

マルチメディアコンテンツの構造化に基づく アプリケーション操作環境の構築

Construction of Application Software Operation Environment Based on Structurizing Multimedia Contents

大西可奈子*1
Kanako ONISHI

渡邊裕子*1
Yuko WATANABE

和泉憲明*2
Noriaki IZUMI

小林一郎*1
Ichiro KOBAYASHI

橋田浩一*2
Kōiti HASIDA

*1 お茶の水女子大学大学院 人間文化創成科学研究科 理学専攻
Graduate School of Humanities and Sciences, Advanced Sciences, Ochanomizu University

*2 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

In order to manage multimedia contents, we introduce constraints that naturally exist in our everyday lives, i.e., the constraints on 'time' and 'place', to organize contents, along with the constraints from each event where the contents were obtained. By using constraints on time and place in addition to the basic property information of a file, we can manage the contents by using events as clues. This allows us to look for target content based on the information about a situation under which the target content was obtained. Since information about time and place is added to property information in the form of natural language expressions, it will be easy for contents to be accessed by natural language. Then, we have also developed a natural language interface to manage the contents.

1. 序論

大量化するマルチメディアコンテンツを整理し、利用するため、コンテンツの管理を可能とする様々なアプリケーションソフトウェアが提供されてきた。しかし、これらはコンテンツをファイル作成の日時や種類等の基本情報のみによって管理するため、必要とするコンテンツにアクセスし易いとは言いがたい。

本論文では、日常生活において誰もが自然に記憶にとどめている時間と場所の情報を使ってコンテンツを構造化し、自然言語で操作可能となる手法を提案する。操作対象のコンテンツを言葉で指示できるように、時間と場所を指示代名詞などの言葉を用いて抽象化する。また、言語によるアプリケーションソフトウェア操作を可能にするため、オントロジーに基づいたアプリケーションソフトウェア操作機能の抽象化を行い、抽象化したコンテンツ構造とアプリケーションソフトウェア操作機能を参照することにより、様々なコンテンツ操作を言葉で指示することを可能とする。さらに、構造化されたコンテンツを自然言語により操作する手法を、PCの操作に不慣れな人でも容易にコンテンツの操作を可能にするシステムとして実装することを目的とする。

2. 日常生活オントロジーに基づいたコンテンツ操作システム

本システムは、特定の規準なしに管理されている大量のコンテンツを日常生活オントロジー（後述）に基づき構造化し、管理したものに対して、自然言語インタフェースを通してユーザからの操作命令を解釈し、実行する。本論文では、日常生活において誰もが共有し、動画や写真を記録した時の意図等が反映されやすい時間と場所の情報を、コンテンツを構造化するための基本情報として用いる。例えば、写真やビデオを撮影する際の意図として、主なものはイベントの記録である。ここで、

時間はイベントの「スケジュール」に相当し、場所はイベントの「開催地」である。このように、イベントを2つの視点から構造化し [1]、この構造化された情報をメタデータとして、各コンテンツに半自動で付与し、コンテンツの管理を行う。本研究では時間と場所の情報に基づいて日常生活のイベントを構造化したものを「日常生活オントロジー (ELO: Everyday Life Ontology)」と呼ぶ。このようにして構成されたイベントは階層構造を持つ。スケジュールに基づいたイベントと地理に基づいたイベントが混在してイベントの階層構造を形成することも可能である。また、ひとつのイベントが複数のイベントに属することもある。これにより様々なつながりを持つ複雑な階層構造が形成され、システムはこの階層構造を手掛かりにしてコンテンツの検索を行う。以下にコンテンツが持つ情報で、本システムが使用するものを記す。

- ・ファイル名
- ・ファイル作成日時
- ・ファイルの種類
- ・イベント名（複数可）

ここでイベント名は、前述したELOに基づいて自然言語で記述され、コンテンツに付与される。

3. アプリケーションソフトウェア操作機能の構造化

3.1 ユースケースに基づく操作機能の構造化

本研究では、アプリケーションソフトウェアが提供する機能を、操作対象のコンテンツを利用するパターン（以下、ユースケースと呼ぶ）により分類する。ここで、ユースケースとは、どのアプリケーションソフトウェアでどのような操作を行うことにより何が達成されたのかを個別の事例として具体的に記述したものとす。例えば、MS Word や MS Outlook, iTunes[2]などのユースケースを階層的に分類したものを図1に示す。

