

Web を介したコミュニケーションの効率化を目指した 表現メディアの変換

3B-05

岡部正幸 Kryssanov V. Victor
科学技術振興事業団

角所考 美濃導彦
京都大学総合情報メディアセンター

1 はじめに

情報インフラの整備, コンピュータ性能の向上とともに, WWW (World Wide Web, 以下 Web と呼ぶ) を利用して提供される情報サービスが急速に普及しており, 今や Web はネットワーク時代における主要な情報流通メディアの一つとなっている。一方で, Web 上に存在する情報量の増大と情報ニーズの多様化を背景に, “余分な情報が多く, 本当に欲しい情報がなかなか見つからない” といった問題も顕著になってきており, Web をより便利で使いやすいものにするためのユーザ支援環境が Web システムに望まれている[1].

Web を利用するユーザは, 目的の情報を探し出そうと, 検索システムなどを用いて情報選択行動を繰り返すが, Web システムと交わされるこの一連のインタラクションは, ユーザからシステムへの情報要求の伝達とシステムからユーザへの要求情報の伝達を目的とした一種のコミュニケーションと捉えることができる。本稿では, Web 上で情報を見つけるための基本的操作 (ディレクトリ検索, キーワード検索, ブラウジング) が, このコミュニケーションのうちのユーザからシステムへの情報要求の伝達における “言語” の役割を果たしていることを示すとともに, その際にユーザとシステム両者の言語解釈の相違によって生じるコミュニケーション効率の低下について考える。さらに, コミュニケーション効率化の観点から, 上の言語解釈の相違に対処するためのアプローチについて提案する。

2 Web 上でのコミュニケーション言語

Web 上でのナビゲーションは, 基本的にユーザによるリンク・メニュー・キーワード等の選択または入力と, それに対するシステム側からの情報提示というインタラクションが繰り返されることによって行われる。これらリンク・メニュー・キーワード等は, 特定の情報内容を簡潔に表現する見出しとして用いられており, システムから情報を引き出すための言葉, つまりユーザの情報要求をシステムに伝えるためのコミュニケーション言語として捉えることができる。

本章では, ユーザと Web システムとのインタラクションの代表的な形態であるディレクトリ検索, キーワード検索, Web サイトのブラウジングを例に, 実際に Web ページ中のどの要素がコミュニケーション言語の役割を果たしているのかを, 言語を構成する単語や文とその解釈, 文の構成規則などといった観点から説明する。

2.1 ディレクトリ検索

ディレクトリ検索は, 内容に基づいて分類・階層化された Web ページ集合をカテゴリ名を手がかりに検索する方法である。図 1 は検索サイト Yahoo! で提供されているディレクトリ階層をたどって “ラスベガス” に関する情報を探した場合に表示される Web ページを示している。一番上の枠で囲った部分には, この Web ページを表示するまでにたどったカテゴリ名が順に表示されている。ユーザはこの “地域情報” から “ラスベガス” までのカテゴリ名を与えられた選択肢の中から順次選択していくことにより, 目的とする情報 (ラスベガスに関する Web ページ集合) を得る。この場合, 各カテゴリ名は, ユー

¹ Representation media translation for efficient communication on the Web

² Masayuki Okabe

³ Kakusho Koh, Michihiko Minoh

⁴ Japan science and technology corporation

⁵ Center for information and multimedia studies, Kyoto university

ザが目的とする情報を表現するための表現要素となっており、それを指定するには与えられた排他的な選択肢の中から1つを選択する必要があることから、コミュニケーション言語における単語に相当し、ユーザがそのようなカテゴリ名を順次選択していく過程は、システムにカテゴリ名という単語を発話しながら一種の発話文を提示している過程であると見なせる。この発話文は任意に作ることはできず、事前にコンテンツ設計者によって与えられた分類階層という制約に基づいて生成されなければならない。したがって、この分類階層という制約は、カテゴリ名という単語間に成り立つ一種のシンタックス（構文、統辞）と見なすことができる。ちなみに、図1中の真中の枠で囲まれた部分は“ラスベガス”の下位カテゴリ名であり、現在の発話文への追加発話単語候補として捉えることができる。また、この文はシステムに対する情報要求を表すものであり、この文に対して提示されるWebページ集合（図1の一番下の枠で囲った部分）はシステムによるこの文の解釈の結果と考えられる。すなわち、ユーザの入力に対するシステムの出力は、システム側の持つ発話文のセマンティクス（意味体系）を与えるものであると

