

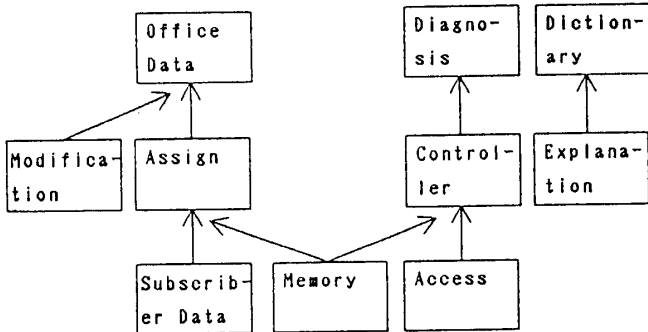
3X-5 キーワード関連構造による マニュアル構成要素検索方式

三谷弘之 千村浩靖
日本電気(株) C&C情報研究所

1. はじめに

マニュアルの作成・管理における課題の1つは、既存のマニュアル構成要素を再利用あるいは引用するための検索を効果的に行なうことである。特に、電子交換機などのシステム製品の場合、マニュアルの分量も多く、その製品のバリエーションによってマニュアルの種類も増え、その機能拡張にしたがってマニュアルの改版も頻繁に行なわれる。このため、効果的な検索方式の重要性が高くなる。

筆者らは、この課題を解決するために、キーワードの関連度に基づくキーワード間の構造をチャート(第1図参照)として表現し、これを用いてマニュアル構成要素を検索する方式を提案する。本方式の特長は、① 適当なキーワードが見つからないときの検索の容易性、② キーワード間の構造の表現の容易性にある。



第1図 キーワードの関連を表わすチャートの例
(英文を対象とした例、カードの中の英字がキーワード)

2. 既存のマニュアル構成要素の検索の要件

既存のマニュアル構成要素の検索では、以下の要件を満たすことが必要である。

- ・再利用あるいは引用するマニュアル構成要素の検索において、適当なキーワードがわからないとき、それが探しやすくてはならない。
- ・マニュアルの改版を頻繁に行なうため、また、それを複数の者で行なうため、キーワードの構

造を表現するチャートの作成および修正が容易でなくてはならない。

本方式は、これらの要件を満たすものである。

3. チャートによるマニュアル構成要素検索方式

本方式は、チャート作成のフェーズと検索のフェーズから成る。

3.1. チャート作成

以下の手順でチャートを作成する。

- ① 書誌情報からのキーワード抽出
- ② 共通なマニュアル構成要素の数をを用いたキーワードの関連度の算出
- ③ 関連度に基づくキーワードの関係付け
- ④ 関係付けに基づくチャートの作成

① キーワードの抽出方法

キーワードは書誌情報から抽出する。チャート作成および修正の容易性のためには、キーワードの自動抽出法(たとえば、金子ほか 1987)を利用すると効果的である。

② キーワードの関連度の算出方法

キーワードの関連度は、キーワード間の意味的近さをキーワードに共通なマニュアル構成要素の数の比率であると定義することによって、算出する。すなわち、以下のとおりである。

$$R_{b/a} = C_{ab} / N_a \dots\dots (1)$$

ただし、 $R_{b/a}$ はキーワード a におけるキーワード b との関連度で、 $0 \leq R_{b/a} \leq 1$ である。 C_{ab}

		キーワード			
		A	B	C	D
文 書	1	○	○	○	X
	2	○	○	X	X
	3	○	○	X	X
	4	○	X	X	X
	5	X	X	X	○
	6	X	X	X	○
	7	X	X	○	○

		キーワード b			
		A	B	C	D
キ ー ワ ー ド	A	\	0.75	0.25	0.00
	B	1.00	\	0.33	0.00
	C	0.50	0.50	\	0.50
	a	0.00	0.00	0.33	\

(a) 文書とキーワードとの関係 (b) キーワード a における関連度: $R_{a/b}$

第2図 キーワード間の関係とその関連度

はキーワードaとキーワードbとをともに含むマニュアル構成要素の数である。N_aはキーワードaを含むマニュアル構成要素の数（以後「キーワードのサイズ」とよぶ）である（第2図参照）。

