値との間の関係式を(1),(2)式に示す。

$$U_1 = 1.9 \times 10^1 - 2.7 \times 10^{-3} \cdot M_g - 8.0 \times 10^{-1} \cdot S_{bf} + 3.6 \times 10^{-1} \cdot C_1 - 5.4 \times 10^{-1} \cdot T_t + 1.3 \cdot T_s + 4.0 \times 10^{-1} \cdot C_t - 4.6 \times 10^{-3} \cdot N_c \quad (1)$$

$$U_t = 1.0 \times 10^1 + 9.8 \times 10^{-3} \cdot M_g - 4.5 \cdot S_{bf} - 1.6 \cdot C_1 - 1.6 \times 10^{-1} \cdot T_t + 1.7 \cdot T_s + 5.9 \times 10^{-1} \cdot C_t - 1.0 \times 10^{-4} \cdot N_c \quad (2)$$

U_1 : 「落ち着いた⇔活気のある」の平均評価

U_t : 「良い⇔悪い」の平均評価

M_g : 「本文の左右マージン」単位mm

S_{bf} : 「本文の文字の大きさ」単位ポイント

C_1 : 「(本文の)段組数」

T_t : 「大見出しの位置」

本文左右を突き抜ける=1、突き抜けない=0

T_s : 「小見出しの位置」

本文左に揃える=0、本文左をはみ出す=1

C_t : 「本文の揃え位置」

センタリング=1、左寄せ=0

N_c : 「1ページの文字数」

式(2)は、見た目の良い文書と感ずるかどうかを表しており、システムによる見た目の良い文書/悪い文書の自動的な評価が可能である。

3. 文書スタイル設計への応用

今回のアンケート調査では、式(1),(2)のように、心理的な文書の評価が文書スタイルを構成する要素の設定値の一次式で表せた。変数の数以上の評価要因の式が求められれば、連立一次方程式として逆関数を求めることができる。構造化文書体系においては、見出しや本文などの文書スタイルを構成する要素が区別され、各要素別の設定が容易である。そこで、求められた逆関数を用いて、文書に対する印象を指示することにより、文書スタイルを構成する要素の設定値を決定することも可能である。

印象を指定するには、対となる言葉がどちらも否定的ではないことが必要である。なぜなら、「良い」と「悪い」では、「悪い」を指示することは、あまり考えられないからである。しかし「良い⇔悪い」のような例は、「良い」文書スタイルが判別できるため、各要素の設定の際に、制約条件式として用いれば、そのまま文書のスタイル品質の向上に役立つものと思われる。

また、心理的な印象の要求入力によって文書スタイルを構成する要素のパラメータの総てが決まれば、文書スタイルが一意に決定できる。

4. おわりに

今回のサンプルでは、7つのパラメータを変化させて心理的な印象を評価し、高い相関関係が得られた。構造化文書においては、これらの相関関係を定量的な知識として、文書スタイルの各要素の設定値の一部を、心理的な印象の指定に置き換えることが可能である。今後は、より多くのレイアウトパラメータと心理的な印象の関係を明らかにすることにより、ほとんどの文書スタイルを心理的な印象で指定することが可能になると考えられる。すなわち、要求入力の指示方法を、よりユーザーに分かりやすい形に変換することにより、WYWIWYGドキュメンテーション機能が実現できると考えている。

参考文献

- [1] 田中、藤本、池尾: 構造化文書体系(ODA)の処理: 文書の基本スタイル設計支援、情報処理学会第44回全国大会論文集(1992)
- [2] 田中、藤本、漆原: 文書表現品質に関する評価尺度について、情報処理学会第42回全国大会論文集(1991)
- [3] C.E.Osgood, et al.: "The Measurement of Meaning", University of Illinois Press(1957)