考えることができる。

2.2 キーワード検索

キーワード検索は、ユーザから与えられるキーワードを手がかりに、Webページ集合を検索する方法である。システムによってキーワード入力だけを用いる、演算子の使用も許す、といった違いはあるものの、いずれも個々のキーワード（や演算子）を単語、それらによって構成される検索式の入力をシステムに対する発話行為と考えることができる。このとき、使用可能なキーワード、キーワードの順序や組み合わせに関するシステム側からの明示的な制約はないことから、2.1と違ってこのときの発話文にはシンタックスはないといえる。一方、それぞれの発話文に対しては、システム独自の基準により順位付けされたWebページ集合が提示されるため、これらの関係を、システム側の持つ発話文のセマンティクスと考えることができる。図2は検索エンジンGoogleに”環境ホルモン”と入力した場合の検索結果を示している。上の枠で囲った部分がシステムに対する発話文、下の枠で囲った部分がその解釈結果である。

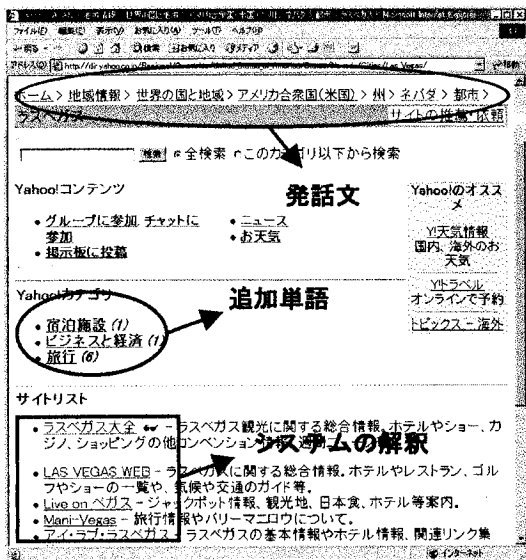


図1 ディレクトリ検索

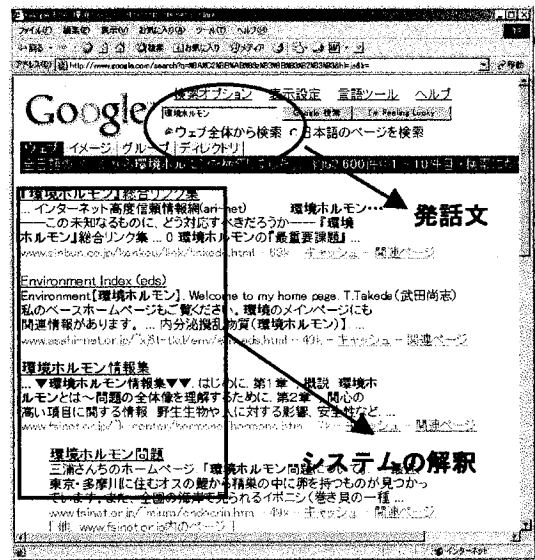


図2 キーワード検索

2.3 Web サイトのブラウジング

ハイパーリンクの選択は Web の最も特徴的かつ基本的な操作である。図 3 は、あるクレジットカード会社のサイトを訪れ、カードの支払日を調べている様子 (枠内の番号は選択順序) を示している。ユーザがまずページ左側にあるメニューから“カードの機能をもっと知りたい”を選び、続いて“ショッピング&キャッシング”を選択すると、右側に“お支払い方法”というアンカーテキストが現れ、このリンク先ページを表示させることにより求める情報にたどり着くことができる。この場合、2.1 と同様に、選択可能な個々のメニューやアンカーテキストが単語に、それらを順次選択していく行為がシステムに対する発話となる。この発話は与えられたメニュー階層とリンク構造による制約の下に行われるため、この制約が、メニュー、アンカーテキストという単語間に成り立つシンタックスと見なせる。また、カードの支払日が記載されたページは、この発話文に対するシステムの解釈の結果と考えられるため、選択されたメニュー、アンカーテキストに対する提示情報の関係を、システム側の持つセマンティックスと見なすことができる。