図1において、各ノードのラベルには、ユースケースを代表する操作語が割り当てられている。ここで、ユースケースは、アプリケーションソフトウェアを操作することにより達成

連絡先: 大西可奈子, お茶の水女子大学大学院 人間文化創成科学研究科 理学専攻 小林研究室, 東京都文京区大塚 2-1-1, TEL:03-5978-5709, onishi@koba.is.ocha.ac.jp

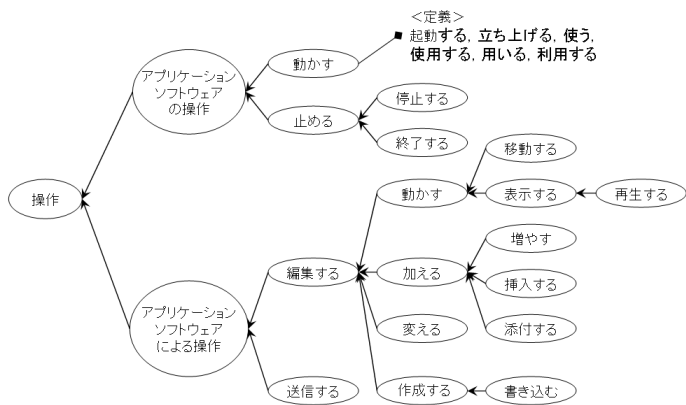


図 1: MS Word, MS Outlook, iTunes のユースケース分類のための機能階層構造

される機能そのものを定義しており、ユーザの操作意図とアプリケーションソフトウェアを結びつける際の仲介役を担うものである。しかし、ユースケースは、異なるアプリケーションソフトウェアの操作機能により具体化されていることがある。このため、ユーザの操作意図からユースケースを代表する操作語を抽出しただけでは、実際の操作を決定することはできない。

各ユースケースが持つ「定義」とは、ユースケースを決定する際に類義語として扱われる言葉を記したもので、この中にある操作意図に対して実際のコマンドの定義も成されている。

また、抽出された操作語に該当するユースケースから実行可能な操作を決定できない場合がある。このようなときには、ユースケースの階層的な分類に従って、ノード間の親子関係や兄弟関係の順に、実行可能な操作を検索することとする。

3.2 ユースケースのモデル記述

アプリケーションソフトウェアの機能をモデル化するために、機能と操作対象との関係を与える。具体的には、ユースケースのモデル記述をオントロジー上のノード間の制約として図 2 のように定義する。これはユーザからの指示により決定した「操作内容」が特定の「操作対象」をとる場合、特定の「アプリケーションソフトウェア」を使用し操作を行うという制約である。図 2 中に現れる「操作対象の階層構造」とは、例えば、図表入りのテキストに対してある操作を行うことができるアプリケーションソフトウェアならば、図表の入っていないテキストも同様の操作を行うことができるという情報を示したものである。

4. 自然言語によるコンテンツ操作システム

4.1 システム構成の概要

提案するシステム構成の概要を図 3 に示す。

自然言語により与えられるユーザの操作命令文は CaboCha/南瓜 [4] によって構文解析された後、解析結果から操作対象となるコンテンツが特定され、コンテンツに対する適切な操作コマンド、および、システムのこれまでの処理履歴がチェックされる。操作対象となるコンテンツの特定は、ELO に基づいて管理されたコンテンツを参照しながら行われる。操作コマンドの特定は、ユーザからの様々な入力表現の意味解釈を支える類語辞書とドメインオントロジーに基づいて構造化されたコマンドセットを参照し、操作時に存在する制約条件（後述）を満たすように決定される。

