

フレームを用いた思考支援手法

清末悌之、中川透

kiyosue@nttvdn.ntt.jp, nakagawa@ntthif.ntt.jp

N T T ヒューマンインタフェース研究所

社内の旅費規程文書を対象にして、文書中に記述された内容を利用者に適切に呈示するための課題について検討している。具体的には、次の3点の課題について検討を行なっている。即ち、1) 文書中の不備な点の明確化、2) 文書の表現の抽象度変換、3) 文書中の適切な情報の抽出、呈示アルゴリズムである。1) については、欠陥を明示化するための文書の表現方法DCTを提案し、2) については、抽象化変換のための表示手法としてフレームを用いる手法を提案し、3) については、利用者が意図していなかったことまでも呈示するための文書中の情報の抽出アルゴリズムについて提案した。

Thinking support method by using frame description

Yasuyuki KIYOSUE, Toru NAKAGAWA

kiyosue@nttvdn,nakagawa@ntthif

N T T Human Interface Laboratories

Some themes to present the appropriate contents described in a regulation document for a user are discussed in this paper. Those are following three concrete subjects, i.e., 1) revelation of document inconsistency, 2) transformation of abstractness of terms in a document, 3) extraction and presentation of appropriate contents from a document. For the first subject, representation method to reveal the document in consistency is proposed here. For the second subject, the way to store the information in a document is proposed. For the last one, two algorithms to extract an appropriate contents from a document are proposed.

1. まえがき

社内の旅費規程文書を対象にして、文書中に記述された内容を利用者に適切に提示するための課題について検討している[1][2]。この研究はそもそも、規程文書はどうして読みにくいのだろうかというところからスタートした。ここでは、その基本的な理由を、文書中の部分記述間の関連の不明確さであるとし、これを前提にして、利用者をサポートする枠組みを提案する。

一方、規程文書中の記述をプロダクションルールへ変換することが容易に想像される。しかし、プロダクションシステムが最適な解答を提示しようと試みるのに対し、規程の利用者は複数の取りうる策を比較しながら、即ち、思考しながら読んでいると考えられる。つまり、プロダクションシステムは利用者側の解釈を行なわせないという意味で、利用者をサポートするための枠組みとしては相応しくない。

これらのことを勘案して、規程文書中の適切な情報を利用者に提示するシステム[3][4]に要求される課題について述べ、これに対処する方策を提案することが本発表の目的である。

2. 対象

ここで、本発表で対象としている旅費規程文書について説明する。旅費規程文書は我が社（NTT）の社内文書であるが、いわゆる規程文書としては、ごく普通の体裁を取っているものと思われる（図1）。ページ数は数十ページの規模であり、規

程文書の中ではそれほど多くはない。その内容は、本文と付表からなり、本文は用語の定義部分と規程要項について記述している部分に分かれている。ここでは規程要項の部分について検討を行なう。本文は、章、条、項、号というツリー状の階層構造を論理的な構造としている。

3. 要求条件

規程文書中の適切な情報を利用者に提示するシステム、即ち、規程文書の利用者をサポートするシステムを構築するにあたって、以下の3つの課題が考えられる。

- 1) 文書中の欠陥、不備の克服
- 2) 抽象度変換の必要性
- 3) 利用者の解釈の支援の必要性

1) は、規程文書の表現体系として確固としたものがなく、規程文書中の要素間の関係について明確な把握がされないまま記述されることに問題がある[2]ことについて述べている。またこのことは、規程文書の利用者に混乱を起し、要素間のハイパーメディア的なリンクを張るための障害にもなっている。

2) は、規程文書が基本的に、なるべく広い範囲の事例に対処できるように、抽象的な記述をしているのに対し、これを利用する際には、極めて具体的な事例が与えられることに起因する、両者の表現のズレが存在することについて述べており、規程文書中の適切な情報を利用者に提示するためには、このズレを補正する必要がある。

3) は、2)の問題が克服されたとしても、規程文書のような対象を扱うシステムでは、利用者が提示した具体例に最もマッチングする要素のみを提示すればよいわけではないことを述べている。利用者が入力した情報をもとに、規程文書中の関連する記述部分を抽出し、これに提示のプライオリティを与えて、利用者の解釈を促すといった手段が考えられる。即ち、利用者へ情報を提示し、利用者がこれを解釈するような環境を構築する必要がある[5]。