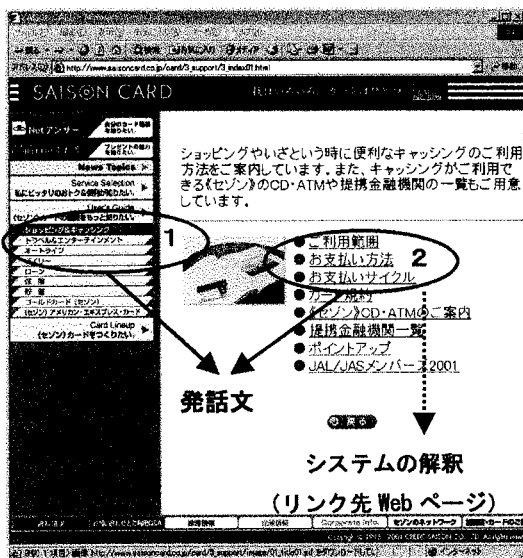


図 3 ブラウジング

3 コミュニケーションの効率性

前章で示したように、Web 上でのナビゲーション機能は、“情報を探すための目印”として付与されているカテゴリ名、キーワード、メニュー項目、アンカーテキストなど (以下では簡単のため“インデックス語”と呼ぶ) を単語として、システムへ情報要求を伝えるための発話文が生成され、システムによる発話文の解釈結果として情報提示が行われる。このように見ると、目的情報を効率よく見つけることは、ユーザが、自分の欲しい情報を伝えるできるだけ簡潔な発話文によって獲得できることであるといえる。ところが、この発話文の単語、シンタックス、セマンティックスは、キーワードを除いてすべて Web コンテンツ設計者によって与えられるものであり、これがユーザにとって自分の欲しい情報を分類・表現する上で適当なものとなっているとは限らない [2]。このように、これらのインデックス語がもつ言語としてのシンタックスとセマンティックスがユーザとシステムとの間で共有されていない状況では、ユーザの入力に対して、ユーザが意図した情報が提示されないといったコミュニケーションギャップが生じる。

本章では、2 章で取り上げた例において実際に起こりうるコミュニケーションギャップの例を示す。

3.1 シンタックスの相違によるコミュニケーション効率の低下

Web 上のナビゲーション操作では、インデックス語の選択肢が情報獲得の効率に影響を与える場合がある。例えばディレクトリ検索において、“ラスベガスに旅行に行きたいので関連情報を集めたい”とした場合、最初に“地域情報”のカテゴリ名よりもむしろ、“趣味とスポーツ”のサブカテゴリである“旅”を選択する人も多いと考えられる。図 4 に示されているように“旅”の下にはラスベガスを連想させるカテゴリ名はなく、実際にそこからは“ラスベガス”のカテゴリにたどり着くことはできない。このように、コンテンツ設計者が用意するディレクトリ階層とユーザが使いやすいと感じるものとの間には相違

がある。

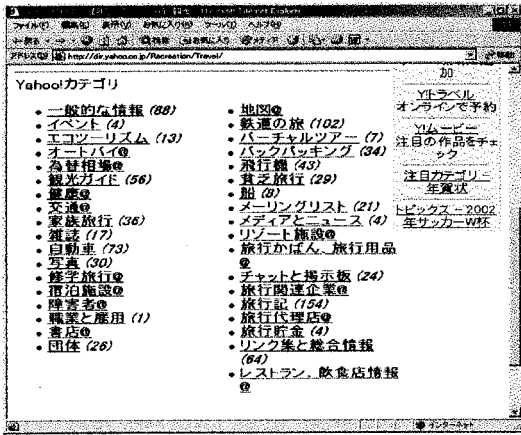


図4 “旅”カテゴリの下位カテゴリ

一方、図5は2章で用いたカード会社のサイトの表紙ページである。左側に5つのメニューが提示されているが、カードの支払日を調べるときに、下の2つのメニュー、“カードの機能を知りたい”、“カードを作りたい”のどちらを選ぶべきか迷うユーザもいると考えられる。実際、後者のメニューからもカードの支払日に関して記述された別のページにたどり着くことができるが、前者のメニューを選んだ場合に比べ、倍以上の回数のメニュー選択を繰り返す必要がある。