③ キーワードの関係付けの方法

前節で算出した関連度を用いて関係付けを行なう。ただし、次の条件を満たすものについてのみ関係付けを行なう。

$$R_{a/b} \geq \mu \text{ または } R_{b/a} \geq \mu$$

および

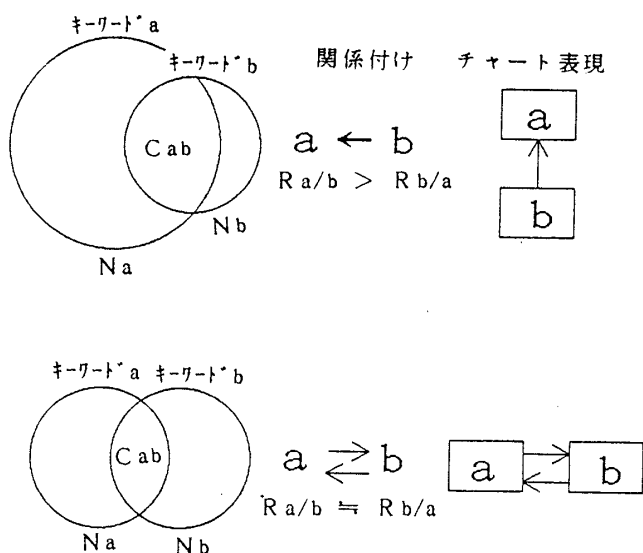
$$C_{ab} \geq \nu \quad \dots\dots\dots (2)$$

ただし、 μ 、 ν はともにキーワード間に関連があるといえるほど充分大きな値であり、経験的に $\mu = 0.3$ 、 $\nu = 10$ である。

以上の条件のもとに、キーワードa、b間の関連度： $R_{a/b}$ 、 $R_{b/a}$ の大小比較によって、関係付けを以下のように行なう。

もし、 $|R_{a/b} - R_{b/a}| \geq \lambda$
 かつ $R_{a/b} > R_{b/a}$ なら、 $a \leftarrow b$
 もし、 $|R_{a/b} - R_{b/a}| < \lambda$ なら、 $a \rightleftharpoons b$
 …………… (3)

ただし、 λ は関連度の差に違いがあるといえるほど充分な値であり、経験的に $\lambda = 0.1$ である（第3図参照）。



第3図 キーワードの関係付けとチャート表現

④ チャートの作成方法

チャート作成には、ISM (Interpretive Structural Modeling) 教材構造化法 (佐藤 1980, 1987) を応用する。ISM教材構造化法とは社会学でのISM法をもとに、教育情報工学の分野で研究開発された構造モデリング法である。この手法を応用する利点は以下のとおりである。

- ・各キーワードは、サイズが大きければ上層の階層に、サイズが小さければ下層の階層に配置される。これは、マニュアル構成要素の検索において検索に使うキーワードの探索と絞り込みに利用できる。
- ・ISM教材構造化法の要素配置アルゴリズムにより、キーワードがクラスタリングされて配置される。そのため、検索に使うキーワードを探しやすくなる。また、このクラスタリングによりキーワードの整理もしやすくなる。
- ・関係付けデータから自動的にチャートを作成できる。したがって、キーワードを自動抽出することにより、チャート作成まで自動的に行なうことが可能となる。すなわち、チャート作成および修正が容易である。

3.2. チャートによるマニュアル構成要素検索方法

マニュアル構成要素の検索には、前節で述べたキーワード間の構造を表わすチャートを用いる。キーワードは、そのサイズによって階層化されて配置されるため、上層の階層のキーワードではそのサイズが大きくなり、下層では小さくなる。したがって、検索キーワードの探索には上層の階層に、検索の絞り込みには下層の階層に移ればよい。

4. おわりに

キーワード関連構造による、既存のマニュアル構成要素の検索方式について述べた。この方式により、① 適当なキーワードが見つからないときの検索が容易、② キーワード間の構造の表現が容易になった。

今後は、チャートの自動作成方法と、キーワードの整理方法を検討していく予定である。

[参考文献]

金子朝男、原良憲、青木祐一郎、会森清、有田裕史、髙原哲 (1987) : 「日本語文章からのキーワード自動抽出」、情報処理学会 第35回全国大会 講演論文集Ⅱ、pp. 1277-1278
 佐藤隆博 (1980) : 「授業設計と評価のデータ処理技法 - ISM教材構造化法とS-P表の活用方法-」、明治図書
 佐藤隆博 (1987) : 「ISM構造学習法」、明治図書
 中村久肇・柴田清 (1988) : 「表層上の手掛かりを用いた日本語テキスト検索・分類システムの試作」、第36回情処全国大会 講演論文集Ⅲ、pp. 2171-2172
 三谷弘之・千村浩靖・藤原尚登 (1988) : 「参照構造の把握に着目した大規模マニュアル管理技法」、第36回情処全国大会 講演論文集Ⅲ、pp. 2245-2246