

Title	カテゴリに基づく製品情報の組織化
Author(s)	有賀, 忠徳
Citation	
Issue Date	2000-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1353
Rights	
Description	佐藤理史, 情報科学研究科, 修士

カテゴリに基づく製品情報の組織化

有賀 忠徳

北陸先端科学技術大学院大学 情報科学研究科

2000年2月15日

キーワード: ワールドワイドウェブ、製品一覧表、製品情報、情報抽出、自動編集.

インターネットの普及により、ワールドワイドウェブ (World Wide Web; 以下 WWW と略記する) 上には多くの製品情報が公開されるようになった。これらの製品情報は、消費者にとって、買物における意思決定の際に大変有用な情報となる。

買物における意思決定には、製品選択とショップ選択の2つの段階が存在する。

第一段階の製品選択は、「どの製品を買うか」を決定する段階である。我々が買物を始める際、例えば「テレビを買おう」のように、既にどのような種類の製品を購入するのが決めているのが普通である。このような製品の種別を、以下では製品カテゴリと呼ぶ。例えば、家電製品ではテレビやビデオなどが、この製品カテゴリとなる。第一段階の製品選択とは、「ある製品カテゴリにおいて、購入する製品を具体的に選び出すこと」と捉えることができる。

購入する製品を決定した後は、第二段階のショップ選択に進む。ショップ選択は、「どこから(どこで)買うか」を決定する段階である。多くの人々は、いくつかのショップで価格やサービスを比較し、できるだけ有利な条件で購入できるショップを選ぶのが普通である。

以上のような買物の二つの段階において重要なことは、「比較する」という作業である。

製品選択の段階では、同じ製品カテゴリに属する複数の製品を比較することが必要である。一方、ショップ選択の段階では、複数のショップにおいて、ある特定製品に関する情報(価格やサービス)を比較することが必要となる。これらの比較作業を簡単に行なえるようにすることが、買物における意志決定支援の重要な要素となる。

しかしながら、現在の WWW は、このような比較支援の機能を、十分には提供できていない。

製品選択段階における比較では、ある製品カテゴリに属する複数の製品の情報が必要となる。しかし、それらは、通常、それぞれの製品を製造している複数のメーカーのウェブサ

イト中に分散して存在している。このため、例えば、テレビの購入を検討している場合、テレビを製造している複数のメーカーのウェブサイトを開覧し、それぞれに対して、そのメーカーが製造しているテレビにどんな製品があり、どのような特徴があるのかを調べなければならない。

ショップ選択の場合も同様である。オンラインショップの場合は、通常の販売店のよう
に、実際に足を運んで価格を調べる必要はない。しかし、それぞれのオンラインショップ
のウェブサイトを開覧し、購入を予定している製品の価格を調査しなければならないこと
には変わりがない。

このような煩雑さは、WWW 上の製品情報が、ユーザにとって望ましい形で組織化さ
れていないことに原因がある。

ユーザにとって望ましい組織化とは、次のような組織化である。

- 製品選択の段階

製品カテゴリ毎に、製品情報（製品の特徴）が一覧できるように組織化されている
こと。このような組織化は、同一製品カテゴリの複数の製品を比較することを容易
にする。

- ショップ選択の段階

特定の製品毎に、各ショップでの価格（やサービス）が一覧できるように組織化さ
れていること。このような組織化により、その製品を最も安く購入できるショップ
が簡単に発見できる。

一方、現在の WWW においては、情報は、提供者毎に組織化されているとみなすこと
ができる。

- 製品情報

その製品を製造しているメーカー毎に組織化されている。

- 価格情報

その製品を販売しているショップ毎に組織化されている。

このような組織化の不整合は、WWW 上の製品情報をユーザにとって使いやすい形態
に整理し直すこと、すなわち、製品情報の再組織化によって解消できると考えられる。

ショップ選択を支援するための製品情報の再組織化にはいくつかの先行研究が存在する
が、製品選択を支援する製品情報の再組織化は、これまでほとんど研究が行われていない。

本論文では、製品選択を支援するために各メーカーサイトから製品情報を抽出し、それら
の製品情報を再組織化するシステムを提案する。本システムは、カテゴリ毎に製品情報を
まとめたカテゴリ別製品一覧表を自動作成することを通して、製品選択時においてユーザ
の意思決定を支援する。

カテゴリ別製品一覧表とは、特定カテゴリの製品に関する情報を提示した一覧表のこと
で、以下の条件を満たしている必要がある。

1. そのカテゴリの製品を網羅する（できるだけ多くの製品を含む）。
2. それぞれの製品に対して、その製品の主要情報を含む。

作成したシステムは、製品情報収集モジュール、実売価格調査モジュール、一覧表生成モジュール、の3つのモジュールと、製品情報データベース、領域知識から構成される。製品情報収集モジュールはメーカーサイトからカテゴリ名を付与した形で製品情報を抽出し、製品情報データベースを作成する。収集した製品情報を用い、実売価格調査モジュールがオンラインショップから実売価格を調査し、製品情報データベースへ保存する。一覧表生成モジュールでは、ユーザの要求を受け、製品データベースから製品一覧表を生成する。

本システムの中核となる製品情報収集モジュールは、カテゴリ URL 収集モジュール、製品情報抽出モジュール、の2つのサブモジュールから構成されている。カテゴリ URL 収集モジュールでは、製品ページへのリンクがまとめられている表であるカテゴリ一覧表を収集し、その中からカテゴリ名とカテゴリ URL の抽出を行う。

製品情報抽出モジュールでは、カテゴリ URL を手がかりに製品ページを収集し、収集した製品ページから製品情報を抽出する。具体的には、カテゴリ URL を出発点として、製品一覧表を用いる方法、製品ページの特徴を利用し得点付けを行う方法、の2つの方法により製品ページを収集する。収集した製品ページからは、仕様表解析、ページ見出し解析、強調文字抽出、により製品情報の抽出を行う。

最終的にシステム全体の評価実験として、家電製品を製造している5つのメーカーにおいて、製品情報収集実験を行った。まず、カテゴリ URL 収集モジュールでは、人手で収集した22のカテゴリ一覧表全てを収集することができた。

製品情報収集モジュールでは、人手で収集した161の製品のうち145を収集することができ、抽出率は90%だった。また、主要属性は161のうち124に対して、正確な値を抽出することができ、抽出率は77%だった。これはメーカーサイトに、グラフィックデータを多用している、バラエティーに富んだレイアウトを含んでいる、など機械化を妨げる要素が多いことを考慮すると良い結果であると言える。

これらの実験結果から、上記の条件を満たすカテゴリ一覧表を生成することができ、製品選択における意思決定の際に有効な情報を提供するシステムを作成することができたといえる。