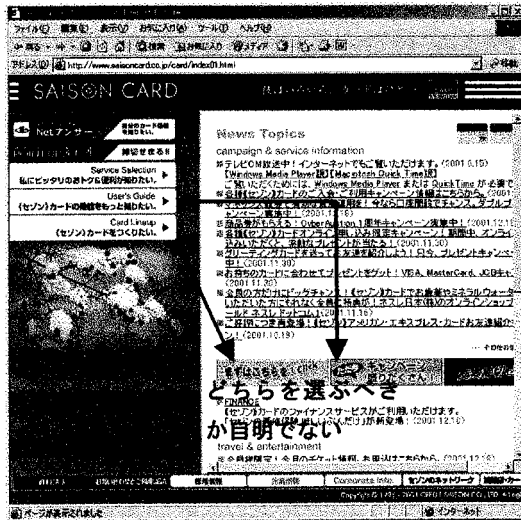


図5 予測の難しいメニュー選択

このようにコンテンツ設計者が暗黙のうちに用いているインデックス語間のシンタックスにより、コミュニケーション効率は大きな影響を受ける。

3.3 セマンティックスの相違によるコミュニケーション効率の低下

情報のアクセス効率を決定する要因はシンタックスだけではなく、単語・文が持つセマンティックスも重要な要素となる。先のインデックス語は、あるクラスの情報を表すラベルといえるが、同じクラスの情報を表すためにユーザが使用するラベルは同一とは限らない。例えば、生化学分野において、「人体内に取り込まれて、人間の内分泌系の働きを攪乱する」という特性を共有する物質のクラスを表現する際には、“環境ホルモン”、“外因性内分泌攪乱化学物質”など複数のラベルが用いられている。このクラスの情報と同じように持っていたても、インデックス語に“環境ホルモン”を用いている Web ページもあれば、“外因性内分泌攪乱物質”しか用いていないものもあるかもしれない。このとき、もしユーザが“環境ホルモン”というキーワードしか思い浮かばなければ後者の情報にはたどり着けないことになる。実際、キーワード検索では、システム内部に必要情報があるにもかかわらず、情報に付与されたインデックス語とユーザからの入力キーワードが異なるため、目的の情報にたどり着けないという結果を招くことも多い[3]。図6は検索エンジンに先の2つのキーワードを入力した結果を示している。上位10ページの中で共通しているページは一つもなく、“環境ホルモン”で検索した結果の中には“外因性内分泌攪乱物質”を用いていないものが多く含まれている。

一方、複数の情報クラスが同じインデックス語を共有している場合、例えば、Windows, Macintosh, Linux という情報クラスは“オペレーティングシステム”というインデックス語を共有していると考えられるが、Windows や Linux の存在を知らず、Macintosh に関する情報クラスを“オペレーティングシステム”というインデックス語によって表現するユーザには、そのインデックス語に対して提示され

る情報のうち、Windows や Linux に関する情報クラスは検索ノイズとなる。図 7 は検索エンジンに“オペレーティングシステム”と入力した場合の結果である。OS そのものに関する説明や様々な種類の OS に関連したページが含まれている。このようにシステムとユーザの間でセマンティックスが異なると、意味を絞り込むために追加の単語入力を行う必要があり、効率の低下を招く。