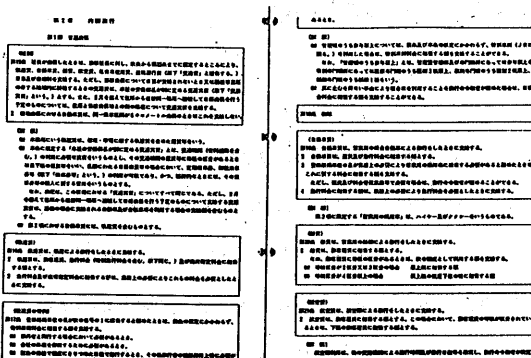


図1 旅費規程文書

以下、これらの課題について具体的に説明するとともに、これらに対する解決策を提案する。

3. 1 文書中の欠陥、不備の克服

まず最初に、規程文書の利用者をサポートするシステムを構築する場合の第一の課題として、文書それ自体の欠陥を挙げることができる。基本的な課題ではあるが、これが利用者の規程文書の理解を阻害していることが多い。ここでいう「文書の欠陥」とは、例えば「例外のみが存在する事例」や「2つの条文がandで結ばれる条件なのか、orの関係なのかが不明確である事例」などのことである。

これらの事例は、しかし、簡単には文書中から見つけだすことはできない。なぜなら、文書の構造、特に各々の記述要素間の関連が陽に示されていないからである。文書中の欠陥とは、また、個々の記述の誤りではなく、この関連そのものの不備であることが多い。ちなみに上記した二つの事例は関連の欠陥そのものである。この不明確な構造は、長年の経験により体得した人だけが、いわゆる規程文書の解釈のエキスパートとして理解できる。

3. 1. 1 DCT表

ここでは、この規程文書中の記述要素間の関連の不備の問題に対処するため、要素間の関係を明確に記述することをめざし、一つの要素を一つの表で表現する手法DCT(Description by Chained Tables)を提案した。DCTによる規程文書の表現の定義を図2に示す。DCTでは、規程文書中の記述要素(単位記述)を一つの表(ここではDCT表と呼ぶことにする)として表現し、この表間の接続、即ち上記した「記述部位間の関連」には以下に述べる3つの関係しか存在しないと規定する。以下図2に従ってDCTを説明する。

図2では、DCT表の定義と、DCT表間の関連について定義がなされている。まず、DCT表の定義について述べる。一つのDCT表は1つの判断対象と2つ以上の判断条件もしくはは

副判断対象、およびそれぞれの条件に対する判断結果から構成される(図2(A))。判断対象は横方向の見出し領域に示され、判断条件もしくは副判断対象は、縦方向の見出し領域に示される。複数の判断条件もしくは副判断対象があれば、適宜行数を増やして(縦方向の見出し領域を分割して)記述する。判断結果は各判断条件もしくは副判断対象の領域の右横に記述する。判断結果の領域には、判断対象が許可されるか否か、もしくは他のDCT表へのポイントが置かれる。

3. 1. 2 DCT表間関係

次に、ここで提案する、DCT表間の3つの関係について述べる。

1) 詳細化:

あるDCT表中で、判断対象に対する判断結果を与えず、別のDCT表で、同じ判断対象について別の判断条件で判断結果を与えるという関係(図2(B))。判断条件を階層状に設定して