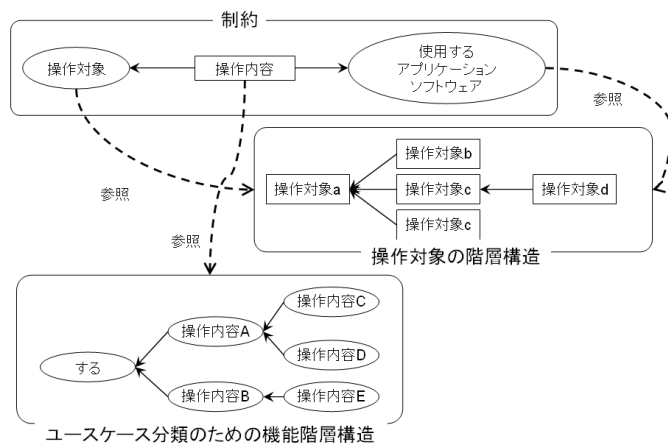


図 2: 制約の定義

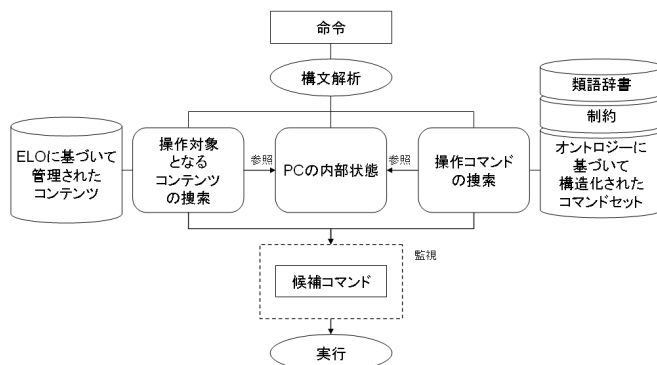


図 3: システム概要

また操作対象となるコンテンツの特定を行う処理と、操作コマンドの特定を行う処理は互いに参照し合い、操作対象コンテンツとコマンドの候補を選出する。システムはこの候補リストを監視し、実行条件が揃ったらコマンドを実行する。これらの処理に関しては 4.4 節で詳しく説明する。

4.2 操作対象となるコンテンツの選出過程

操作対象となるコンテンツの選出の流れを図 4 に示し、各過程の内容を以下に示す。

Step1. 着目しているイベントの抽出

これまでの処理履歴を参照し、前回の着目イベントを抽出する。

Step2. 指示されているイベントの同定

‘次に撮った動画’等、時間的順序関係からコンテンツを指示する命令の場合、システムは、現在のコンテンツの指示状況を PC の内部状態から参照し、ユーザの発話を理解する。順序関係には非常に多量の指定がある。例えば動画の場合は、

- ・ 次の動画
- ・ 次に撮った動画（日時の順序関係）
- ・ 次に行った場所で撮った動画（移動の順序関係）

等の指示方法がある。特に‘次の動画’という指示の場合は、現在選択されているコンテンツがどのような目的で選択されているか、という部分まで考慮する必要がある。

コンテンツの条件を指示する語として‘イベント名’がある場合、それをユーザの指定するイベントとする。さらに、この場合、これまでの操作履歴からこれまでに作成・再生したファ