図 6 同じ情報を異なるキーワードを用いて検索した結果

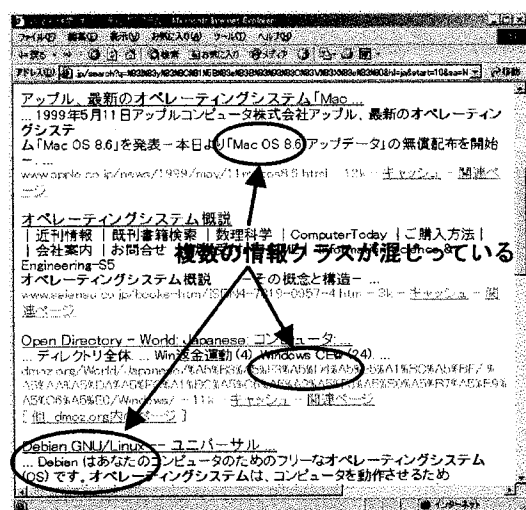


図 7 複数の情報クラスがキーワードを共有している場合の検索ノイズ

以上見てきたように、インデックス語集合が持つシンタックスとセマンティックスがユーザとシステムの間で異なっている場合には、コミュニケーション効率を損なう危険性がある。コミュニケーション効率が低い場合、システムにユーザの情報要求が正しく伝わらず、ユーザにとっては欲しい情報がなかなか見つけられないという状態を招いてしまう。コンテンツ設計者はユーザにとって利用しやすいサイト構成を行おうと努力するが、利用効率を調査するコストは高く時間もかかる。また、ユーザ要求は多様で、時間とともに変化するという理由からユーザにとって使いやすく、かつ複雑な構造をもつ Web サイトをあらかじめ構築するのは困難な作業であるといえる。このような理由から、ユーザ・システム間に存在するシンタックスとセマンティックスの相違を、ユーザとのインタラクションを通じて埋めることのできる適応システムが必要とされている。

4 表現メディア変換

情報ナビゲーションのための言語体系 (インデックス語が持つシンタックスとセマンティックス) は、ユーザが自分の欲しい情報を表現するためのメディアとしての役割を果たしていることから“表現メディア”と呼ぶことができる。ユーザ・システム間のコミュニケーションが効率よく行われるには、当初コンテンツ設計者から提供される表現メディアを、ユーザに適したものに変換する (または変換されたように見せる) システムが必要である。

本章では、まず、情報探索効率の向上を目指した従来研究の中から、ユーザ・システム間のコミュニケーション効率向上のための表現メディア変換技術と捉えることができるものとして「IndexFinder」システム[4,5]、「TalkMine」システム[6,7]について紹介し、さらに現在我々が試作を進めているリンク先読み情報付加システムの特徴について述べる。

4.1 適応 Web サイト

IndexFinder は適応 Web サイト研究のケーススタディとして構築されたシステムである。適応 Web サイ

トとは、ユーザにとって利用効率の高いサイト設計を前もって行うのは困難なため、サイト内ページのアクセス履歴を分析することにより Web ページ間のリンク構造を変化させる、ページの構造（トピックの配置）をフィルタリングするなどして利用効率を上げるための自動的な構造変化を行うサイトのことをいう。IndexFinder は何らかの関連性をもった Web ページへの直接リンクがはられたインデックスページを自動合成しコンテンツ設計者に提案するシステムである。

リンク構造は情報獲得のための発話文を生成する上での一種のシンタックスを表すものであるため、適応 Web サイトはユーザ、またはユーザ集合が持つシンタックスに適応させるためのシステムと捉えることができる。

4.2 適応検索システム

キーワード検索では、入力されたキーワードの各 Web ページ内における重要度を基に優先的に提示するページが決定されるため、各ページに対応付けるキーワードの選択と重要度の計算は検索性能に大きな影響を与える。検索エンジンを代表とする Web 上のキーワード検索システムの多くは、Web ページ中に現れる単語のみを用いた全文検索システムである。このため、3 章で述べたように概念的に等価なキーワードでも、そのキーワードが現れない Web ページを獲得することはできない。TalkMine システムはユーザコミュニティの分析とユーザフィードバックを通じて、Web ページ内における各キーワードの重みを適応的に変化させる。これにより、例えば Web ページ中に現れないキーワードであっても、情報内容を表す適当なものであれば、重みを与えてやることにより、そのキーワードによる検索が可能になる。

2 章で述べたように、入力する発話文の解釈はシステムから提示される Web ページと見なせることからキーワードと Web ページの間の重要度に関する対応関係は、検索システムが持つセマンティックスとすることができる。よって、TalkMine システムは、このセマンティックスをシステム利用者に適し

たものへと適応させるシステムであるといえる。

4.3 リンク先ページの先読み情報の可視化

IndexFinder と TalkMine システムはいずれもコンテンツ提供者が持つデータベースの構造を変化させるものであったが、我々は個々のユーザに適したナビゲーション構造への変換を目指している。先の 2 つのシステムでもユーザ一人一人への対応は可能であるが、ユーザごとにカスタマイズされたシステムが必要となるためコストが非常に高くつく。そのため、我々はサーバから受け取る情報をユーザに提示する前に変換するアプローチ — コンテンツ内の表現メディアがもつシンタックス・セマンティックス自体を変化させるのではなく、各ユーザに適したシンタックス・セマンティックスをもつ表現メディアへの変換規則を獲得する — が有効であると考えている。このアプローチの特徴は、サーバ側に蓄積された表現メディアを変えることなく、ユーザに表現メディアが変換されているように見せることである。図 8 はこのアプローチの概要図である。ユーザとのインタラクションによって獲得された表現メディアの変換規則が Web システムに与える命令、またはシステムから返される結果の翻訳の機能を果たす。

