

4K-12

# オントロジーを用いた検索支援システム —オンラインマニュアルへの適用—

石橋正章 武田正之

東京理科大学理工学研究科情報科学専攻

## 1 はじめに

特定の対象領域のオントロジーを利用する事で、その対象領域のドキュメントに関して、ドキュメントの内容の意味をある程度理解した上でのドキュメントの検索および分類が可能であると考えられる。

本研究は、オントロジーを利用し、いかに有効かつ効率的なドキュメントの検索を行うかという点に関して、UNIXのオンラインマニュアルについて、コマンド名によらず、目的を告げるだけで必要なコマンドについてのマニュアルを得る方法について進めていったものである。ここでは、本研究の概要について述べる。

## 2 オントロジーの利用

UNIXのオンラインマニュアルはmanコマンドにより見ることが可能であるが、あるもの(コマンドやライブラリ関数など)の利用方法を調べるためには不自由することはないのであるが、あることがしくて、それを達成するためにはどうしたらよいのか、といった目的での利用には不自由することもある。これは、マニュアルとコンピュータの利用者の目的とを結び付けるものが存在しないからで、オントロジー(対象領域の概念化の仕様)の利用は、マニュアルとコンピュータの利用者の目的とを結び付けるための有効な手段であると考えた。

また、UNIXのオンラインマニュアルという対象領域に関するオントロジーの定義は、UNIX

初心者の、UNIX関連の知識を補完するものとなりうる。これは、断片的な言葉からでも自分が何を行ないたいかということをはっきりさせるための手助けとなるものであると言い換えることができる。

## 3 システムの概要

本システムの利用対象者は、UNIXの経験があまりない者を対象として、設計した。検索の対象は、manpath以下のディレクトリman1に存在するユーザーコマンドのマニュアルである。利用者は、これらのコマンド名について知らなくてもよい。

利用者は、何をしたいかをシステムに告げるだけでよい。このときの入力操作、対象の順の英文で行なう。例えば「ファイルをコピー」したければ“copy file”とすればよく、この入力によりコマンドcpのマニュアルが利用者に提示される。また、オントロジーの定義に存在する範囲でこれと同意の表現についても同じ結果が得られる。

検索要求から、目的のコマンド名を得る過程について述べる。システムはまず利用者の入力をオントロジーを利用して解析する。この結果をもとに、各コマンドと概念間の関係の対応を利用して入力に対応するコマンドを検出し、そのマニュアルを利用者に提示するという順序である。

## 4 対象領域の概念化の仕様

システムの実現のために、UNIXの一般的なユーザコマンドを概念化した。各コマンドは操

作と対象の関係で表されるものとして定義して操作と対象の概念とその関係からなるオントロジーを定義した。概念のクラスとして、操作と対象そして、対象どうし関係を表すもの3種類を定義した。

これらの各概念は、定義名、クラスのタイプ、単語あるいはその連続した文字列の実例の定義、を含みこれらを利用者の入力の解析に用いている。また、特別に各対象には、概念の包含関係、対象の持つ属性(例えば、ファイルならばサイズや名前といったもの)、許される操作の集合、を定義することができる。また、対象以外のクラスの定義は先に述べた対象と共通の3つの定義のみを可能にした。また、対象どうし関係を表す単語(句)とは、“A to B”や“A from B”といった表現の“to”や“from”に相当するものである。

現在稼働中のシステムでは、各コマンドと、1つの操作と1つ以上の対象が文法どおりに記述したものとを対応づけたものを利用しているのであるが、ここでクラスのタイプを利用して各コマンドの定義を行ない、これを利用してコマンド名を割り出すことで各コマンドの定義から文法に依存した部分を取り去ることができる。

## 5 検索要求の解析

利用者が入力をしてから回答を得るまでの処理について説明する。図1がその概要を示したものである。まず入力された文字列からオントロジーを利用して目的の操作を抽出する。ここでは、オントロジーから操作のクラスを抜きだし、それらの実例と入力の中の単語(句)を比較して、マッチした概念の定義名を列挙していく。さらにその各操作が許される対象をオントロジーから抽出し、それ以降の記述でマッチする対象を同様な方法で検出する。そして、こうして得られた各定義名の列を、4章で説明したオントロジーで定義されたコマンドと概念の列の対応を参照して、対応するコマンドを選び出し、そのマニュアルを提示する。

また、対応するコマンドが複数検出された場合、現在のところ優先度の定義の利用や学習による、候補のしぼり込みは実現されておらず、利

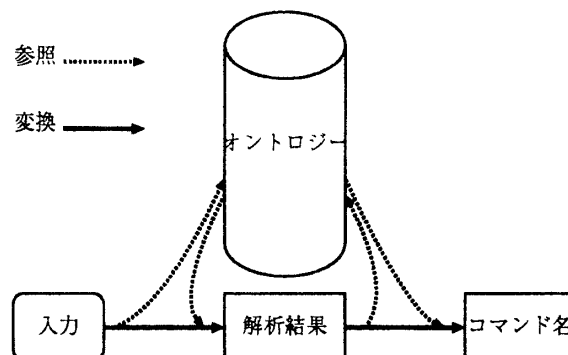


図1: 検索要求の解析過程

用者の判断に任される。

## 6 おわりに

これまでに、文献[3]で説明されている基本的なコマンドについて検索が可能になっている。これについての評価は来年度の演習で初心者を対象にできたらと考えている。

改良すべき点については、現在のシステムでは、入力に対して複数のコマンドが対象として挙がった場合について、最良の選択を自動的に行なうところまでは実装されておらず、各概念間の関係に優先度をつけた解析の実現が求められる。

## 参考文献

- [1] 森田、速水：情報フィルタリングシステム - 情報洪水への処方箋 - 情報処理 8月(1996)
- [2] 岩爪、武田、西田：インターネットからの情報獲得と統合化：人工知能学会(第10回)論文集 p37-40
- [3] 武田、児玉、他：UNIX ワークステーション演習：東京理科大学情報科学科