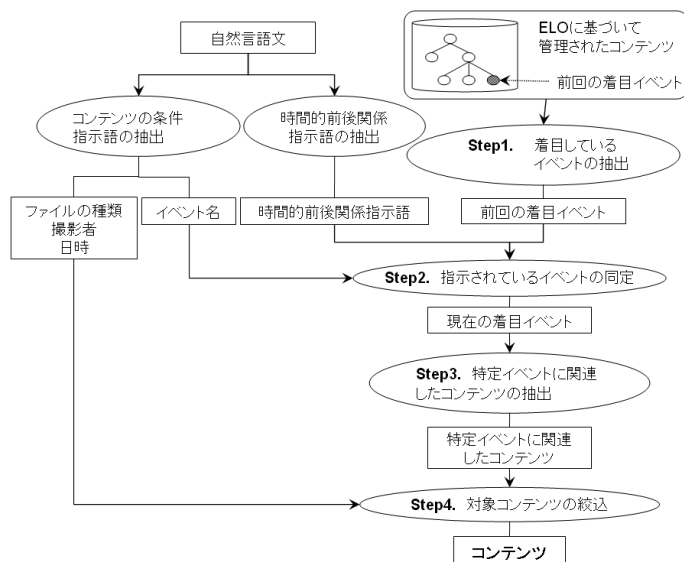


図 4: 操作対象となるコンテンツの選出概要

イルリストを参照し、マッチするものがあればそのファイルリストも操作対象コンテンツとする。これにより、以前、ユーザが操作したファイルリストをそのまま再生することもできる。

Step3. 特定イベントに関連したコンテンツの抽出

Step2 までで特定されたイベントに関連したコンテンツをすべて抽出し、それらを操作対象コンテンツとする。

Step4. 対象コンテンツの絞込

各単語と、‘日時’、‘撮影者名’、‘ファイルの種類’の3項目でマッチングを行う。マッチする項目があれば、それを満たさないコンテンツは操作対象コンテンツから除外する。

4.3 アプリケーションソフトウェアおよびコマンドセットの選出過程

操作コマンド候補の決定法を図5に示す。

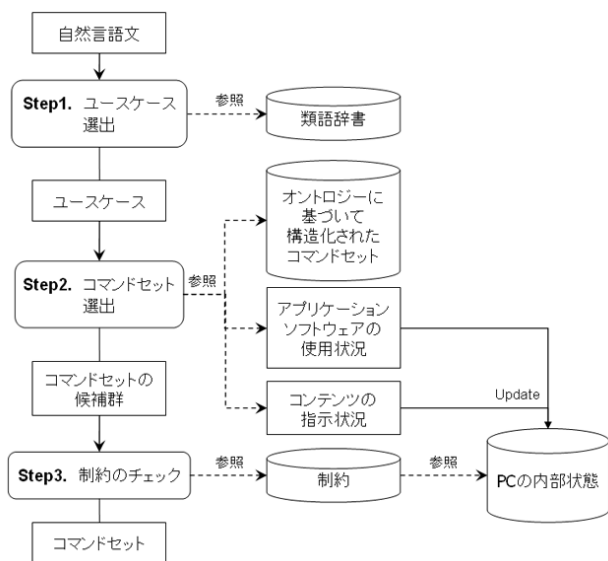


図 5: アプリケーションソフトウェアおよびコマンドセットの選出概要

Step1. ユースケース選出

構文解析によって取り出された単語は類語辞書を参照し、意味的に最も近いと思われるユースケースを決定する。図5において、類語辞書とは意味や表記の揺れを吸収するための辞書で、作成には類語検索大辞典日本語シソーラス [3] を参考に構築した。この辞書を参照することにより、ユーザの操作意図をある一つの表記に決める。例えば、「追加する、付ける、付け加える、付加する」は類語辞書により全て「加える」として扱われる。

次に、このユースケースをプロパティとする部分をドメインオントロジーから参照し、操作コマンドが結びついている場合は次の段階へ進む。そうでない場合は、該当プロパティの上位概念をプロパティとする部分をドメインオントロジーから参照し、操作コマンドが結びついているかをチェックする。これを繰り返して、操作コマンドが見つかった時点で Step2 へ進む。

Step2. コマンドセット選出

なんらかのコンテンツが選択されている場合、および、新たにコンテンツが指定された場合、それらのコンテンツを操作できるコマンドを実行候補コマンドとする。対象コンテンツが操作対象外であった場合は、対象コンテンツの概念階層構造を参照し、該当プロパティの上位概念を扱えるコマンドを探し、存在すればそれを実行候補コマンドとする。それでも適当なコマンドが見つからなかった場合は、プロパティの階層構造から、現在参照しているプロパティの上位概念をプロパティとする部分をドメインオントロジーから参照する。そして上述した手順を繰り返し、該当コマンドを見つける。