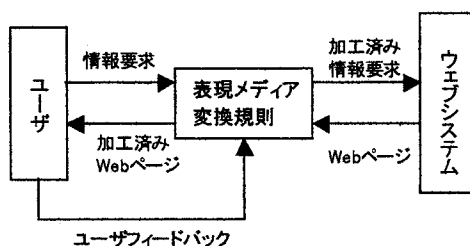


図 8 表現メディア変換システムの構成

筆者らによる検索エンジンのフィルタシステム [8]は、検索エンジンが返す結果から必要なページのみを選び出すシステムであり、図 8 に当てはめると、ユーザフィードバックを利用して生成される検索ノ

イズの除去フィルタが表現メディアの変換機能を果たしている。現在、我々はこのアプローチをハイパーリンクの先読み情報(リンク先ページの内容要約、過去の参照履歴など)の獲得へ適用することを試みている。先読み情報の付加は、図 8 における加工済み Web ページを生成することにあたる。Weinreich[9]らは、リンク選択時におけるユーザの認知的負荷が大きいことを指摘しており、リンクの先読み情報を適切に与えることが必要であると述べている。リンクの先読み情報リンクの先読み情報は、文を生成するために与えられたシンタックスと文生成における単語選択の優先順位をユーザに知らせるものでなければならない。ユーザの情報ニーズに合わせた適切な情報の選択が必要である。我々は表現メディアがもつ言語的性質を利用することにより、適切な先読み情報の獲得が行えると考えている。

5 おわりに

本稿では、ユーザが Web 上で情報を見つけるために行う操作を一種の言語と見なし、Web コンテンツに付与されたナビゲーション情報が持つシンタックスとセマンティックスについて具体例を通して説明した。また、ユーザとコンテンツ設計者の間に存在するこのシンタックスとセマンティックスの相違が両者のコミュニケーション効率を低下させ、ユーザに情報が見つげにくいと感じさせる要因になっていることを指摘した。さらに、この相違に対処するための研究事例と我々のアプローチについて説明した。

現在、我々はハイパーリンクの先読み情報提示システムに必要な Web コンテンツがもつシンタックスとセマンティックスのモデル化を行うとともに、インタラクションを通じてユーザに適したモデルに変換する方法、この変換に基づく先読み情報の生成方法の構築を目指している。

謝辞

本研究は、科学技術振興事業団戦略的基礎研究事業「デジタルシティのユニバーサルデザイン」の助成を受けた。

参考文献

- [1] リチャード・S・ワーマン, それは情報ではない—無情報爆発時代を生き抜くためのコミュニケーション・デザイン, エムディエヌコーポレーション (2001)
- [2] 藤田節子, インデックスのテクニック, 共立出版 (2001)
- [3] 徳永建伸, 言語と計算—5 情報検索と言語処理, 東京大学出版会 (1999)
- [4] Perkowitz, M. and Etzioni, O., Adaptive Web Sites, *Communications of the ACM*, Vol.43, No.8 (2000)
- [5] Perkowitz, M. and Etzioni, O., Towards adaptive web sites: Conceptual framework and case study, *Artificial Intelligence* (2000)
- [6] Rocha, L.M., TalkMine and the Adaptive Recommendation Project, In *the Proceedings of the ACM Digital Libraries*, pp.242-243 (1999)
- [7] Rocha, L.M., Adaptive Webs for Heterarchies with Diverse Communities of Users, Paper prepared for the workshop *From Intelligent Networks to the Global Brain: Evolutionary Social Organization through Knowledge Technology* (2000)
- [8] Okabe, M. and Yamada, S., Interactive Web Page Filtering with Relational Learning, *The First Asia-Pacific Conference on Web Intelligence* (2001)
- [9] Weinreich, H. and Lamersdorf, W., In the Proceedings of the 9th International WWW Conference (2000)