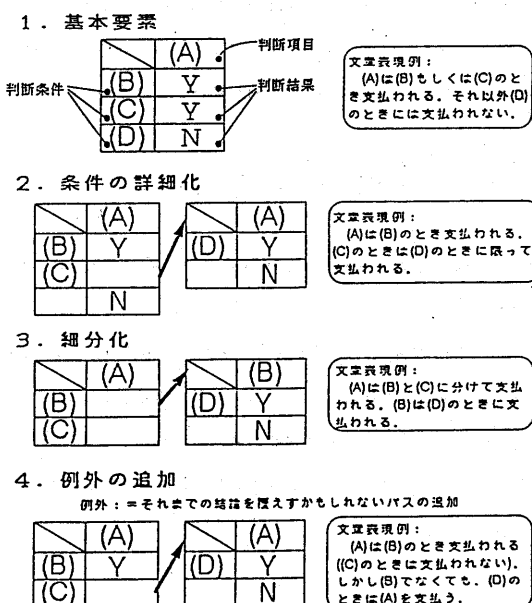


図2 DCT表による旅費規程文書の記述

いる時の記述、もしくは、それまでの判断条件が大まかすぎて、より細かく規定しなければならなくなったときなどに相当する。

2) 細分化:

あるDCT表中で、判断対象に対する判断条件を与えず、判断対象をより細かい判断対象に分割し、即ち、判断対象を副判断対象に細分化し、それぞれを新たな判断対象として別のDCT表で判断結果を与えるという関係(図2(C))。判断対象を階層状に設定しているときの記述、もしくは、それまでの判断対象が大まかすぎて現状の世の中に対応できなくなり、より細かい判断対象ごとに判断結果を与えなければならなくなったときに相当。

3) 例外追加:

あるDCT表で、それまでは許可しないという判断結果を与えていたところに判断結果を与えず、新たに別のDCT表で、同じ判断対象について別の判断条件で判断結果を与えるという関係(図2(D))。

大まかな判断で不許可になっていた判断結果に特例を認めざるを得なくなった時の記述に相当する。詳細化の関係では、今まで許可されていたものをなるべく絞る方向でDCT表間関係が設定されていたのに対し、例外追加の関係では、今まで許可されていなかったものをなるべく許可できるようにDCT表間関係を設定する。まさしく、例外追加の場合に相当する。

3. 1. 3 DCT表を用いた文書の欠陥のチェック

上記DCT表現を用いて、文書中の欠陥を発見する手法を提案する。

文書の欠陥の発見は、DCT表の許されない連結パターンの検出として位置づけられる。ここではそれを、禁止されたDCT表現と呼び、実際に起こる可能性のある連結パターンを図3にいくつか挙げる。以下図3の各禁止された連結パターンについて説明する。

1) 複数の非被参照表の存在:

他のDCT表から詳細化、細分化、例外追加の参照が行なわ

れていないDCT表が二つ以上存在することである。これは文書中に2つ以上の読み始める箇所が存在することに相当する。すべての文書には読み始める箇所は1箇所しか存在しない。なぜなら、結論は1つでも、読み始める箇所が場合によって異なることがあるのは、記述に冗長性が存在することに他ならない。このことは規程文書の利用者を混乱させる要因となる。

2) 孤立した表(群)の存在:

他のDCT表から参照が行なわれてもいないし、他のDCT表への参照も行なっていないDCT表、もしくは参照はなされていてもある閉領域に閉じているようなDCT表群が1つ以上存在することである。このような記述が存在する場合は、規程文書を要素間の関連に拘泥せずに最初から読み下さないかぎり知ることはできない。利用者が規程文書中の情報を獲得することを阻害する原因となる。

3) 網状の表連結:

複数のDCT表の参照関係がループ状をなしている場合である。この場合は、判断結果を得ることができない。

4) 複数の詳細化を行なう表の存在:

一つのDCT表の一つの詳細化をおこなう判断結果から、複数のDCT表へ連結がなされているときである。この場合は、利用者がこれ以降の判断を進めることができない。

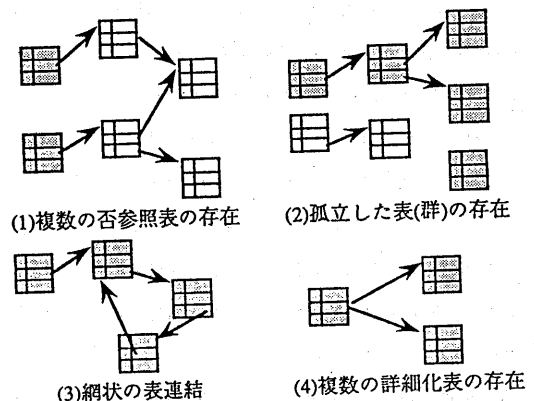


図3 DCT表の禁止された連結パターン

3. 2 抽象度変換

利用者側の具体的な質問に対して、規程文章は比較的抽象的な記述がなされているので、文章中の適切な記述箇所を利用者へ呈示するためには、記述の抽象度の変換を行わなければならない。ここでは、記述の抽象度を、使用されている用語の抽象度の差ととらえ、これを変換するためのしくみを提案する。