Step3. 制約のチェック

ひとつの操作を実行する際に付随する操作を、その操作を達成するために必要な制約として導入する。現在、本システムで扱っている制約は次の2つである。

- ・アプリケーションソフトウェアの起動以外の動作を行う場合は、その操作に必要なアプリケーションソフトウェアが起動していることが必要である。起動していない場合は、該当アプリケーションソフトウェアを起動してからコマンドを実行する。

- ・メール以外の対象を送信するには、対象コンテンツをメールに添付しなければならない。

実行コマンド候補と現在のアプリケーションソフトウェアの状況を考慮し、上記制約を満たさない場合は、制約にしたがってコマンドを増やす。例えば、起動していないアプリケーションソフトウェアを使って操作を行う場合は、コンテンツ操作の前に「アプリケーションソフトウェアを起動する」というコマンドを増やす。

4.4 コンテキストとしての内部状態とコマンドセット実行

システムの状態は、‘現在選択されているコンテンツ’、‘現在選択されているコンテンツが属するイベント’、‘命令の履歴’、‘前回行った操作’の4つで監視される。図6にそれぞれの具体的な内容を示す。

現在選択されているコンテンツとして、コンテンツがアプリケーションソフトウェアによって使用されている現在の状況を監視する。現在選択されているコンテンツが属するイベントが複数ある場合は、該当コンテンツを選択するためにユーザが着目したイベントを、現在選択されているコンテンツが属するイベントとして記録する。また命令の履歴として、これまでにユーザが入力した自然言語文を、前回行った操作として、その命令に対してシステムが行った内容を記録する。

このような PC の内部状態情報は、ユーザの命令を実際の操作に変換する際に用いられる。例えば、図6に示すような

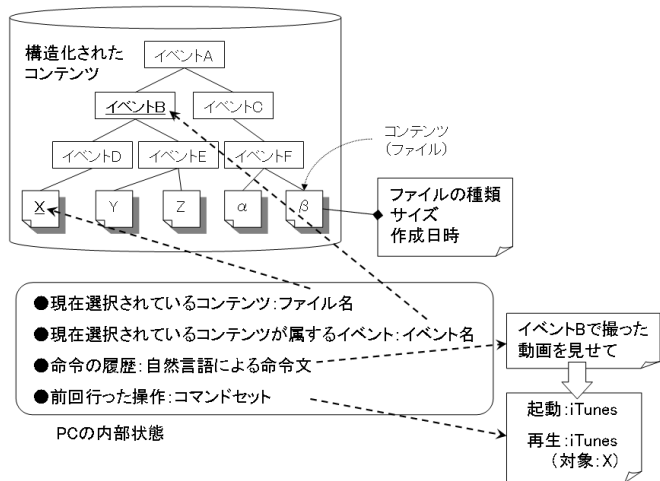


図 6: PC の内部状態

イベント構造に属するコンテンツに対して、ユーザが「イベント B で撮った動画を見せて」という命令を行う。それに対応する操作として、システムは iTunes を起動し動画 X を再生したとする。この状態で PC の内部状態として保持される内容は、図 6 に示すとおりである。この時、ユーザが続いて「次の動画を見たい」と指示したとする。イベント B に属する動画で X の次に撮影されたものが Y であるため、システムは動画 Y を再生する。しかし、「次のイベントで撮った動画を見たい」と命令された場合、システムは、X が属するイベント B の次に発生したイベント C の最初のコンテンツ が該当コンテンツであると判断する。これは本システムが、ユーザからの指示によるコンテンツ移動は、イベントの階層はなるべく変えずに行うというルールを守っているためである。