用語の抽象度を変換するためには、抽象度を測るための基準(ピボット)が必要である。この基準は概念と呼ぶこともでき、これを明確しておかないと、異なる概念で抽象化される可能性がある。ここでの概念はまた、あるカテゴリ中の最も抽象的な表現であるとも考えることもできる。ここでは、旅費規程を対象にして、質問文中に現われる可能性のある概念を抽出し、これをスロットの名前(slot name)とし、このスロットの集合体としてのフレームで表現する手法を提案する。これを図4に示す。このスロットの値(slot value)に具体値が与えられることで、質問がなされることが考えることができる。規程文書中の各単位記述

(検索される最小単位)にも、このフレームが割り当てられ、スロットの値を埋めることができる。当然のことながら、一つの単位記述が、すべてのスロットの値を与えなくてもよいし、全てのフレーム中の全てのスロットの値が与えられていなくてもよい。また単位記述から導かれたスロットの値は、利用者側の質問から得られたスロットの値よりは概念の抽象度が低い(より抽象的な)表現である。

両者の表現を一致させるためには、抽象度の変換が必要であり、ある概念の中で用いられる用語間の関係を記述しておくシソーラスが必要である。ここでは、旅費規程文書用に、人物、場所、手段、時間についてのみの抽象度変換を考える。例えば、図4中、出張元というスロット名(概念)に対して「自宅」というスロットの値が与えられる質問がされた時、このスロットが場所に関する抽象化のためのシソーラスを持っていれば、自宅→職員データベースの(出張者)の自宅の欄の記述、という抽象度の変換が可能である。(距離に関係する概念の具体値を得るために、この抽象度変換は必要となる。)

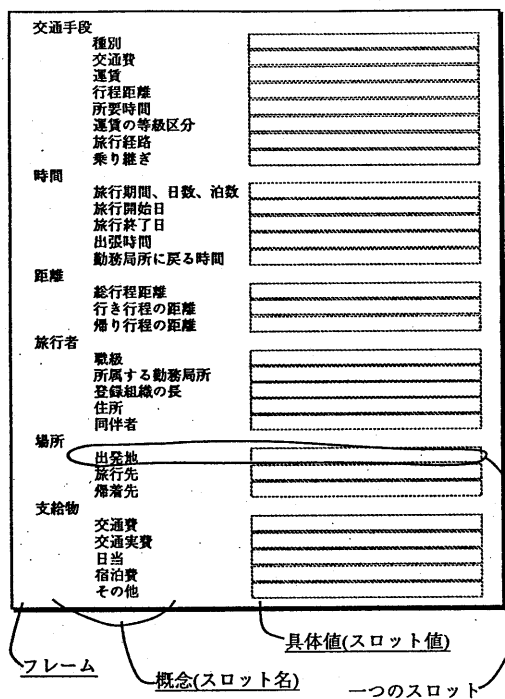


図4 規程文書のフレームによる表現

3. 3 利用者の解釈の支援

最後に、規程文書を利用するときに最も大事なことは、利用者が意図していなかったようなことまでをシステム側が呈示してくれることである、と考える。しかし、全く関係がないことを呈示してくれても意味がない。ここでは、利用者の質問に関係があるが、利用者が意図していなかったと思われる情報を呈示してくれる機構を、抽出アルゴリズムとして提案する。ここでは以下の2つの抽出アルゴリズムを提案する。これらの抽出アルゴリズムは、上記したフレーム表現を利用する。

ここで提案する、第一の抽出アルゴリズムを図5に示す。このアルゴリズムを用いれば、利用者が意図していなかった条件で規定が加えられていることを知ることができる。ここでの処理の流れを以下に述べる。

0. まず、利用者の質問文から作られた一個のフレームと、規程文書中の文章から作られたP個のフレームを仮定する。前者を

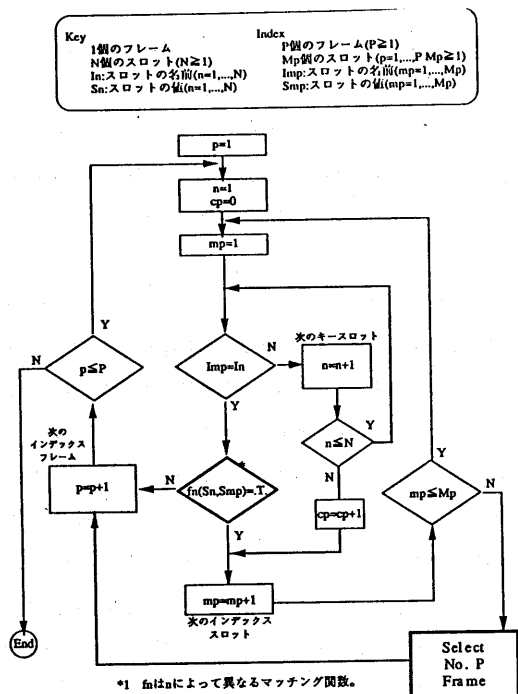


図5 第一の抽出アルゴリズム

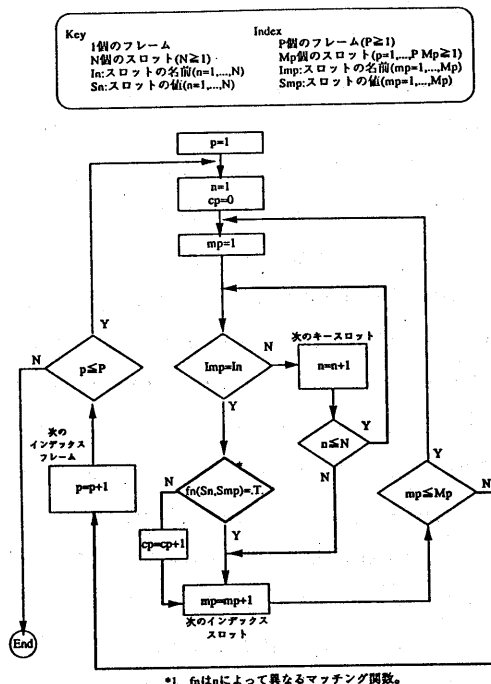


図6 第二の抽出アルゴリズム

キーフレーム、後者をインデックスフレームと呼ぶことにする。

キーフレームにはN個のスロットが存在しているものとし、p番目 ($p = 1, \dots, P$) のインデックスフレームにはMp個のスロットが存在しているとする。前者をキースロット、後者をインデックススロットと呼ぶことにする。

1. あるインデックスフレームpに着目する。

2. p中のあるインデックススロットmp ($mp = 1, \dots, Mp$) に着目する。

3. mpの名前Impが、あるキースロットnの名前Inに等しくなければ、次のキースロット ($n+1$) について3. の繰り返し返し。

4. 全てのキースロットについて3. を行なっても等しいキースロットの名前が存在しなければこのインデックススロットをあきらめ、カウンタCpの値を一つ増やして、次のインデックススロットについて2. 以降繰り返し返し。

5. すべてのインデックススロットについて4. を行なっても等しいキースロットの名前が存在しなければこのインデックスフレームを選択し、次のインデックスフレームについて1. 以降繰り返し返し。

以降繰り返し返し。

6. 3. で、あるキースロットの名前と等しい名前を持つインデックススロットが存在する場合は、両者の値 (キースロットの値Sn、インデックススロットの値Smp) についても比較を行ない、等しければ次のインデックススロットについて2. 以降繰り返し返し。

7. 6. で、SnとSmpが等しくなければ、このインデックスフレームをあきらめ、次のインデックスフレームについて1. 以降繰り返し返し。

8. 選択されたインデックスフレームの中で、Cpの多いものから順に呈示。

第2の抽出アルゴリズムを図6に示す。このアルゴリズムを用いれば、利用者は意図しているが、許可されないことの原因になっていることは知らない事項が存在していることを知ることができる。ここでの処理の流れを以下に述べる。