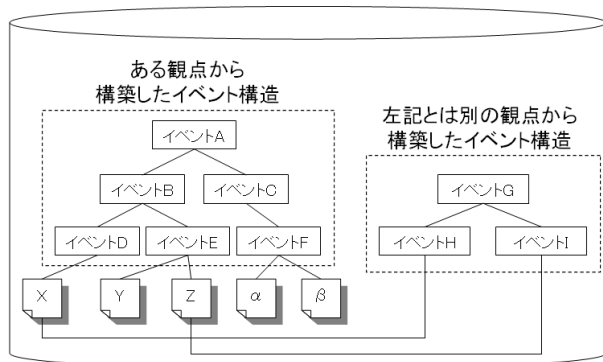


図 7: 複数のイベントに関連するコンテンツ

また、コンテンツは複数のイベントに属してもよい。例えば図 7 に示すようなイベント構造に属するコンテンツが存在するとする。この状態でコンテンツ X は、イベント A、または B、または D、または H に属する X として存在する。そして例えば「イベント H で撮った動画」という指定でコンテンツ X が選択された場合、次の命令が「次の動画」だった場合は、イベント D に属するコンテンツ Z となるが、「次のイベントの動画」だった場合は、イベント I に属するコンテンツ Z となる。このように同じ X を選択するのであっても、ユーザがどの階層に着目して選択を行いたいのか、ということにより状態が変化し、それが次の操作に生かされる。

そして、実行コマンド候補が以下の条件を満たすとき、システムはコマンドを実行する。

- ・操作アプリケーションソフトウェアが使用可能
- ・実行候補コマンドが対象コンテンツを操作可能

5. システムの実装結果に関する考察

本研究では、時間と場所に基づいて管理されたコンテンツを、自然言語を用いて操作する方法を提案した。具体的には、時間と場所によって定義されるスケジュールがイベントとなり、そのイベントの名前が、コンテンツにタグ属性のような形で保持される。このようにして保持されたイベント名を利用することにより、自然言語によるコンテンツ利用が可能となった。また、このようなコンテンツを操作するためのアプリケーションソフトウェアを自然言語で操作するため、本研究では、アプリケーションソフトウェアの機能を、「操作内容」、「操作対象」、「使用するアプリケーションソフトウェア」の 3 つの観点からモデル化した。これにより、本システムはユーザの命令に柔軟に対応することができるようになった。

しかし、現在は時間と場所から取得したイベント名を、そのままコンテンツ利用に用いているのみであるため、コンテンツの利用方法が限られてしまっている。例えば、イベント名として「ランチ」を保持するコンテンツと、イベント名として「夕食」を保持するコンテンツとは、「食事」という共通のイベントに関連すると思われる。このようなつながりを、追加情報として利用することができれば、さらに便利なコンテンツ操作ができる。このように、時間、場所、およびイベント名から新たなイベントを派生させ、コンテンツに関連する新しい情報を利用できるようにすることは今後の課題である。

6. 結論

本研究では、ELO に基づき構造化して管理されたコンテンツを自然言語で操作する手法の提案と、その自然言語インタフェースの開発を行った。ELO に基づき構造化することにより、コンテンツを管理する際、自然言語によるコンテンツへのアクセスが可能となる。すなわち、本論文で提案する自然言語インタフェースシステムを備えたコンテンツ管理法により、コンテンツの指定からアプリケーションソフトウェアの操作まで全て自然言語で指示することが可能となった。また、システムは PC の内部状態を常に監視し、把握する。この情報はユーザの命令を操作に置き換える際に用いられ、これが状況に応じたコンテンツ選択と操作を可能にする。

今後は、コンテンツ構造が持つ情報から新たな情報を引き出すことにより、コンテンツ同士の関係性を使ったコンテンツ選択を目指す。また、辞書に登録されている語彙数の不足を補うことと、操作可能な機能を充実させることにより、より使いやすいインタフェースの開発を行う。

参考文献

- [1] 渡邊 裕子, 大西可奈子, 和泉憲明, 小林一郎, 橋田浩一: 共有スケジュールの構造化に基づくコンテンツ管理, 第 21 回人工知能学会全国大会, 3H6-2, 2007.6.18-22
- [2] iTunes: <http://www.apple.com/jp/itunes/>
- [3] 大修館書店 類語検索大辞典 日本語ソーラス 山口翼編
- [4] 奈良先端科学技術大学大松本研究室: CaboCha/南瓜 <http://chasen.org/taku/software/cabocha/>