0. ~4. 第一の抽出アルゴリズム0. ~4. と同じ。

5. 第一の抽出アルゴリズム5.と同じ。ただし、選択はしない。

6. 第一の抽出アルゴリズム6.と同じ。

7. 6.で、 S_n と S_{mp} が等しくなければ、カウンタ C_p の値を一つ増やして、次のインデックスフレームについて1.以降繰り返し。

8. C_p の少ないものから順に呈示。

4. 実験と考察

ここでは、まず文書中の欠陥や不備な点を見だしやすくする表現方法を提案した。実際、規程文書を読んでいて、直感的に不備を感じた箇所を陽に表現することができた。具体的には、ここで対象として取り上げている旅費規程文書中には、例外について記述している条項が突発的に記述されていた。この文書の表現方法は、しかし、人間がチェックすることを想定して提案されている。効率のよいチェックを行なうには、機械的に行なう手法および、そのペリファイのためのユーザインタフェースについても検討を行なう必要がある。

次に、今度は文書中に欠陥がない場合を想定して、利用者の質問に対して、適切な情報を規程文書中から抽出呈示する手法について検討した。そのためにはまず、質問と文書中の記述の抽象度の違いを吸収する必要があることを述べた。また、このために文書中の記述をフレームとよぶ一種のテンプレートを用いて表現する手法を提案した。ここでは、このフレームを設計する手法については述べていない。実用に供するシステムを構築する場合は、この入力変換部分を自動化するという課題が存在する。

最後に、上記フレーム表現を用いて利用者を支援する情報を抽出し呈示するアルゴリズムについて述べた。ただし、このアルゴリズムでは、全く一致するスロットが存在しない場合でも利用者側に呈示することになるため、対象とする文書側のフレーム群にある特定のトピックスに限るなどの工夫が必要である。

以上提案したフレーム表現と抽出アルゴリズムの有効性を確認するために、ワークステーション上にインプリメントした。

ここでは、入力処理について考慮することはせず、図4に示したフレームをそのまま利用者に呈示して、スロットの値の入力を求める。ただし、デフォルトで決まるスロット値はその旨表示される。フレームの値の入力が完了したあとは抽出処理が始まり、選択された規程文書中の情報が呈示される。このインプリメントで、フレーム表現と抽出アルゴリズム基本的な有効性が確認できた。

以上、ここで提案したシステムに似た既存のシステムにデータベースがあるが、これが利用者の意図を尊重し、これに則した出力を行なおうと努めているのに対し[6]、ここで提案した思考支援システムは、蓄積されている情報から検索者の意図していなかったものまで出力することを目的としている。

5. まとめ

規程文書を対象に、その中に存在する規程文書として相応しくない箇所を明らかにする表現方法を提案するとともに、規程文書の利用者をサポートするシステムについて検討した。このシステムの特徴は、利用者が意図していなかったことを呈示するところにある。ここではこれを思考支援システムと呼んだ。

今後は、文書中の不備を抽出する手法と、利用者サポートのシステムとをリンクさせ、その有効性を定量的に確認するための評価手法について検討する予定である。

謝辞 日頃ご指導いただくヒューマンインタフェース方式研究部川嶋部長ならびにご討論いただいたマルチメディア処理方式研究グループの方々に深く感謝します。

参考文献

- [1]清末、中川「連結した表表現による文書の無矛盾性チェックの提案」第4回人工全大5-9,1990
- [2]清末、中川「連結した表を用いる文書中情報の表現の提案」第40回情処全大5C-2,1990
- [3]島「通信技術文書体系化システム-COSMOS」信学研資DE89-35,1989
- [4]「技術文書の知的検索」日経AI,pp.8,1990.9.24

[5]木内ほか「知的検索を目指したConcept Browser」情報研資F

I13-4,1989

[6]長尾ほか「大規模知識ベースの構築と利用」第4回人工知能

全大パネルディスカッション,1990